



HAL
open science

Vers une modélisation dynamique de l'activité cérébrale pour la conception d'interfaces cerveau-machines asynchrones

Cedric Gouy-Pailler

► **To cite this version:**

Cedric Gouy-Pailler. Vers une modélisation dynamique de l'activité cérébrale pour la conception d'interfaces cerveau-machines asynchrones. Traitement du signal et de l'image [eess.SP]. Université Joseph-Fourier - Grenoble I, 2009. Français. NNT: . tel-00440836

HAL Id: tel-00440836

<https://theses.hal.science/tel-00440836>

Submitted on 11 Dec 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Thèse

**Vers une modélisation dynamique de l'activité
cérébrale pour la conception d'interfaces
cerveau-machines asynchrones**

présentée par

Cédric GOUY-PAILLER

En vue de l'obtention du grade de

**Docteur
de l'Université Joseph Fourier**

dans la spécialité

**Biotechnologie, instrumentation, signal et imagerie pour
la biologie, la médecine et l'environnement**

Soutenance publique le 1^{er} Octobre 2009

Composition du jury :

Dr. Jean-Philippe LACHAUX, Président du jury

Dr. Maureen CLERC, Rapportrice

Prof. François CABESTAING, Rapporteur

Dr. Olivier BERTRAND, Examineur

Dr. Marco CONGEDO, Directeur de thèse

Prof. Christian JUTTEN, Directeur de thèse

Travaux réalisés au sein du laboratoire GIPSA-lab, UMR CNRS 5216

École Doctorale Ingénierie pour la Santé, la Cognition et l'Environnement (EDISCE)

*À mes parents, à mes sœurs,
À Julie,*

Remerciements

Les trois années écoulées depuis le début de ma thèse furent jalonnées de nombreuses rencontres, discussions, réunions et séminaires avec des personnes aux caractères, origines et domaines de compétences très divers. Scientifiquement ou humainement, mes travaux leur doivent beaucoup et les quelques lignes qui suivent ont pour but de leur exprimer toute ma gratitude.

En premier lieu, je veux remercier tout particulièrement mes deux directeurs de thèse Marco et Christian. Depuis le début de cette aventure et la première réunion qui nous a permis de définir le sujet jusqu'à l'aboutissement de ces travaux, vous avez toujours su m'épauler et me communiquer votre enthousiasme afin de mener à bien mes recherches. Vos idées fécondes ont inspiré l'ensemble de mes travaux et vos expertises respectives ont largement contribué à améliorer la qualité de mes recherches. Merci encore pour vos relectures pointilleuses de mes différents écrits, même lorsque les délais impartis étaient très réduits. J'ai énormément appris scientifiquement à votre contact, mais j'aimerais également vous témoigner ma grande admiration pour vos qualités humaines et votre disponibilité. Celles-ci ont largement contribué au plaisir que j'ai toujours pris à m'investir dans mes travaux. Je n'oublierai également jamais tous les bons moments que j'ai passés en votre compagnie. Merci encore pour tout !

Je voudrais également remercier tous les membres de mon jury pour avoir accepté de lire et critiquer mes travaux. Un grand merci à Maureen Clerc et François Cabestaing pour leur travail minutieux de relecture et leurs précieuses suggestions qui ont contribué à améliorer la qualité du manuscrit. Un grand merci à Jean-Philippe Lachaux pour avoir accepté de présider le jury, pour ses questions très pointues sur les aspects cognitifs de mes travaux. Merci enfin à Olivier Bertrand pour avoir pris le temps de relire le manuscrit et de participer à distance à la soutenance malgré un emploi du temps extrêmement contraint.

J'ai eu lors de ces trois années le plaisir de collaborer avec un grand nombre de personnes, certaines parties de ce manuscrit sont issues de ces différentes collaborations. J'aimerais tout particulièrement remercier Reza Sameni pour m'avoir tant appris sur le plan scientifique aussi bien que moral ; Sophie Achard pour les diverses collaborations très enrichissantes que j'ai eues avec elle, le magnifique séjour à Cambridge qu'elle m'a permis de réaliser et les nombreuses discussions autour des repas à la cafétéria ; et Jérémie Mattout avec qui j'ai pu échanger sur diverses parties de mon travail. Je tiens à remercier tout spécialement Bertrand Rivet, j'aurais aimé avoir un peu plus de temps pour collaborer avec toi mais je te suis d'ores et déjà très reconnaissant pour tout ce que j'ai appris au cours des discussions que nous avons pu avoir. Ma collaboration avec Gert Pfurtscheller et Clemens Brunner fut l'une des plus fructueuses de mes travaux, je tiens à les remercier chaleureusement pour le temps qu'ils ont passé à répondre à mes questions, leurs conseils judicieux et leurs corrections minutieuses. J'ajoute que ça a été pour moi un grand honneur de pouvoir collaborer avec l'un des pères des interfaces cerveau-machines. J'ai enfin eu beaucoup de plaisir à collaborer avec Jérôme Lemoine, Pierre-Olivier Amblard et Alexandre Barachant sur des thèmes connexes à mon sujet de thèse et que je souhaite pouvoir continuer à travailler avec vous.

Je remercie tous les membres du projet Open-ViBE. Vous avez tous largement contribué à ma formation dans le domaine des interfaces cerveau-machines et j'ai toujours pris beaucoup de plaisir à participer aux différentes réunions de travail dans une ambiance à la fois studieuse et chaleureuse. Merci donc à Anatole Lécuyer, Yann Renard, Antoine Souloumiac, Denis Chêne, Bernard Hennion, Karim Jerbi, Ornella Plos, Vincent Delannoy, Virginie Attina, Guillaume Gibert, Emmanuel Maby, Claude Dumas, Sophie Heinrich, Nathan Weisz, Aurélien Van Langenhove (j'en oublie sûrement). J'adresse un merci spécial

à Antoine Souloumiac pour ces lumières au sujet de la diagonalisation conjointe approchée et à Bernard Hennion pour les innombrables excellents moments que j'ai passés en sa compagnie et tout ce que j'ai pu apprendre à son contact.

Merci à tous mes compagnons de bureau. Vous m'avez aidé à réviser certaines des règles du français, mais surtout votre insatiable bonne humeur et votre gentillesse ont rendu ma thèse très agréable. Un grand merci donc à Ladan et Leonardo (ainsi qu'à nouveau Reza et Jérôme).

Merci à tous mes collègues du département Steeve, Jérôme, Olivier, Florent et les membres du groupe sigmaphy, Barthélémy, Nicolas, Brice, Rosilde, Sophie M., et spécialement pour les accueils toujours chaleureux que j'ai reçus dans le bureau des espagnols Jordi, Ion et Pepa. Je n'oublie pas bien sûr que cette thèse a aussi été rendue possible grâce aux assistances informatiques de Jean-Marc et Hervé qui ont toujours répondu à mes requêtes en des temps records. Je tiens à remercier tout spécialement Lucia pour son aide et son insatiable bonne humeur et Marie-No pour sa gentillesse et pour toutes les fois où elle m'a simplifié les démarches administratives. Je voudrais enfin saluer Joël Gardes et Alexandre Vienne qui ont été pour moi des sources d'inspiration toujours efficaces.

J'ai bien sûr une pensée pour mes parents et mes sœurs qui m'ont soutenu et supporté pendant toutes ces années. Ce travail n'aurait pas été le même sans leur aide, merci aussi pour leur aide lors de la relecture et la préparation du pot de thèse. Merci également à tous mes amis qui ont participé à la relecture de ce document, Elise, Sandra et Caroline. Merci à Danièle pour son aide pour la préparation du pot de thèse et son soutien tout le reste du temps.

Je veux remercier par dessus tout ma chère et tendre Julie, pour sa patience au cours de ces deux dernières années, et notamment lorsque j'ai quelquefois consacré un peu trop de temps à l'écriture de ma thèse.

Je conclurai ces remerciements par un hommage à Line Garnero qui avait initialement accepté de rapporter mon travail de thèse mais nous a quitté prématurément. Outre les différents travaux que cette grande scientifique a réalisés dans le domaine général des neurosciences, je voudrais rappeler tout particulièrement que Line Garnero a été l'une des premières personnes en France à s'intéresser aux interfaces cerveau-machines, encadrant notamment la thèse de Michel Besserve qui fût l'une des premières thèses françaises dans le domaine des interfaces cerveau-machines.

Table des matières

TABLE DES MATIÈRES	IV
TABLE DES FIGURES	IX
Liste des tableaux	XV
Liste des algorithmes	XVII
ABRÉVIATIONS	XIX
I INTRODUCTION	1
1 Contexte	2
2 Applications « défense »	3
3 But de la thèse	4
4 Approche générale	5
5 Organisation du manuscrit	5
6 Références du chapitre	7
II INTERFACES CERVEAU-MACHINE : SITUATION ACTUELLE	9
1 Introduction	10
2 Mesurer l'activité cérébrale	10
2.1 Techniques d'imagerie cérébrale	11
2.1.1 Bases de neurophysiologie	11
2.1.2 ICM et réponses hémodynamique ou métabolique	13
2.1.3 ICMs et réponse électromagnétique	14
2.2 Résumé	18
3 Indices électrophysiologiques et interfaces cerveau-machine	18
3.1 P300	19
3.2 Potentiels évoqués stationnaires	20
3.2.1 Visuels	20
3.2.2 Somatosensoriels	21
3.2.3 Auditifs	21
3.3 Modulation des rythmes cérébraux	21
3.3.1 Rythmes sensorimoteurs	23
3.3.2 Rythmes « cognitifs »	24
3.4 Potentiels corticaux lents	24
4 Différents types d'interfaces cerveau-machine	24
4.1 Point de vue du sujet	25
4.1.1 Les tâches mentales	25
4.1.2 « Interrupteurs utilisateurs »	25
4.2 Point de vue du système	26
4.2.1 Les commandes	26
4.2.2 « Interrupteurs systèmes »	26
4.2.3 Stimulations et retour sensoriel	27
4.3 Quelques exemples	28
4.3.1 « Thought Translation Device »	28
4.3.2 « P300 Speller »	28

	4.3.3	Contrôle d'un curseur et rythmes sensorimoteurs	29
	4.3.4	Imagination de mouvement (Graz)	29
	4.3.5	Résumé	29
	4.4	Définitions	29
5		Méthodologie	31
	5.1	Caractériser une activité cérébrale	31
	5.1.1	Prétraitements	32
	5.1.2	Domaine de représentation des caractéristiques	37
	5.1.3	Utilisation des successions d'activations	38
	5.2	Apprentissage et classification	39
	5.2.1	Présentation générale	39
	5.2.2	Analyse discriminante linéaire	42
	5.2.3	Régression logistique	42
	5.2.4	« Support Vector Machine »	43
	5.2.5	Évaluation et sélection de modèles	45
6		Enjeux actuels et problématiques de la thèse	46
7		Conclusion	47
8		Références du chapitre	47
III		FILTRAGE SPATIAL : EXTRACTION DE SOURCES NON-STATIONNAIRES	55
	1	Introduction	56
	2	Non-stationnarité et imagerie motrice	58
	3	Séparation de sources non-stationnaires	63
	3.1	Formulation théorique	63
	3.2	« Common Spatial Pattern »	64
	3.2.1	Deux classes	64
	3.2.2	Extension à plus de deux classes	65
	3.3	Utilisation de la diversité entre classes : DCA	65
	3.4	Utilisation des variations d'énergie dans une tâche : DCA-MS	66
	4	Paradigme expérimental	67
	5	Méthode	68
	5.1	Phase d'entraînement	68
	5.2	Phase de test	69
	5.3	Validation croisée	69
	5.4	Transfert inter-session	69
	6	Résultats	70
	6.1	Validation croisée	70
	6.1.1	Analyse de variance dans le cas $T_s = 60$	70
	6.1.2	Résultats détaillés pour $T_s = 60$	71
	6.1.3	Influence de T_s	71
	6.2	Transfert session à session	72
	6.2.1	Analyse de variance	73
	6.2.2	Résultats détaillés	73
	6.3	Comparaison avec la « BCI Competition 2008 »	74
	6.4	Artefact numérique ou information réelle	75
	7	Utilisation du domaine fréquentiel	75
	8	Discussion	77
	9	Conclusion	79
	10	Références du chapitre	79
IV		ELIMINATION DES ARTEFACTS OCULAIRES	83
	1	Introduction	84

2	Méthode	86
2.1	Détection des périodes de contamination	86
2.1.1	Méthode semi-automatique	86
2.1.2	Information sur $\Phi_{\mathcal{H}_0}(t)$	86
2.1.3	Cas général	87
2.1.4	Résumé	88
2.2	Transformation linéaire	88
2.3	Séparation entre le signal et le bruit	89
2.3.1	Mise à zéro	89
2.3.2	Débruitage par ondelettes	89
2.3.3	Retour à l'espace des capteurs	90
2.4	Extension itérative	90
3	Évaluation à partir de données simulées	91
3.1	Génération des données	91
3.2	Méthode d'évaluation	92
3.3	Résultats	93
4	Données réelles	93
4.1	Détection des périodes d'activité oculaire	94
4.2	Projection linéaire	95
4.3	Signaux débruités	96
5	Application : classification de signaux non-bruités	97
5.1	Données et méthode	97
5.2	Résultats	99
6	Discussion et conclusion	100
6.1	De l'importance de débruiter les données	100
6.2	Implémentation de la méthode	100
6.3	Conclusion	101
7	Références du chapitre	102
V	FILTRAGE DES ÉTATS MENTAUX PAR MODÈLE DE MARKOV	105
1	Introduction	106
2	Formulation théorique	107
2.1	Modèle d'évolution des états mentaux, cadre variationnel bayésien	107
2.1.1	Le cadre variationnel bayésien	109
2.1.2	Reformulation variationnelle du problème	110
2.1.3	Filtrage variationnel	112
2.2	Résumé de l'algorithme de filtrage des probabilités	114
2.3	Illustration du filtrage variationnel bayésien	115
3	Méthode générale pour les ICMs	117
3.1	Phase d'entraînement	117
3.2	Phase d'évaluation	117
3.3	Résumé	117
4	Apprentissage de l'état « repos »	118
4.1	Utilisation des périodes de repos	118
4.2	Définition implicite	119
5	Données expérimentales	120
5.1	« BCI Competition III, dataset 4b »	120
5.2	« BCI Competition IV, dataset 1 »	121
6	Résultats	121
6.1	Méthode d'évaluation	121
6.1.1	Erreur quadratique moyenne	121
6.1.2	Matrice de confusion	122

6.1.3	Caractéristique de fonctionnement du récepteur	123
6.1.4	Proportion de points mal classés	123
6.2	« BCI Competition III »	123
6.2.1	Choix des paramètres par validation croisée	123
6.2.2	Apprentissage de l'état de repos	123
6.2.3	Utilisation du modèle de Markov	125
6.3	« BCI Competition IV »	127
6.3.1	Choix des paramètres	127
6.3.2	Apprentissage de l'état de repos	127
6.3.3	Utilisation du filtrage variationnel	128
7	Discussion et conclusion	129
7.1	Utilisation des bandes de fréquences	130
7.2	Améliorations dues au filtrage variationnel	130
8	Références du chapitre	132
VI	VERS LA COADAPTATION CERVEAU-MACHINES	135
1	Introduction	135
2	Quantifier l'apprentissage du sujet	136
2.1	Quantifier la réalisation de chaque essai	137
2.2	Convergence de l'apprentissage des filtres spatiaux	140
2.3	Application aux interfaces asynchrones	141
3	Classification continue adaptative	143
3.1	Modéliser les probabilités en sortie du classifieur	143
3.2	Application aux interfaces asynchrones	145
4	Discussion et conclusion	146
5	Références du chapitre	148
VII	CONCLUSION ET PERSPECTIVES	151
1	Filtrage spatial pour les ICMs	151
2	Utilisation de la dynamique des états mentaux	152
3	Vers des modèles dynamiques plus précis	153
4	Conclusion générale	154
A	CLASSIFICATION EN GRANDE DIMENSION	155
1	Fonction de coût pour l'apprentissage	155
2	Illustration du fonctionnement	156
3	Quantité de données d'apprentissage et régularisation	156
4	Conclusion	157
5	Références du chapitre	158
B	LA DIAGONALISATION JOINTE EN PRATIQUE	159
C	NOTE SUR R ET LA PARALLÉLISATION DES TÂCHES (À MOINDRE COÛT)	161
1	Utilisation d'une bibliothèque de calcul algébrique optimisée	161
2	Utilisation de tous les cœurs d'une même machine	162
3	Parallélisation multi-machines et multi-processeurs	165
4	Conclusion	166
	PUBLICATION DE L'AUTEUR	167

Table des figures

I.1	Schéma de principe d'une interface cerveau-machine dans lequel sont représentés les différents blocs de traitement nécessaires à la conception d'un ...	5
II.1	Représentation schématique d'une interface cerveau-machine. Le sujet accomplit une certaine tâche mentale. L'activité induite par cette tâche ...	10
II.2	a) Une représentation des différents lobes du cortex. b) Cette figure montre la densité relative des vaisseaux sanguins dans le cortex du primate. Le ...	12
II.3	Représentation schématique des principaux générateurs de champs électromagnétiques ainsi que les principaux systèmes d'imagerie permettant de ...	14
II.4	Le système 10-20 décrivant les positions standards des électrodes EEG, d'après Zani et Proverbio (2002).	17
II.5	Représentation schématique des neurones pyramidaux, principaux générateurs de l'EEG de scalp. Ils reçoivent les projections excitatrices au niveau ...	17
II.6	D'après Zani et Proverbio (2002). Génération des potentiels évoqués stationnaires en fonction de la fréquence du stimulus visuel. En haut à gauche, ...	21
II.7	Unité de temps en abscisse : seconde, en ordonnée : unité arbitraire. Electroencéphalogramme humain lors de différents états de conscience. (a) état ...	23
II.8	Densité spectrale de puissance typique d'un EEG. Le spectre est estimé par la méthode du périodogramme et a été lissé à l'aide d'un noyau de Daniel.	23
II.9	Communication cerveau-machines. Ce schéma détaille le fonctionnement d'une interface cerveau-machine en séparant le rôle du sujet et du ...	25
II.10	Critères permettant de classer les différents types d'interfaces cerveau-machines en fonction des commandes du système. Dans chacun des quatre ...	27
II.11	Quatre différentes configurations des relations qui lient les variables propres au sujet ($u_1(t)$ indique les moments d'envoi d'une commande par le ...	30
II.12	Exemple simulé à partir de deux lois normales, chacune en dimension deux. A gauche, les tirages (500 points) représentant les valeurs mesurées au ...	35

II.13	a) Exemple de classification en dimension deux. Les points correctement classés sont dessinés en noir et les points mal classés sont en rouge. Les ...	41
II.14	Frontières obtenues pour le même problème que précédemment avec les méthodes de l'analyse discriminante linéaire et l'analyse discriminante ...	43
II.15	Erreur de prédiction sur les données d'entraînement et de test en fonction de la complexité du modèle utilisé.	46
III.1	Homunculus de Penfield. La partie motrice, à droite, montre l'organisation de la commande motrice des muscles du corps humain. La partie ...	59
III.2	Quelques profils possibles des variances pour des fréquences différentes et des sources distinctes. A partir de chaque moment indiqué sur l'axe des ...	61
III.3	a) Coefficients des filtres (gauche) correspondant à la première source de la figure précédente ainsi que distribution spatiale de l'activité associée ...	62
III.4	a) Schéma représentant le déroulement temporel de la réalisation d'une tâche d'imagerie motrice. b) Montage EEG des 22 capteurs. La distance entre ...	67
III.5	Schéma bloc résumant les différentes phases de traitement dans les étapes d'entraînement et de test.	70
III.6	Figure montrant l'influence du nombre d'essais dans le jeu de données d'entraînement. Seuls les sujets 3 et 7, session 2 sont présentés ici.	72
III.7	Résumé des performances obtenues par les trois méthodes considérées pour chacun des sujets dans les cas $2 \rightarrow 1$ (Sess. 1) et $1 \rightarrow 2$ (Sess. 2).	73
III.8	Schéma bloc résumant les différentes phases de traitement dans les étapes d'entraînement et de test dans le cas de l'utilisation de l'information ...	76
III.9	Analyse de l'utilisation des bandes de fréquences par le classifieur. Pour chacune des courbes représentant l'utilisation des bandes pour classifier ...	77
IV.1	Exemple d'enregistrement EEG comportant 21 capteurs. Les signaux sont contaminés par des artefacts oculaires (le sujet devait bouger les yeux). ...	85

IV.2	Distributions de probabilités de $\Phi(t)$ dans deux conditions distinctes. Sur les deux graphes, on a également indiqué le seuil ℓ qui permet d'obtenir ...	87
IV.3	Représentation schématique résumant la procédure décrite. Les données EEG sont d'abord projetées, puis débruitées et enfin l'inverse de la ...	91
IV.4	Performances de la méthode présentée dans ce chapitre par rapport aux performances d'une méthode basée sur l'analyse en composantes ...	94
IV.5	Illustration des trois phases différentes au début de la session. En haut, le sujet avait les yeux ouverts, au milieu, il gardait les yeux fermés et en ...	95
IV.6	La figure de gauche est un graphe quantile-quantile qui permet de vérifier la normalité de la distribution de $\varphi(t)$ lorsque le sujet n'a pas d'activité ...	96
IV.7	Illustration montrant l'opération de seuillage de l'énergie. Sur fond bleu sont représentées les parties du signal qui sont considérées comme de ...	96
IV.8	Trois premières et trois dernières composantes extraites par décomposition en valeurs propres généralisée. Attention : on a utilisé des échelles ...	97
IV.9	Sur chacune des quatre figures sont représentés les signaux originaux (haut), signaux débruités (milieu) ainsi que les résidus (bas). a et c) ...	98
IV.10	Résumé des performances obtenues par les deux méthodes considérées pour chacun des sujets dans les cas $2 \rightarrow 1$ (Sess. 1) et $1 \rightarrow 2$ (Sess. 2). DCA ...	99
IV.11	Analyse de l'utilisation des bandes de fréquences par le classifieur. Pour chacune des courbes représentant l'utilisation des bandes pour classifier ...	100
V.1	Modèle graphique décrivant les interactions entre les différentes variables du modèle. l_t représente une variable latente du système et peut être ...	108
V.2	Densités marginales d'une distribution de Dirichlet en dimension 2 pour différentes valeurs de κ .	112
V.3	Illustration du fonctionnement de l'algorithme à partir de probabilités issues de données réelles. De l'expérience qui comportait deux tâches, nous ...	116

V.4	Schéma bloc résumant les différentes phases de traitement dans les étapes d'entraînement et de test en incluant le filtrage des probabilités par la ...	119
V.5	Utilisation des périodes d'inactivité afin de définir l'état de repos.	119
V.6	Utilisation de la matrice de confusion pour évaluer les performances d'un système de classification. A gauche, les proportions de vrais négatifs (Vrai ...	122
V.7	En haut : en rouge la courbe ROC correspondant à la méthode de l'apprentissage implicite du repos dans le cas où le paramètre $\mathbb{P}(\mathcal{I})$ varie entre ...	125
V.8	En haut : en bleu, le point correspondant à la méthode du repos appris ; en orange la même séquence probabilités a été filtrée en utilisant le filtrage ...	126
V.9	Courbes ROC pour les quatre derniers sujets. L'apprentissage de l'état de repos est implicite. Les précisions ne sont pas indiquées sur ces figures ...	129
V.10	Coefficients non nuls de la régression logistique. Les coefficients (rectangles gris dont la valeur est codée selon la valeur absolue du coefficient) sont ...	131
V.11	Probabilités ainsi que les décisions associées en sortie du classifieur (haut) ainsi qu'après filtrage variationnel. La discrimination implique seulement ...	132
VI.1	Scores d'apprentissage pour l'ensemble des essais de la session d'entraînement relatifs à quatre sujets (parmi sept). Nous rappelons que les sujets a, ...	139
VI.2	Étude de la convergence de l'apprentissage des filtres lors de la session d'entraînement en fonction du nombre d'essais de chacune des tâches ...	141
VI.3	Influence de l'élimination des mauvais essais sur les performances. Seuls les résultats des sujets f et g sont présentés ici. Ces résultats sont présentés ...	142
VI.4	Illustration de l'estimation de la moyenne et de la variance d'un processus Gaussien comportant un saut à $t = 5000$. Le facteur d'oubli est ...	145
VI.5	En haut à droite, courbes ROC correspondant aux quatre déclinaisons proposées. Les courbes sont obtenues en faisant varier entre zéro et un la ...	147

VI.6	Classification continue pour chacune des quatre déclinaisons pendant une période d'une durée de 200 s. Les tâches théoriquement réalisées par le ...	148
A.1	Influence de la valeur du paramètre de régularisation sur l'entraînement du classifieur pour les quatre sujets réels de la quatrième compétition BCI. ...	157
A.2	Taux de classification par rapport au maximum obtenu pour chacun des couples de valeurs (taille des données d'apprentissage, valeur de la ...	158
B.1	Convergence de plusieurs algorithmes de diagonalisation conjointe fournis dans la bibliothèque jointDiag. On a effectué 50 diagonalisations ...	160
C.1	Temps de calcul pour une multiplication matricielle en fonction de la taille des matrices. Les trois configurations comparées sont : le logiciel R avec ...	163

Liste des tableaux

II.1	Différents systèmes d'imagerie, évalués en fonction de critères définis pour les Interfaces Cerveau-Machines. Les lignes dont le fond est ...	18
III.1	Résumé de l'analyse de variance pour les performances en validation croisée. L'influence de chacun des facteurs de la première colonne est évaluée ...	71
III.2	Performances détaillées obtenues par la meilleure méthode (troisième colonne) pour chacun des sujets et chacune des sessions.	72
III.3	Résumé de l'analyse de variance pour les performances en transfert session à session. L'influence de chacun des facteurs de la première colonne est ...	73
III.4	Mesures Kappa obtenues par les trois meilleurs compétiteurs ainsi que par les trois méthodes évaluées ici (CSP, DCA et DCA-MS). En est ...	75
V.1	Erreurs quadratiques moyennes pour chacune des deux méthodes d'apprentissage de l'état de repos.	124
V.2	Valeurs des différents paramètres choisis arbitrairement (nombre de filtres spatiaux) ou fixés par validation croisée (meilleure fenêtre et paramètre ...	127
V.3	Erreurs quadratiques obtenues pour chaque méthode d'apprentissage de l'état de repos et chaque sujet. Ce tableau contient également les résultats ...	128
V.4	Proportions de points mal classés dans le cas où le seuillage est basé sur les probabilités brutes en sortie du classifieur ou bien sur les probabilités ...	129
C.1	Performances en termes de temps de calcul et de mémoire RAM utilisée par deux versions différentes d'un même traitement. La version 1 cœur ...	165
C.2	Performances en termes de temps de calcul et de mémoire RAM utilisée par deux versions différentes d'un même traitement. La version 1 cœur ...	166

Liste des algorithmes

V.1	Filtrage variationnel bayésien	114
V.2	Apprentissage des paramètres lors de la phase d'entraînement. Lors de cette phase, le paradigme est synchrone, nous disposons donc des indications	118

Notations mathématiques

Fonctions et opérateurs

$\det(\cdot)$	Déterminant d'une matrice
$(\cdot) * (\cdot)$	Produit de convolution
$(\cdot) \circ (\cdot)$	Produit de Hadamard
$\langle \cdot, \cdot \rangle$	Produit scalaire
$ \cdot $	Cardinal d'un ensemble
$ \cdot $	Module d'un scalaire
$(\cdot)'$	Dérivée
$(\cdot)^T$	Transposition
$(\cdot)^*$	Conjugaison
$\hat{(\cdot)}$	Estimée
$\Gamma(\cdot)$	Fonction gamma
$\psi(\cdot)$	Fonction digamma
$\mathcal{N}(\cdot, \cdot)$	Loi normale
$\mathcal{D}i(\cdot)$	Loi de Dirichlet
$\mathcal{M}u(\cdot)$	Loi multinomiale
$\mathcal{G}(\cdot, \cdot)$	Loi gamma
$\mathbb{E}\{\cdot\}$	Espérance mathématique
$\mathbb{P}(\cdot)$	Probabilité d'un événement

Ensembles et espaces

\mathbb{N}	Ensemble des entiers naturels
\mathbb{Z}	Ensemble des entiers relatifs
\mathbb{R}	Ensemble des réels
\mathbb{C}	Ensemble des complexes
$L^2(\mathbb{R})$	Ensemble des fonctions de \mathbb{R} dans \mathbb{R} d'énergie finie
$\chi_{\mathbb{F}}(\cdot)$	Fonction indicatrice dans \mathbb{F}

Scalars, vecteurs et matrices

p_t	vecteur de probabilités estimée à l'instant t
$p_{i,t}$	$i^{\text{ème}}$ composante de p_t estimée à l'instant t
$T_{i,t}$	$i^{\text{ème}}$ colonne de la matrice T_t
$t_{i,j,t}$	Élément de la $i^{\text{ème}}$ ligne et $j^{\text{ème}}$ colonne de T_t
$\mathbf{y}(t)$	Vecteur colonne de signaux temporels
$y_i(t)$	$i^{\text{ème}}$ composante du vecteur $\mathbf{y}(t)$
A	Matrice

Abréviations

ACI	Analyse en Composantes Indépendantes
AFM	Association Française contre les Myopathies
AR	autorégressifs
BP	Bereitschaft Potentials
CSP	« Common Spatial Pattern »
DCA	Diagonalisation Conjointe Approchée
DGA	Délégation Générale pour l'Armement
ECoG	électrocorticographie
EEG	électroencéphalographie
EOG	électrooculographie
ERD	Event-Related Desynchronization
ERP	Event-Related Potentials
ERS	Event-Related Synchronization
ICM	Interface Cerveau-Machine
iid	indépendantes et identiquement distribuées
INRIA	Institut National de Recherche en Informatique et en Automatique
INSERM	Institut National de la Santé et de la Recherche Médicale
IRMf	Imagerie par Résonance Magnétique fonctionnelle
LENA	Laboratoire Neurosciences Cognitives et Imagerie Cérébrale, CNRS, Paris, France
MEG	magnétoencéphalographie
PE	Potentiels Évoqués
RKHS	reproducing kernel Hilbert Space – espace de Hilbert à noyau auto-reproduisant
RSB	Rapport Signal sur Bruit
SAS	Séparation Aveugle de Sources
SLA	Sclérose Latérale Amyotrophique
SPIR	Spectroscopie Proche de l'Infra-Rouge
SVM	Support Vector Machine
TTD	Thought Translation Device

Chapitre I

INTRODUCTION



If the brain were simple enough for us to understand it, we would be too simple to understand it.

— Ken Hill

TOUT COMMENÇA par un rêve. Quand Hans Berger, alors spécialiste de l'étude vasculaire cérébrale, reçut une lettre de sa sœur lui faisant part d'un rêve qu'elle avait fait dans lequel elle le voyait se casser la jambe dans une chute de cheval, il n'imaginait alors pas toutes les conséquences qu'aurait cette lettre sur l'étude du cerveau humain. Mais le rêve devint prémonitoire et celui-ci se cassa effectivement la jambe, le rendant par là même convaincu de la réalité de possibilités de télépathie¹ entre les êtres humains. Il se mit alors à chercher des signaux physiques émanant du cerveau qui pourraient être transportés au-delà des limites du crâne humain. Il eut l'idée d'utiliser un galvanomètre² pour mesurer des différences de potentiels à la surface du crâne. Ayant éliminé la possibilité que les différences observées puissent provenir des variations de pression sanguine ou encore de la peau du scalp, il devint convaincu de la réalité de variations électriques dues à l'activité cérébrale mesurable à la surface du crâne. Après cinq ans d'expérimentations, il publia le premier article ([Berger, 1929](#)) témoignant de variations du potentiel électrique mesurées à la surface du crâne. Il mesura la période de ces oscillations et nomma ce rythme α (alpha) car c'était le premier qu'il observait. Le deuxième rythme observé par Dr. Hans Berger fut ensuite appelé β (beta) et était environ deux fois plus rapide. L'analyse de l'activité cérébrale par électroencéphalographie (EEG) était née.

Si l'idée d'utiliser l'activité cérébrale pour contrôler un système électronique ou communiquer avec celui-ci a nourri plusieurs générations d'écrivains de science-fiction, il faut attendre 1960 et l'introduction du terme « cyborg »³ dans un article scientifique ([Clynes et Kline, 1960](#)) pour que soit formalisée cette idée. C'est ensuite en 1973 qu'ont lieu les premières expérimentations réelles visant à utiliser l'activité cérébrale mesurée par EEG pour contrôler un système électronique ([Vidal, 1973](#) et [Vidal, 1977](#)).

¹ Le terme est employé ici dans le sens d'une communication entre deux cerveaux humains par des voies autres que les voies de communication sensorielles classiques.

² Appareil de mesure de l'intensité d'un courant électrique faible. Le galvanomètre doit son nom à Luigi Galvani (qui en a eu l'idée) mais c'est en 1820 que Johann Schweigger a été le premier à construire un tel appareil.

³ Le terme « cyborg » fut employé pour décrire l'interaction entre composants artificiels et biologiques au sein d'un même organisme.

1 Contexte

Depuis ces travaux pionniers, les concepts sous-jacents à la communication entre le cerveau et les machines se sont développés et formalisés. On appelle aujourd'hui Interface Cerveau-Machine (ICM) l'ensemble des systèmes visant à établir un moyen de communication direct, *i.e.* sans aucune utilisation des voies motrices périphériques, entre le cerveau humain et un système électronique quelconque. Une telle définition englobe par exemple les implants cochléaires qui constituent la première et la plus large entreprise afin de rendre une ICM accessible au grand public. Le principe d'un implant cochléaire est de convertir l'ensemble de l'environnement sonore du patient en un signal électrique destiné au nerf auditif. En 2001, près de 30000 patients dans le monde entre 12 mois et 80 ans bénéficiaient d'un tel système (Nicolelis, 2001), avec des résultats très hétérogènes entre les sujets. Mais c'est depuis le début des années 1990 que la recherche sur les ICMs a pris tout son essor. Les chercheurs y ont vu le moyen idéal pour restaurer chez certains types de patients des capacités de communication et de contrôle. Certains patients atteints de maladies motrices graves n'ont en effet plus aucune possibilité de communication ou de mouvement alors que leur cerveau est en parfait état de fonctionnement. La Sclérose Latérale Amyotrophique (SLA) est un exemple d'une telle maladie. C'est une maladie neuro-dégénérative qui, au fur et à mesure de sa progression, provoque la mort des neurones moteurs de l'organisme, rendant impossible toute activité nécessitant l'usage d'un muscle. D'autres maladies neuro-dégénératives ainsi que les accidents vasculaires cérébraux peuvent conduire les patients vers le même état d'enfermement. En convertissant l'activité cérébrale volontaire de ces patients en commandes permettant d'agir sur des systèmes électroniques, les ICMs représentent pour eux un espoir prometteur afin de leur fournir un moyen de communication (Poulain et al., 2007 et Berger et al., 2007).

Parallèlement à ce but médical, l'engouement pour les ICMs se développe actuellement dans le grand public. Un rapport commandé par Microsoft Corp.[®] (Harper et al., 2008) auprès d'une quarantaine de chercheurs dans le monde pour imaginer les interactions homme-machines dans les années 2020 prédit une place de plus en plus importante aux ICMs, leur développement accompagnant un mouvement à long terme tendant à rendre toujours de plus en plus floue la limite entre humains et ordinateurs. Les compagnies de jeux vidéos y voient une nouvelle possibilité d'interaction révolutionnaire en ajoutant ainsi un degré de liberté supplémentaire à leurs produits¹. L'un des exemples le plus caractéristique est la société Emotiv[®] fondée en 2004, dont le but est de créer un casque EEG pour les joueurs. Cependant les différentes démonstrations qui ont eu lieu n'ont pour l'instant pas convaincu les journalistes et spécialistes du domaine (Heingartner, 2009). Afin de mesurer la difficulté de la tâche, il suffit de citer le temps de réaction moyen d'un joueur pour appuyer sur un bouton (souris ou clavier par exemple) : 190 ms. Au moment de la rédaction de cette thèse on peut aussi noter la sortie d'un nouveau système d'acquisition de données électroencéphalographiques, électrocardiographiques et électrooculographiques mis au point par la société Starlab[®] et destiné à des applications simples et ludiques².

Le système actuellement le plus utilisé pour mesurer l'activité cérébrale des sujets est l'EEG. L'une des raisons de cette situation est le faible coût de ce type de systèmes en comparaison des autres systèmes d'imagerie. Mais ce système d'enregistrement de l'activité cérébrale est loin d'être parfait. L'analyse des signaux recueillis et leur transcription en une commande n'est pas une tâche aisée et nécessite la collaboration entre des équipes pluridisciplinaires. Parmi les principales compétences nécessaires à la mise en place d'une ICM, on peut citer les suivantes :

¹ Rappelons que l'industrie du jeu vidéo a engendré 18 milliards de dollars au cours de l'année 2007 sur le continent américain. Elle est aujourd'hui considérée comme plus importante que l'industrie du cinéma.

² Pour plus d'informations, voir le [Site Internet officiel de la compagnie Starlab basée en Espagne à Barcelone](#)

- ▷ **Neurophysiologie** : Utiliser l'activité générée par le cerveau pour la transcrire ensuite en une commande nécessite de connaître les caractéristiques de ce qu'on cherche à mesurer, *e.g.* à quel endroit du cerveau est-on susceptible de détecter l'activité cible ;
- ▷ **Technologie d'imagerie cérébrale** : concevoir une interface cerveau-machine nécessite d'acquérir des signaux représentatifs de l'activité du cerveau. Cette partie implique donc l'utilisation de systèmes de mesure de l'activité cérébrale adaptés ;
- ▷ **Ergonomie** : ce point souvent négligé est en fait crucial dans le cas de systèmes destinés à un public large. Ce domaine de compétence s'intéresse par exemple au type de retour que l'on peut renvoyer à l'utilisateur pour faciliter l'apprentissage du contrôle du système ;
- ▷ **Traitement du signal** : les signaux EEG contiennent les signaux relatifs à la tâche mentale que l'on cherche à reconnaître, mais pas uniquement... ils sont aussi sensibles aux perturbations extérieures, à l'activité de fond du cerveau, aux activités musculaires et aux mouvements des yeux. Pour toutes ces raisons, des compétences en traitement du signal sont nécessaires pour faire ressortir, à partir des enregistrements, les signaux spécifiques à la tâche que l'on cherche à reconnaître ;
- ▷ **Apprentissage** : lorsque les signaux sont nettoyés des contaminations indésirables, il est alors nécessaire d'entraîner le système à reconnaître les tâches mentales « commandes » parmi l'ensemble de l'activité cérébrale du sujet. Dans ce but, les ICMs font un usage intensif des techniques d'apprentissage statistique.

Notons enfin que la recherche concernant les ICMs basées sur des techniques d'imagerie non-invasives¹ en France est relativement récente. C'est au sein du laboratoire parisien le LENA (Laboratoire Neurosciences Cognitives et Imagerie Cérébrale, CNRS, Paris, France) qu'a été soutenue la première thèse française sur les ICMs (Besserve, 2007)². Les travaux de thèse présentés dans ce manuscrit s'inscrivent dans le cadre du premier projet français financé par l'ANR sur le thème des ICMs. Nous avons ainsi eu la chance de participer aux prémices du projet et de partager les engouements suscités par les différentes réunions trimestrielles, autant d'occasions d'échanger autour d'un même thème de recherche au sein d'une assemblée largement pluridisciplinaire. Open-ViBE (Arrouet et al., 2005) regroupait initialement les experts en informatique de l'INRIA (Institut National de Recherche en Informatique et en Automatique, Rennes), les experts en neurophysiologie de l'INSERM (Institut National de la Santé et de la Recherche Médicale, Lyon), les ergonomes de l'AFM (Association Française contre les Myopathies, Paris) et certaines équipes de recherche d'Orange[®]. Le consortium initial a rapidement été élargi pour englober le laboratoire GIPSA-lab (Grenoble Images Parole Signal Automatique) ainsi que le CEA (Commissariat à l'Énergie Atomique, Saclay).

2 Applications « défense »

Cette thèse a été financée par la DGA (Délégation Générale pour l'Armement). Ceci m'a valu, dans la plupart des conférences auxquelles j'ai assisté, la question suivante : mais qu'est-ce qui peut bien intéresser la DGA dans des travaux de thèse sur les ICMs ? Bien

¹ On parle d'une technique non-invasive lorsqu'aucune intervention chirurgicale n'est nécessaire pour mettre en place le système de mesure de l'activité cérébrale.

² Cette thèse a été co-encadrée par Line Garnero et Jacques Martinerie.

que la réponse ne soit connue que de la DGA, j'ai tenté dans ce paragraphe d'éclaircir ce point et de situer les recherches sur les ICMs dans le contexte de la défense nationale¹.

Au moment de rédaction de cette thèse, on pouvait ainsi lire dans un journal américain² une brève concernant un projet de financement par la division Darpa du Pentagone sur le thème suivant : tout d'abord, tenter à l'aide de capteurs EEG d'associer à chaque mot d'une personne une activité cérébrale caractéristique ; ensuite déterminer si ces activités sont consistantes d'un sujet à un autre ; et enfin, si c'est le cas, construire un prototype capable de détecter de telles activités et les transmettre aux personnes se situant dans un rayon d'action proche (en l'occurrence les alliés pendant un combat). Ce projet est bien sûr loin d'être à un stade suffisamment avancé pour déterminer de son succès mais il montre bien l'attitude des responsables de la défense vis-à-vis des recherches sur le cerveau. En suivant les liens de l'article on peut alors avoir accès à un rapport (Williams, 2008) non-classé fourni au département de la défense américain par l'association JASON³ et qui traite en partie de l'intérêt des ICMs pour les forces militaires. Ce groupe d'experts s'est intéressé à la maturité des technologies existantes et en a déduit des recommandations. Nous traduisons ici un extrait du résumé qui nous semble assez caractéristique de l'état des ICMs dans le contexte militaire aussi bien que civil : « Les applications (bonnes ou mauvaises) des ICMs visant à la modification des performances humaines stimulent les imaginations. Mais les capacités réelles actuelles de ces interfaces dans le domaine médical restent loin des espoirs qu'on leur attribue. Même si de véritables améliorations dans la qualité de vie de certains patients ont pu être observées, [...], les performances atteintes restent loin derrière les performances usuelles. Ceci est dû à la fois au stade précoce de développement de ces technologies et également à notre connaissance trop incomplète du fonctionnement du système nerveux central. » Deux scénarii sont ensuite proposés dans le rapport. Dans le premier, les auteurs voient les ICMs comme le moyen de contrôler ou diriger un équipement extérieur. Dans le deuxième, les ICMs sont envisagées pour améliorer les retours sensoriels venant de l'extérieur ou augmenter les performances d'un combattant. En conclusion, le rapport recommande de suivre avec attention les avancées du domaine compte tenu des applications potentielles mais rejettent l'idée d'une technologie suffisamment mature pour être utilisée actuellement.

3 But de la thèse

Les travaux présentés dans cette thèse se sont concentrés sur l'aspect « traitement du signal » relatif à la conception d'une interface cerveau-machine. Ils s'inscrivent dans l'ensemble des travaux visant à fournir des méthodes efficaces pour rendre les ICMs utilisables dans la vie de tous les jours. La plupart des systèmes ICMs impose à l'utilisateur le moment auquel la tâche mentale qui permet d'envoyer une commande doit être réalisée. Ce mode de fonctionnement facilite grandement l'analyse des signaux EEG recueillis car le système sait à quel moment il doit chercher la tâche mentale « cible », mais il s'avère aussi lent et très fastidieux à l'usage. C'est pourquoi une partie importante des efforts de recherche actuellement consacrés au domaine se concentre sur le développement de systèmes « asynchrones »⁴ : ce n'est pas l'ordinateur qui impose au sujet à quel moment il doit effectuer sa tâche mentale mais plutôt le sujet qui doit avoir la possibilité d'effectuer sa tâche mentale lorsque qu'il le désire. Ceci impose que le système soit capable d'analyser en temps réel l'activité cérébrale du sujet et de reconnaître la tâche mentale recherchée

¹ Ces propos n'engagent que l'auteur de cette thèse.

² <http://www.wired.com/dangerroom/2009/05/pentagon-preps-soldier-telepathy-push/>

³ JASON est un groupe de scientifiques indépendant qui conseille le gouvernement américain en matière de science et de technologie, il a été créé en 1960 et aurait entre 30 et 60 membres (source : <http://en.wikipedia.org>).

⁴ Nous reviendrons dans les chapitres suivants sur la définition de ce terme.

parmi la multitude des activités cérébrales mesurées. Nos travaux visent donc à proposer des méthodes pour les trois problématiques suivantes :

- i. permettre à l'utilisateur d'envoyer sa commande lorsqu'il le désire ;
- ii. augmenter les capacités du système à reconnaître correctement une tâche mentale ;
- iii. rendre le système robuste à certains changements qui peuvent s'opérer lors de l'utilisation du système, *i.e.* adapter le système aux possibles changements du sujet dans la manière dont celui-ci accomplit sa tâche mentale.

4 Approche générale

Afin de répondre à ces différentes problématiques, notre approche est basée sur les différentes connaissances neurophysiologiques qui peuvent être introduites afin de simplifier la reconnaissance de tâches mentales. L'ensemble des méthodes proposées dans cette thèse utilise le postulat suivant : une tâche mentale (*e.g.* imagination de mouvement) est une séquence d'états mentaux, chacun étant caractérisé par, d'une part la zone cérébrale impliquée dans la tâche mentale, et d'autre part les bandes de fréquences pour lesquelles une modification du spectre peut être observée. Utilisant ces trois types d'information *a priori* (temporelle, spatiale et fréquentielle), nous avons tenté de simplifier le travail du système de reconnaissance afin qu'il puisse reconnaître une tâche mentale spécifique parmi le flot continu de l'activité cérébrale mesurée par EEG.

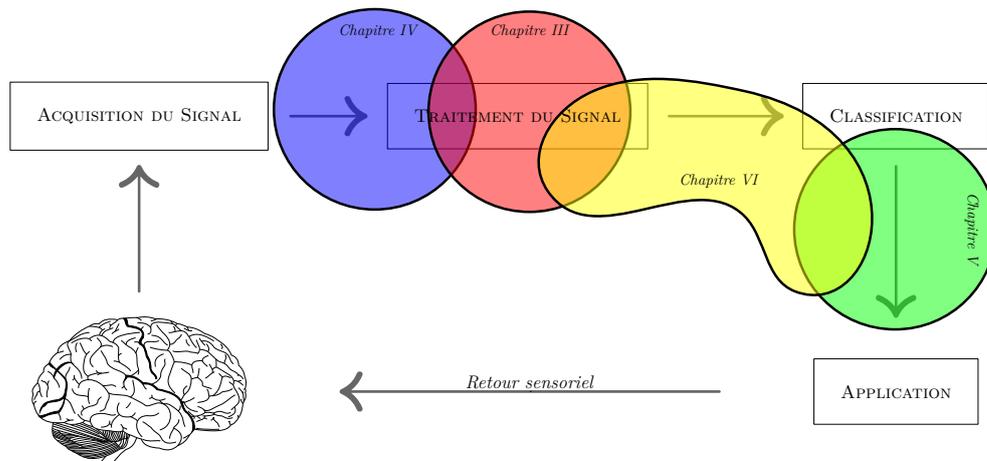


Figure I.1 Schéma de principe d'une interface cerveau-machine dans lequel sont représentés les différents blocs de traitement nécessaires à la conception d'un système. L'organisation du présent manuscrit a été superposée à ce schéma afin d'illustrer la place des différents travaux présentés dans le flot de traitement d'une interface cerveau-machine.

5 Organisation du manuscrit

Chapitre II : bien qu'il soit impossible de passer en revue l'ensemble des travaux antérieurs qui ont porté sur le domaine des interfaces cerveau-machines, nous présentons dans ce chapitre les éléments primordiaux à la base de la conception des systèmes. Après avoir présenté les différentes techniques d'imagerie qui peuvent être utilisées pour concevoir des ICMs, nous nous intéressons aux phénomènes neurophysiologiques actuellement connus et utilisés dans le domaine et associons à ces différents phénomènes des paradigmes courants, *i.e.* les conditions expérimentales qui sont utilisées pour permettre au

sujet de provoquer différents phénomènes neurophysiologiques « cibles ». Cette partie nous conduit à poser le problème de la définition des différents paradigmes utilisés et à mettre en exergue l’ambiguïté de certains termes. Afin d’y remédier, nous proposons un schéma global comportant un ensemble de variables permettant de caractériser un système ICM. Enfin, nous présentons les principes de base de traitement du signal qui sont utilisés pour l’interprétation de l’activité cérébrale des sujets. Conformément à l’approche que nous avons définie, nous nous concentrons sur les aspects des méthodes visant à utiliser les domaines temporel, fréquentiel et spatial des phénomènes neurophysiologiques.

Chapitre III : le filtrage spatial permet de passer d’une activité cérébrale globale et peu spécifique mesurée au niveau d’un ensemble de capteurs à un ensemble restreint de composantes représentant de manière efficace les activités neurophysiologiques « cible ». Nous montrons dans ce chapitre qu’une des méthodes couramment utilisée dans le contexte des ICMs peut être interprétée selon un critère utilisé pour la séparation de sources. Le cadre commun s’interprète comme une méthode d’extraction de sources non-stationnaires.

Son but est d’utiliser au maximum les différences qui existent entre les activités relatives à chacune des tâches proposées au sujet. Plus précisément, nous proposons d’adapter le cadre de l’extraction de sources non-stationnaires afin d’utiliser deux types distincts de diversité : tout d’abord la diversité entre les tâches tente d’utiliser le fait que deux tâches différentes mettent en jeu des zones cérébrales distinctes ; de plus, nous montrons qu’il est également possible d’utiliser la succession des activations lors de la réalisation d’une unique tâche mentale. En utilisant des données réelles, nous évaluons nos méthodes à l’aide de validations croisées et des performances lors du transfert entre sessions (on utilise les méthodes entraînées lors d’une première session pour classer les données enregistrées au cours d’une seconde session). Ce chapitre est innovant de par l’approche unificatrice des méthodes de filtrage linéaires utilisées dans le domaine des ICMs. Il propose des méthodes nouvelles qui sont validées par un bon taux de classification lors du transfert entre sessions. Enfin, nous concluons le chapitre en proposant une méthode capable d’utiliser également les informations du domaine fréquentiel. Bien que les données considérées dans ce chapitre ne soit pas issues de paradigmes asynchrones, la problématique abordée est exactement la même et les méthodes proposées sont directement applicables aux systèmes ICMs dans lesquels le sujet est libre d’envoyer sa commande au moment où il le désire. Les méthodes sont ainsi utilisées dans les chapitres qui suivent.

Chapitre IV : S’il est possible d’obtenir des taux de classification relativement élevés dans des conditions expérimentales très contrôlées, il faut néanmoins constater que le nombre de systèmes ICMs réellement utilisés par des patients s’avère très faible. Une des raisons qui peut expliquer cette situation réside dans la contamination des signaux par des artefacts extérieurs très puissants. Ceux-ci troublent en effet la bonne interprétation des signaux EEG, conduisant à un contrôle du système très entravé. Nous considérons donc que le débruitage des signaux est une partie essentielle des traitements nécessaires à la bonne interprétation de l’activité EEG. Ce chapitre aborde de manière originale et simple le problème de la contamination des signaux par les artefacts oculaires. La méthode que nous présentons, qui est une adaptation particulière d’un cadre général proposé par Dr. Reza Sameni ([Sameni et al., 2009](#) et [Sameni, 2008](#)), vise à débruiter simplement et en temps réel les signaux. En proposant deux variantes de la méthode, nous montrons qu’il est possible de parvenir à un débruitage quasi-parfait des signaux en utilisant l’une d’entre elles basée sur la transformée en ondelettes des signaux (un débruitage parfait consiste à enlever l’ensemble des signaux contaminant en ne modifiant strictement rien des signaux cérébraux EEG). Si les différences engendrées par l’utilisation de notre méthode pour l’augmentation du taux de classification dans une expérience d’ICM synchrone et en environnement contraint s’avèrent très faible, nous pensons néanmoins que notre méthode est d’un intérêt important afin de rendre les ICMs utilisables dans un environnement libre.

Chapitre V : ce chapitre est consacré plus spécifiquement aux interfaces cerveau-machines qui permettent à l’utilisateur d’envoyer sa commande lorsqu’il le désire. Afin

d'analyser l'activité cérébrale en temps réel, nous proposons d'aller au-delà d'une analyse fenêtre par fenêtre des signaux et proposons d'utiliser un modèle de Markov afin de prendre en compte les dépendances temporelles existant entre les différents états mentaux. Dans un cadre variationnel Bayésien, nous obtenons les équations très simples qui permettent d'inférer, en temps réel, l'état mental du sujet à partir d'une observation constituée d'un ensemble de probabilités pour que le sujet soit dans un état mental défini. Cette approche nouvelle dans le cadre des ICMs s'avère particulièrement intéressante en raison de la possibilité d'inférer l'ensemble des variables du modèle au fur et à mesure de l'expérience. Il est notamment possible d'inférer la matrice de transition et d'observer son évolution dans le temps. Cette formulation est couplée à l'ensemble des méthodes développées dans les chapitres précédents pour la classification de tâches mentales dans un contexte asynchrone. Enfin, nous proposons de coupler l'ensemble du système proposé à un algorithme adaptatif permettant d'effectuer un réapprentissage des tâches par le système. La méthode proposée est largement évaluée à l'aide de données provenant de deux « BCI Competitions ». Nous montrons que la méthode proposée dans ce chapitre engendre pour certains sujets des améliorations substantielles de performances.

Chapitre VI : nous abordons deux problèmes d'une importance primordiale pour la conception de systèmes ICMs réels. Le premier concerne l'évaluation de la qualité de réalisation des tâches par le sujet pendant la phase d'entraînement. Nous proposons une application originale du modèle du **chapitre V** qui utilise l'apprentissage automatique de la matrice de transition afin de donner une unique mesure de la qualité de chacun des essais réalisés lors de l'apprentissage. Dans un deuxième temps, nous proposons un système adaptatif afin de prendre en compte les possibles inconstances du sujet dans la manière de réaliser les tâches mentales. Ce système est basé sur l'estimation de la distribution des probabilités en sortie du classifieur.

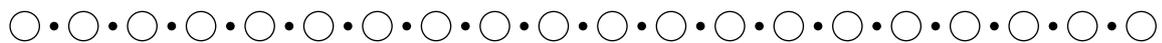
6 Références du chapitre

- Arrouet, C., Congedo, M., Marvie, J., Lamarche, F. et Lécuyer, A. et al. (2005). Open-ViBE: a 3D Platform for Real-Time Neuroscience. *Journal of Neurotherapy*, 9(1):1–22.
- Berger, H. (1929). Über das elektroencephalogramm des menschen. *Archiv für Psychiatrie und Nervenkrankheiten*, 87(1):527–570.
- Berger, T. W., Chapin, J. K., Gerhardt, G. A., McFarland, D. J. et Principe, J. C. et al. (2007). International assessment of research and development in brain-computer interfaces. Rapport technique, World Evaluation Technology Center, Inc.
- Besserve, M. (2007). *Analyse de la dynamique neuronale pour les Interfaces Cerveau-Machines : un retour aux sources*. Thèse de doctorat, Université Paris Sud.
- Clynes, M. E. et Kline, N. S. (1960). Cyborgs and space. *Astronautics*, 1960(1):26–27, 74–75.
- Harper, R., Rodden, T., Rogers, Y. et Sellen, A. (2008). Being human: Human-computer interaction in the year 2020. Rapport technique, Microsoft Corporation.
- Heingartner, D. (2009). Mental block. *IEEE Spectrum*, 46(1):42–43.
- Nicolelis, M. A. (2001). Actions from thoughts. *Nature*, 409(6818):403–407.
- Poulain, A., Piovan, R., de Varax, A. et Letellier, M. (2007). Etude prospective sur les technologies pour la santé et l'autonomie. Rapport technique, ALCIMED.
- Sameni, R. (2008). *Extraction of Fetal Cardiac Signals from an Array of Abdominal Maternal Recordings*. Thèse de doctorat, Institut Polytechnique de Grenoble.
- Sameni, R., Jutten, C. et Shamsollahi, M. B. (2009). A deflation procedure for subspace decomposition. *IEEE Trans. Signal Process.*. Submitted (in review process).
- Vidal, J. J. (1977). Real-time detection of brain events in EEG. *Proc. IEEE*, 65(5):633–641.

- Vidal, J. J. (1973). Toward direct brain-computer communication. *Annu. Rev. Biophys. Bioeng.*, 2(1):157–180.
- Williams, E. (2008). Human performance. Rapport technique, JASON, The MITRE Corporation.

Chapitre II

INTERFACES CERVEAU-MACHINE : SITUATION ACTUELLE



1	Introduction	10
2	Mesurer l'activité cérébrale	10
2.1	Techniques d'imagerie cérébrale	11
2.2	Résumé	18
3	Indices électrophysiologiques et interfaces cerveau-machine	18
3.1	P300	19
3.2	Potentiels évoqués stationnaires	20
3.3	Modulation des rythmes cérébraux	21
3.4	Potentiels corticaux lents	24
4	Différents types d'interfaces cerveau-machine	24
4.1	Point de vue du sujet	25
4.2	Point de vue du système	26
4.3	Quelques exemples	28
4.4	Définitions	29
5	Méthodologie	31
5.1	Caractériser une activité cérébrale	31
5.2	Apprentissage et classification	39
6	Enjeux actuels et problématiques de la thèse	46
7	Conclusion	47
8	Références du chapitre	47

CE CHAPITRE présente les Interfaces Cerveau-Machines (ICMs) telles qu'elles existent actuellement, leur principe est expliqué en détail et nous décrivons les différentes étapes de traitement qui sont nécessaires à leur conception. Les méthodes couramment utilisées pour la partie traitement du signal sont décrites. Après avoir défini le concept général d'une ICM dans la [section 1](#), nous décrivons les systèmes d'imagerie qui peuvent être utilisés pour mesurer l'activité cérébrale du sujet ([section 2](#)). Après avoir expliqué les raisons qui font de l'électroencéphalographie (EEG) le système actuellement le plus adapté à la conception d'ICMs, nous expliciterons les principes de base de cette mesure afin de faire ressortir les problèmes d'interprétation des signaux inhérents à ce système d'acquisition. La [section 3](#) décrira ensuite les différents phénomènes neurophysiologiques qui peuvent être utilisés pour concevoir une ICM. Ceci nous conduira alors à présenter une taxonomie des différents types d'ICM ([section 4](#)) en décrivant les systèmes mis au point par les principales équipes du domaine. Une brève revue des différentes méthodologies du domaine sera proposée dans la [section 5](#). Ce chapitre sera conclu par une description détaillée des différents enjeux actuels du domaine qui servira de base à la problématique

de cette thèse. Le nombre de travaux dans le domaine a augmenté exponentiellement au cours des 20 dernières années, rendant quasiment impossible une revue exhaustive. Nous avons néanmoins tenté de fournir au lecteur les bases de compréhension afin de faciliter la lecture des chapitres suivants. Notre travail s'est également concentré sur les aspects plus nouveaux abordés dans cette thèse : c'est-à-dire le concept d'interfaces cerveau-machines asynchrones et les bases neurophysiologiques sur lesquelles s'appuient notre travail. Une revue plus spécifique des travaux précédents relative aux différents chapitres de la thèse sera fournie au début de chaque chapitre. Le lecteur intéressé par une revue plus exhaustive du domaine pourra, en fonction de ses intérêts propres, se tourner vers [Lebedev et Nicolelis \(2006\)](#), [Nicolelis \(2001\)](#), [Vaughan et al. \(2003\)](#), [Vaughan et Wolpaw \(2006\)](#), [Vaughan et al. \(1996\)](#), [Wolpaw et al. \(2000\)](#), [Wolpaw et al. \(2002\)](#) et [Birbaumer et Cohen \(2007\)](#) pour une vue d'ensemble généraliste du domaine, consulter [Curran et Stokes \(2003\)](#) et [Curran et al. \(2004\)](#) s'il désire avoir des informations sur les aspects neuropsychologiques et cognitifs liés aux domaines des ICMs, se référer à [Cunnington et al. \(1996\)](#), [Decety \(1996a\)](#), [Jeannerod \(1995\)](#), [Jeannerod \(1994\)](#), [Lotze et Halsband \(2006\)](#), [Michelon et al. \(2006\)](#), [Pfurtscheller et Neuper \(2006\)](#) et [Pfurtscheller et Neuper \(2001\)](#) pour des considérations plus neurophysiologiques, ou enfin se plonger dans les travaux relativement exhaustifs de [Bashashati et al. \(2007\)](#), [Besserve et al. \(2007\)](#), [Lotte \(2008\)](#), [Lotte et al. \(2007a\)](#) et [Wolpaw et al. \(2006\)](#) pour se faire une idée des différentes méthodes utilisées du point de vue du traitement du signal et de la classification.

1 Introduction

Une ICM est un système qui permet le contrôle direct d'un ordinateur ou de tout autre système électronique (prothèse ou robot) à partir des simples signaux émis par le cerveau. Afin de parvenir à un tel but, plusieurs aspects ont été étudiés depuis une trentaine d'années. Une ICM nécessite en effet l'assemblage de plusieurs briques de base pour pouvoir fonctionner. Ces briques sont explicitées de manière schématique sur la [figure II.1](#), elles vont guider l'ordre de présentation de cette revue.

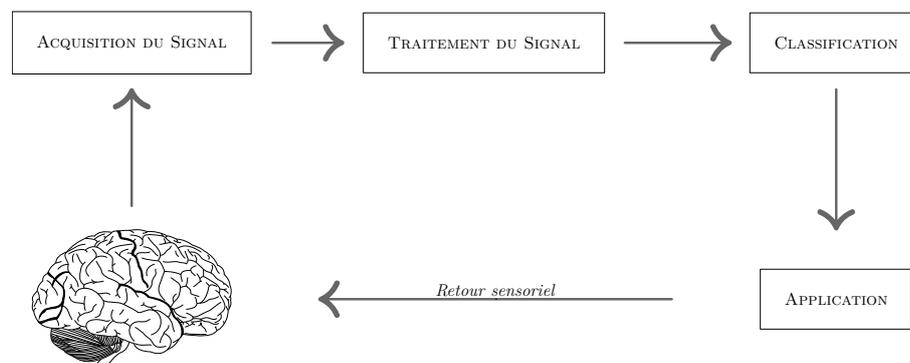


Figure II.1 Représentation schématique d'une interface cerveau-machine. Le sujet accomplit une certaine tâche mentale. L'activité induite par cette tâche mentale est ensuite évaluée, traitée puis convertie en une « commande mentale » qui est envoyée au système électronique. La plupart des systèmes fournissent alors un retour perceptif au sujet afin qu'il puisse modifier et/ou corriger sa commande.

2 Mesurer l'activité cérébrale

Afin de pouvoir utiliser les signaux du cerveau pour envoyer une commande mentale, il faut tout d'abord être capable de mesurer l'activité cérébrale du sujet. Bien que les travaux présentés dans cette thèse soient résolument davantage tournés vers l'EEG, nous

allons néanmoins présenter les différentes techniques envisageables pour mesurer l'activité cérébrale. De manière générale, on peut s'intéresser à plusieurs critères pour décrire les différentes méthodes d'imagerie dans l'objectif de construire une ICM.

- **Critères de confort**
 - a. invasif/non-invasif : un système est dit invasif lorsque sa mise en place requiert une intervention chirurgicale ;
 - b. facilité d'emploi : l'utilisation nécessite-t-elle une préparation longue et fastidieuse ?
 - c. longévité : ce critère est particulièrement important dans le cas des systèmes invasifs, il quantifie l'intervalle de temps avant qu'une nouvelle intervention soit nécessaire afin de changer ou remettre en place le système ;
- **Critères temporels**
 - d. résolution : une ICM nécessite la plupart du temps une résolution temporelle grande afin que le temps entre deux mesures de l'activité cérébrale soit faible, permettant ainsi de caractériser temporellement de manière précise la signature de la tâche mentale. Ce critère influe également sur les bandes de fréquence qui vont pouvoir être utilisées. Ainsi, une résolution temporelle fine pourra permettre d'analyser le signal jusqu'à des bandes de fréquence élevées ;
 - e. retard (ou latence) : on peut mesurer l'activité cérébrale d'un sujet toutes les millisecondes, ceci ne garantit en rien que le retard entre le changement d'état mental et la réaction du système sera elle aussi courte. En effet, si l'on mesure un signal qui commence à changer plusieurs secondes après un changement d'état mental, on aura une excellente résolution temporelle mais la réponse du système aura toujours un retard de plusieurs secondes sur l'état mental « réel » du sujet. En télécommunication cette nuance se caractérise par deux grandeurs distinctes : le délai de transmission et le débit de transmission ;
- **Critères spatiaux**
 - f. résolution : ce critère caractérise la capacité du système d'acquisition à distinguer des activités cérébrales spatialement proches.
 - g. étendue : capacité du système à mesurer l'activité globale ou localisée du cerveau.

Le type de signal mesuré (e.g. signaux électriques, magnétiques ou réponse hémodynamique) a été volontairement écarté de cette liste car il nous semble plutôt être la raison des caractéristiques du système plutôt qu'un critère permettant de le caractériser. Il nous guidera cependant pour présenter les différentes techniques d'imagerie utilisées en ICM.

2.1 Techniques d'imagerie cérébrale

On peut évaluer l'activité cérébrale d'un sujet de manière indirecte en mesurant les champs électromagnétiques produits par l'activation des neurones ; ou bien en évaluant les variations métaboliques ou hémodynamiques induites par l'activité cérébrale (variations du taux d'oxygénation du sang, consommation de glucose ou production de chaleur). Ces différentes caractérisations de l'activité cérébrale des sujets conduisent à plusieurs techniques de mesure. Après avoir expliqué les inconvénients des techniques basées sur les réponses métaboliques ou hémodynamiques, nous détaillerons plus en détail les possibilités de mesures de l'activité cérébrale du sujet à partir de l'activité électromagnétique des neurones.

2.1.1 Bases de neurophysiologie

Si le cerveau est depuis longtemps suspecté d'être le siège de la pensée (les querelles entre Platon et Aristote en témoignent), ce n'est que récemment que les hommes ont eu la possibilité d'étudier le cerveau humain en état de fonctionnement. Les approches

indirectes de la neuropsychologie du XIX^{ème} siècle, fondées tour à tour sur l'étude des caractéristiques externes du cerveau pour en déterminer les capacités¹ ou l'analyse des cerveaux endommagés afin de comprendre le rôle de la zone cérébrale « manquante »², n'ont été supplanté par les techniques d'imagerie fonctionnelle cérébrale qu'à la fin du XX^{ème} siècle. Et malgré les améliorations substantielles dans les techniques d'imagerie, il semble que le cerveau recèle encore de multiples secrets. Mais afin de comprendre comment interpréter les données fonctionnelles issues des différentes méthodes d'imagerie, nous allons dans un premier temps décrire quels signaux sont susceptibles d'être mesurés dans le cerveau.

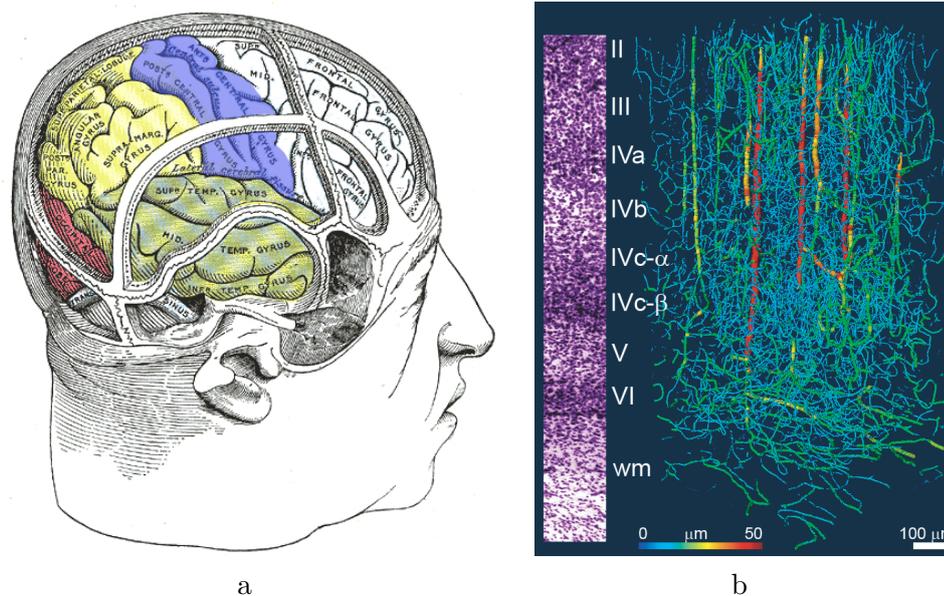


Figure II.2 a) Une représentation des différents lobes du cortex. b) Cette figure montre la densité relative des vaisseaux sanguins dans le cortex du primate. Le diamètre des vaisseaux sanguins est codé par couleurs. La surface corticale est en haut et la matière blanche en bas de la coupe. Sur la gauche de la figure, une coupe de Nissl montre la densité des neurones dans la même zone de cortex. Bien que la représentation soit trompeuse quant à la proportion relative des vaisseaux sanguins et des neurones, la proportion de vaisseaux sanguins est en fait de 3% par rapport à celle des neurones. (a) Krauledat (2008), (b) Logothetis (2008).

Le cerveau est l'organe le plus complexe du corps humain, il est composé en moyenne de 10^{11} neurones, chacun connecté en moyenne à 10^4 synapses. La matière grise, aussi appelée cortex, occupe en moyenne un volume 1,1 à 1,5 fois plus gros que celui de la matière blanche, qui est au centre du cerveau. De plus, le nombre de cellules gliales est environ 10 à 50 fois plus élevé que celui des neurones. Le tout est alimenté par un réseau complexe de vaisseaux sanguins (voir figure II.2). Parmi cet amas formant l'un des réseaux le plus complexe du monde, on connaît essentiellement le fonctionnement microscopique des neurones et sa conséquence sur le métabolisme (augmentation de la consommation d'oxygène qui provoque la réponse hémodynamique). C'est aux neurones que l'on attribue l'essentiel du fonctionnement cérébral, ils sont constitués d'un corps cellulaire (soma), d'un axone et d'un ensemble de dendrites. L'axone est considéré comme la « sortie » du neurone, connecté via des synapses³ à des dendrites, « entrée » d'autres neurones adjacents.

¹ L'expression « La bosse des maths » est un vestige de cette époque.

² Approche dont l'une des plus célèbres découverte concerne la célèbre « zone de Broca » (du nom de son découvreur), dont la fonction dans les capacités de langage (aphasie de Broca) a été découverte après que le patient « Tan » (nom donné d'après l'un des seuls mot qu'il pouvait encore prononcer) ait été hospitalisé suite à un accident qui lui avait perforé cette fameuse zone.

³ Les synapses jouent le rôle de convertisseurs électriques-chimiques-électriques.

Le potentiel d'action est le moyen de communication entre les neurones, il se propage le long des axones jusqu'aux synapses et peut éventuellement donner lieu à un potentiel post-synaptique si le potentiel de la membrane atteint une valeur supérieure à un certain seuil (variable dans le temps mais généralement de l'ordre de $-50 \mu\text{V}$). On sait (Nuñez, 2005) que les activités neuronales peuvent être approximées par des dipôles électriques pré ou post-synaptiques. Ces activités électrochimiques s'accompagnent également d'une consommation d'énergie qui est à la source de la réponse métabolique.

Différentes techniques d'imagerie cérébrale tentent de rendre compte de la complexité de ce système, chacune avec ses avantages et ses inconvénients, nous allons nous concentrer sur celles qui ont été utilisées en ICM.

2.1.2 ICM et réponses hémodynamique ou métabolique

Quelques travaux se sont intéressés à la possibilité d'utiliser les réponses hémodynamiques et métaboliques pour la conception d'ICMs.

2.1.2.1 Imagerie par résonance magnétique fonctionnelle

Technique récente utilisée pour la première fois dans les années 1990 (Ogawa et al., 1990), l'IRMf (Imagerie par Résonance Magnétique fonctionnelle) permet de mesurer de manière non-invasive une grandeur quantifiant le niveau d'oxygénation du sang dans le cerveau. Bien que le lien entre ce signal et l'activité électrique des neurones soit encore loin d'être complètement élucidé (Logothetis et al., 2001), certains auteurs se sont récemment intéressés aux possibilités d'utiliser un tel signal pour contrôler une ICM (Weiskopf et al., 2004) ou encore pour découvrir les mécanismes neuronaux impliqués dans le contrôle d'une ICM (Hinterberger et al., 2005 et Hinterberger et al., 2004). Mais comme le mentionne Birbaumer et Cohen (2007), la latence entre le signal mesuré et l'état réel du sujet est d'environ 3s ce qui signifie que le sujet obtient un retour continu de l'activité mentale qu'il a généré il y a 3s. Bien que ce retard pourrait paraître rédhibitoire, Birbaumer (2006b) et Birbaumer et Cohen (2007) argumentent en faveur de cette technique pour le futur des ICMs en s'appuyant sur une vision comportementaliste de l'apprentissage (Watson, 1913 et Curran et Stokes, 2003). Ils considèrent que le sujet peut, sous réserve que la récompense qui lui est fournie pour sa tâche soit suffisante, s'adapter à ces conditions.

L'avantage de l'IRMf est sa résolution spatiale, on peut en effet obtenir une carte d'activation des voxels¹ du cerveau d'une résolution très fine (typiquement moins d'un centimètre d'arête). La résolution temporelle ainsi que l'encombrement et le coût de tels appareils rendent leur utilisation actuellement peu adaptée aux problématiques des ICMs. Enfin, son utilisation pose des problèmes de complexité du dispositif expérimental à mettre en œuvre.

2.1.2.2 Spectroscopie proche infra-rouge

La Spectroscopie Proche de l'Infra-Rouge (SPIR), de même que l'IRMf, permet également d'obtenir une mesure indirecte du taux d'oxygénation dans le sang mais utilise une méthode optique de mesure. Une lumière infra-rouge générée par une LED est envoyée sur le scalp. L'intensité de la lumière recueillie par les détecteurs est directement proportionnelle au taux d'oxygénation du sang. L'utilisation de la réponse hémodynamique pose les mêmes problèmes que dans le cas de l'IRMf du point de vue de la latence temporelle mais son utilisation, son coût et sa mise en place sont beaucoup plus adaptés aux exigences

¹ Voxel désigne un élément de volume dans un modèle en trois dimensions, ce mot provient de l'extension de pixel (« picture element », *i.e.* élément de base d'une image) à la troisième dimension (« volume element », *i.e.* élément de base d'un volume).

des ICMs. Plusieurs publications ont montré la possibilité d'utiliser ce type d'imagerie cérébrale pour concevoir des ICMs (Rovati et al., 2007, Sitaram et al., 2007 et Coyle et al., 2004) mais ce système reste à l'heure actuelle peu répandu.

2.1.3 ICMs et réponse électromagnétique

La transmission d'information par le biais des neurones s'accompagne d'une modification des équilibres ioniques à l'intérieur ainsi qu'aux abords des neurones. Différents systèmes permettent d'accéder à ces quantités, de manière plus ou moins invasive et mesurant différents types de potentiels. Nous présentons ici les principaux systèmes utilisés en ICM, utilisant la réponse électrophysiologique accompagnant une activité cérébrale. La figure II.3 permet de résumer ces différents systèmes de manière schématique.

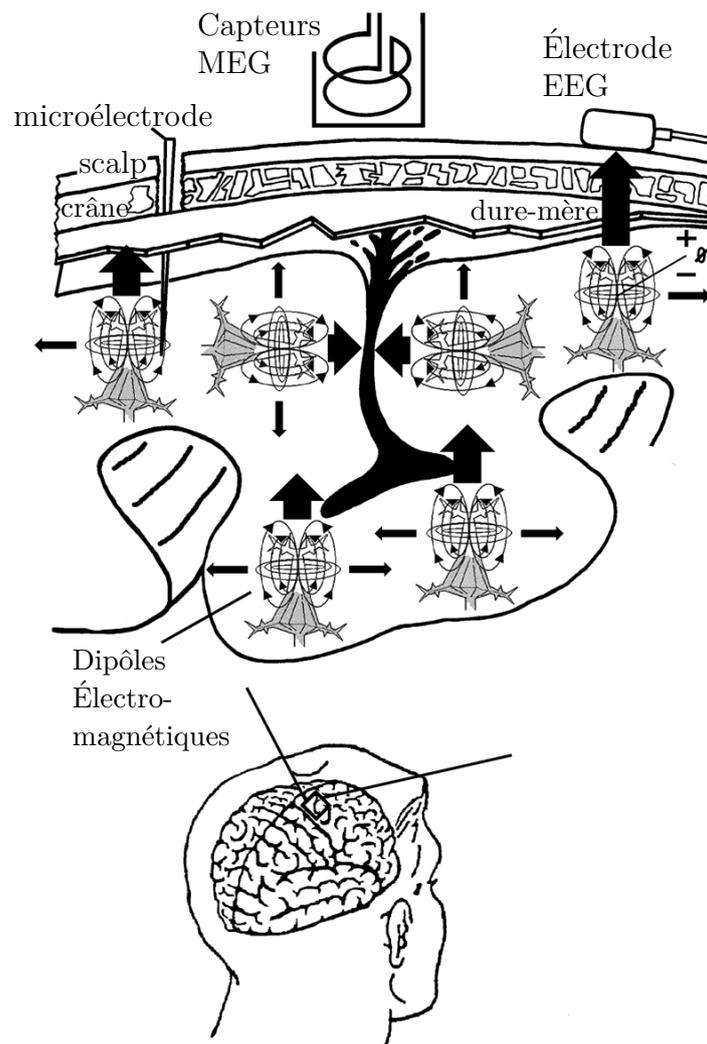


Figure II.3 Représentation schématique des principaux générateurs de champs électromagnétiques ainsi que les principaux systèmes d'imagerie permettant de les mesurer. En haut, l'agrandissement montre un repliement caractéristique du cortex (circonvolution). La coupe montre un ensemble de neurones orientés horizontalement ou verticalement à la surface du crâne. Les flèches larges en noir montrent la direction de propagation du champs électrique dû au dipôle tandis que les flèches plus fines montrent les directions de propagation du champs magnétique dû au dipôle. Modifiée et adaptée à partir de Zani et Proverbio (2002).

2.1.3.1 Microélectrodes

Les récents développements dans le domaine de la micro-électronique ont permis l'émergence de capteurs de plus en plus petits. Dans le cas du cerveau, ceci a donné naissance à une famille de capteurs appelés microélectrodes. Les microélectrodes peuvent se présenter sous la forme de fils sur lesquels les capteurs sont espacés de quelques centimètres ou encore sous la forme de matrices d'électrodes, destinées à être posées directement à la surface ou même à l'intérieur du cerveau afin de mesurer l'activité électrique produite par les neurones du cortex. La dimension des dispositifs utilisés pour effectuer les mesures conditionnent l'échelle des signaux que l'on peut espérer acquérir (Lebedev et Nicolelis, 2006, Nicolelis, 2001 et Hochberg et Donoghue, 2006). Si les dimensions nanométriques de certaines microélectrodes permettent d'avoir accès à l'activité d'un unique neurone (potentiel d'action), il est difficile de recueillir des activités très spécifiques avec des capteurs millimétriques et ceux-ci permettent plutôt de mesurer l'activité d'assemblées importantes de neurones (potentiels de champs locaux). La qualité des signaux enregistrés à l'aide de ces dispositifs est excellente, l'effet de la propagation des signaux des neurones jusqu'aux capteurs est largement diminué, le rapport signal sur bruit et la spécificité des signaux s'en trouvent par conséquent améliorés. Les microélectrodes sont de plus insensibles aux artefacts musculaires et aux mouvements des yeux qui viennent contaminer les signaux recueillis à l'extérieur de la boîte crânienne. Bien que ce type de capteurs alimente une grande partie des espoirs du domaine des ICMS, leur utilisation chez l'homme s'est avérée moins fructueuse que chez certains animaux comme le singe (Hochberg et al., 2006). Parmi les expériences très encourageantes dans le domaine, le dispositif BrainTV (Lachaux et al., 2007) a permis de grandes avancées dans l'interprétation de l'activité cérébrale de certaines zones du cerveau chez les patients épileptiques à l'aide de microélectrodes implantées.

Les microélectrodes sont des capteurs très invasifs car elles doivent être insérées à l'intérieur du cortex. Ceci pose d'importants problèmes de biocompatibilité : le dispositif doit être changé ou réinstallé à intervalles réguliers. Enfin, ce type de capteurs bénéficie actuellement d'une grande attention car on peut naturellement le modifier pour le transformer en périphérique bidirectionnel : une partie passive pour mesurer l'activité et une partie active pour stimuler les neurones et ainsi influencer sur leur fonctionnement. La stimulation neuronale a été utilisée avec succès dans un contexte médical depuis les années 90 (Benabid et al., 1991) en tant que méthode de traitement pour les sujets atteints de la maladie de Parkinson.

2.1.3.2 Électrocorticographie

L'ECOG (électrocorticographie) est également une technique d'imagerie invasive. Les capteurs sont typiquement constitués d'une matrice d'électrodes régulièrement espacées, destinée à être posée à la surface du cortex. A la différence des microélectrodes, la matrice ne nécessite pas d'être enfoncée dans le cortex et résout donc une partie des problèmes de biocompatibilité soulevés par les microélectrodes. Les signaux enregistrés, bien que moins spécifiques que dans le cas de microélectrodes micrométriques, restent très peu bruités et ne contiennent pas d'artefacts musculaires. Les signaux ECOG peuvent donc être analysés jusqu'à des fréquences de 100 Hz ou 200 Hz (fortement perturbées par les artefacts musculaires dans le cas d'enregistrements EEG).

Ces diverses qualités en font une technique très intéressante pour la conception d'interfaces cerveau-machines. Plusieurs expériences menées sur des patients épileptiques implantés avec des matrices ECOG ont montré que ceux-ci ont été capables rapidement d'influencer volontairement et significativement les signaux ECOG en utilisant l'imagerie motrice (Graumann et al., 2004 et Hill et al., 2006), l'utilisation conjointe de l'imagerie

motrice ou de la parole (Leuthardt et al., 2006 et Leuthardt et al., 2004), le calcul mental (Ramsey et al., 2006), ou encore l'imagerie auditive (Wilson et al., 2006). Enfin, dans la dernière compétition BCI 2008¹, l'un des jeux de données a montré qu'il était possible de prédire la position des doigts de la main à partir des signaux électrocorticographiques.

2.1.3.3 Magnétoencéphalographie

La MEG (magnétoencéphalographie) est une technique d'imagerie non-invasive. Elle est basée sur la mesure des variations du champ magnétique à la surface du crâne, provoquées par l'activation des neurones pyramidaux. Bien que cette technique soit très similaire à l'EEG, le développement de la MEG n'a commencé qu'à la fin des années 70. Ceci s'explique par la précision des capteurs nécessaires à la mesure des variations infinitésimales du champ magnétique ($\sim 10^{-15}$ Tesla). De plus, le coût de l'appareil ainsi que la nécessité de refroidir les capteurs avec de l'hélium liquide rendent son expansion limitée à l'heure actuelle. Néanmoins, plusieurs études ont montré la possibilité de concevoir des ICMS basées sur la MEG en utilisant l'imagerie motrice (Kauhanen et al., 2006) ou bien l'attention auditive (Mellinger et al., 2007). Ces études ont montré que le grand nombre de capteurs généralement disponibles en MEG permettait d'avoir une excellente résolution spatiale ainsi qu'une vitesse comparable aux systèmes basés sur l'EEG. Enfin, il est important de noter qu'à la différence de l'EEG qui capture préférentiellement les activités des assemblées de neurones pyramidaux formant un dipôle perpendiculaire à la surface du crâne, la MEG permet d'avoir uniquement accès aux dipôles qui sont orientés parallèlement à la surface du crâne (voir la [figure II.3](#) pour une représentation schématique de cette différence). Cette différence fait donc de la MEG et de l'EEG deux techniques d'imagerie complémentaires. De même que dans le cas de l'IRMf, il faut noter que la complexité d'utilisation d'un tel dispositif est assez importante. Ceci le rend donc peu adapté pour la diffusion au sein d'un large public des ICMS.

2.1.3.4 Électroencéphalographie

Nous présentons enfin l'EEG (électroencéphalographie) qui est la technique de loin la plus utilisée dans le domaine des ICMS. Son faible coût et sa simplicité d'utilisation par rapport aux autres techniques d'imagerie en ont fait un système de choix. Les capteurs EEG sont des électrodes Ag/AgCl (Argent/Chlorure d'argent) d'un rayon d'environ 5 mm, généralement placés à des positions standardisées (voir [figure II.4](#)). Ils permettent de capter principalement l'activité des neurones pyramidaux orientés perpendiculairement à la surface du crâne. On estime qu'il faut environ 10^6 neurones orientés dans le même sens et fonctionnant de manière synchrone pour que le potentiel post-synaptique engendré puisse être observable au niveau du scalp à l'aide de capteurs EEG (Nuñez, 2005). On considère que seuls les potentiels post-synaptiques des dendrites des neurones pyramidaux peuvent ainsi être détectés à l'aide de l'EEG. Les potentiels que ces dendrites génèrent durent en effet relativement longtemps (entre 10 et 100 ms) et l'organisation spatiale de celles-ci est relativement uniforme (perpendiculairement à la surface corticale). Ces deux propriétés particulières permettent aux potentiels générés par les dendrites des neurones pyramidaux d'être suffisamment synchrones pour que la somme qui en résulte puisse être détectée au niveau du scalp. En raison des propriétés évoquées, on modélise ce type d'activité par un dipôle électrique. L'atténuation spatiale théorique résultant de ce modèle est alors proportionnelle à $1/r^2$ (r étant la distance par rapport à la source électrique).

¹ Quatrième occurrence de cet événement : l'idée est de confronter la communauté mondiale ICM du point de vue des méthodes de traitement du signal. Voir le site Web de la dernière compétition pour plus de détails ([Site internet officiel de la compétition Interfaces Cerveau-Machines, édition IV](#)).

2.2 Résumé

Nous avons présenté différentes techniques qui peuvent être utilisées dans le cadre des ICMs pour mesurer l'activité cérébrale. Les travaux présentés dans cette thèse sont plus particulièrement dédiés à l'EEG mais les principes généraux d'analyse qui seront présentés par la suite s'appliquent de la même manière dans le cas d'une autre technique d'imagerie. Afin de résumer les avantages et inconvénients de chaque méthode d'imagerie du point de vue des ICMs, nous avons consigné l'ensemble des informations dans le [tableau II.1](#).

Système de mesure	Confort			Temporel		Spatial	
	invasif	facilité	longévité	résolution	latence	résolution	étendue
IRM	Non	–	–	–	--	+++	+++
SPIR	Non	+	–	+	--	+	+
MEG	Non	–	–	+++	+++	++	++
μ élec. Arr.	Oui	+	+	+++	+++	+++	--
ECoG	Oui	+	++	+++	+++	+++	–
EEG	Oui	+	–	+++	+++	+	+

Tableau II.1 Différents systèmes d'imagerie, évalués en fonction de critères définis pour les Interfaces Cerveau-Machines. Les lignes dont le fond est rassemblent les systèmes qui mesurent la réponse hémodynamique. Les lignes dont le fond est rassemblent les systèmes qui mesurent la réponse magnétique. Les lignes dont le fond est rassemblent les systèmes qui mesurent la réponse électrique.

3 Indices électrophysiologiques et interfaces cerveau-machine

La fiabilité d'une ICM dépend de la capacité du système à extraire de l'ensemble de l'activité cérébrale du sujet les informations pertinentes caractérisant l'état mental de celui-ci. En raison de la complexité du cerveau et des limites imposées par les systèmes d'imagerie décrits précédemment, on ne peut avoir l'ambition d'analyser l'ensemble de l'activité cérébrale du sujet pour tenter de décoder sa volonté¹ mais seulement d'associer à une tâche mentale bien connue et facilement identifiable une commande prédéterminée. Nous présentons donc succinctement les principaux phénomènes neurophysiologiques utilisées en ICM. Ce sont ces phénomènes neurophysiologiques que l'on cherche à mesurer. Ceux-ci doivent être extraits de l'activité cérébrale de fond par des transformations mathématiques afin de donner des « indices électrophysiologiques ». Les indices électrophysiologiques ont pour objectif de résumer de manière succincte (typiquement quelques variables dépendant du temps) l'activité cérébrale du sujet pour être facilement convertis en commandes.

Le grand nombre de phénomènes électrophysiologiques utilisé en ICM a rapidement donné lieu à des tentatives de taxonomies ([Curran et Stokes, 2003](#), [Wolpaw et al., 2002](#) et [Lotte, 2008](#)). De nombreux auteurs distinguent les PE (Potentiels Évoqués)² des activités cérébrales oscillatoires³. Cette distinction donne lieu à beaucoup de confusions et n'est pas réellement justifiée par des arguments neurophysiologiques. En effet, on associe

¹ Les interfaces cerveau-machines n'ont en effet pas pour but de décoder les pensées.

² Dans la littérature des ICMs, ce terme a souvent été associé à des activités cérébrales réflexes et inconscientes en réponse à un stimulus extérieur ([Curran et Stokes, 2003](#) et [Wolpaw et al., 2002](#)).

³ A l'opposé, celles-ci désignaient plutôt les activités volontairement générées par l'utilisateur, accompagnant un processus cognitif déterminé, (e.g. rotation de lettre, imagerie motrice, calcul mental).

souvent les modulations de certaines oscillations cérébrales (e.g. rythme μ (mu) associé au mouvement) à une modification (augmentation ou diminution de la densité spectrale de puissance dans la bande considérée) de l'activité de fond du cerveau tandis que les potentiels évoqués (e.g. type P300) sont plutôt considérés comme une réponse indépendante de l'activité de fond qui vient s'ajouter à celle-ci. Or cette distinction semble artificielle dans la mesure où le P300 s'accompagne également d'une modification de l'activité cérébrale de fond (Niedermeyer et Lopes da Silva, 2004, Chapitre 50). Nous préférons donc adopter la vision décrite dans Niedermeyer et Lopes da Silva (2004), Chapitre 50, page 991, qui considère que l'activité cérébrale résulte pour une partie de l'activité de fond et pour une autre partie de l'activité provoquée par des événements (ERP ou Event-Related Potentials). Cette définition ne sous-entend aucunement que ces deux termes s'ajoutent mais laisse le champ libre à au moins deux hypothèses distinctes. La première considère que certains neurones silencieux pour l'activité de fond se mettent en activité lors de la présentation de certains stimuli ou lorsque le sujet réalise une certaine tâche. Dans une deuxième hypothèse, les ERPs viennent modifier l'activité de fond, l'augmentant ou la diminuant selon des phénomènes complexes de rétroactions positives et négatives impliquant des structures telles que le thalamus. Ces deux hypothèses sont encore débattues pour plusieurs phénomènes neurophysiologiques mais il semble que la deuxième hypothèse soit aujourd'hui privilégiée (Niedermeyer et Lopes da Silva, 2004) pour la plupart d'entre eux. Enfin, par le terme « événement », il faut entendre n'importe quel phénomène exogène¹ ou endogène² donnant lieu à une modification de l'activité cérébrale. Bien que la distinction entre événements endogènes et exogènes semble relativement claire au premier abord, une analyse plus fine de chaque phénomène neurophysiologique montre que cette distinction est moins évidente qu'il n'y paraît. Dans la description sommaire des différents phénomènes qui va suivre, nous nous concentrerons donc sur les critères de classification suivants :

- déclenchement endogène et/ou exogène : l'événement déclenchant la réponse neurophysiologique provient-il du sujet seul ou bien d'un élément extérieur ?
- réponse asservie en temps et/ou phase : ce critère permet de caractériser le lien entre le déclenchement et la réponse.

3.1 P300

Le P300 est une variation positive de l'EEG. Il est visible environ 300 ms après la présentation d'un stimulus rare ou surprenant. Afin de déclencher ce type de réponse chez le sujet, on lui présente successivement dans un ordre aléatoire deux types de stimuli : les stimuli « cibles » sont ceux qui sont visés par le sujet, les stimuli « non-cibles » sont neutres pour le sujet. Il est également important que le stimulus « cible » apparaisse beaucoup moins fréquemment que les stimulus « non-cibles » afin que son apparition puisse provoquer un effet de surprise chez le sujet et ainsi générer un P300. Ce phénomène a été pour la première fois exploité par Farwell et Donchin (1988) afin de permettre au sujet d'épeler des lettres. Ils ont décrit le « P300 Speller » dans lequel le sujet était assis en face d'un écran sur lequel une matrice de 6×6 lettres était présentée. Les lignes et les colonnes de cette matrice clignotaient successivement et le sujet devait se concentrer sur une des lettres et incrémenter un compteur mental lorsque la lettre qu'il avait choisie était allumée. En moyennant son activité cérébrale, il était alors possible de savoir quelle était la ligne et la colonne « cible » en faisant apparaître dans l'EEG un P300. Ce phénomène a depuis été largement utilisé dans des paradigmes similaires. Ce type de réponse est donc déclenché

¹ Provenant de l'extérieur du corps.

² Dont la cause est interne au sujet.

de manière exogène. Néanmoins dans le paradigme présenté initialement par [Farwell et Donchin \(1988\)](#), il y a aussi une partie endogène puisque le sujet doit incrémenter un compteur mental lorsqu'il voit le stimulus « cible » apparaître. Bien que le temps entre le stimulus et la réponse soit variable (entre 250 et 350 ms), on peut considérer que cette réponse est asservie en temps par rapport au stimulus. La durée de la réponse est par contre assez courte ($\sim 50 - 100$ ms). Cette réponse est également asservie en phase : on peut moyenner directement l'EEG en se calant sur l'instant du stimuli pour obtenir la réponse P300 (bien qu'une légère inconstance d'environ 50 ms de la latence de la réponse puisse être observée).

3.2 Potentiels évoqués stationnaires

Lorsque l'on présente au cerveau des stimuli de manière répétée et régulière, on peut observer une réponse stationnaire dans des zones cérébrales qui dépendent des caractéristiques des stimuli (voir la [figure II.6](#) qui illustre le phénomène des potentiels évoqués stationnaires en faisant varier la fréquence de stimulation). Cette dénomination a été donnée à cette réponse par analogie avec le fonctionnement d'un circuit électrique et en opposition avec le caractère transitoire de la réponse P300. Ces réponses stationnaires sont également asservies en temps et en phase : on peut, de même que pour le P300, moyenner les réponses EEG suivant le stimulus pour obtenir la réponse évoquée stationnaire. Notons néanmoins que le caractère stationnaire de cette réponse rend une analyse en fréquence de la réponse EEG aussi pertinente qu'un moyennage temporel. Historiquement, c'est d'ailleurs plutôt dans le domaine de Fourier qu'ont été caractérisés pour la première fois les potentiels évoqués stationnaires ([Regan, 1966](#)). Dans une expérience basée sur des stimuli visuels, celui-ci démontra que le cerveau pouvait atteindre un régime stationnaire dans lequel l'amplitude et la phase des harmoniques de la réponse sont à peu près constantes en fonction du temps. Bien qu'en apparence, ce type de réponse soit évidemment issu d'un déclenchement exogène, l'utilisation de ce phénomène dans des paradigmes destinés aux ICMs met la plupart du temps en jeu des processus endogènes. En effet, afin de permettre au sujet d'effectuer un choix entre plusieurs commandes possibles, on associe dans la même expérience des stimuli à différentes fréquences. Le sujet doit alors se concentrer sur le stimulus de son choix afin de provoquer la réponse évoquée stationnaire correspondante. Ce type de paradigme met donc en jeu, non seulement la réponse stationnaire au déclenchement exogène, mais également des processus attentionnels lorsque le sujet passe d'un stimulus à un autre. Or ces processus attentionnels sont éminemment endogènes. Plusieurs types de stimuli destinés à provoquer des potentiels évoqués stationnaires ont été utilisés dans le cadre des ICMs.

3.2.1 Visuels

Les potentiels évoqués stationnaires visuels sont des oscillations observables au niveau des électrodes occipitales (cortex visuel) évoquées par des stimulations répétitives visuelles. Comme une stimulation à une certaine fréquence engendre dans le cerveau une réponse à cette même fréquence (et ses sous-harmoniques) ([Herrmann, 2001](#)), il est assez aisé, en utilisant le domaine spectral, de détecter sur quelle fréquence le sujet se concentre. Les systèmes utilisant ce phénomène dans le cadre des ICMs sont décrits dans [Gao et al. \(2003\)](#). Ce type de système a l'inconvénient d'être très fatigant à l'utilisation (fixation d'un ensemble de cibles clignotant à des fréquences variant entre 15 et 60 Hz). De plus, comme la concentration sur une cible clignotante s'accompagne généralement de mouvements des yeux, on peut s'interroger sur l'avantage d'un tel système sur une interface à base d'un oculomètre.

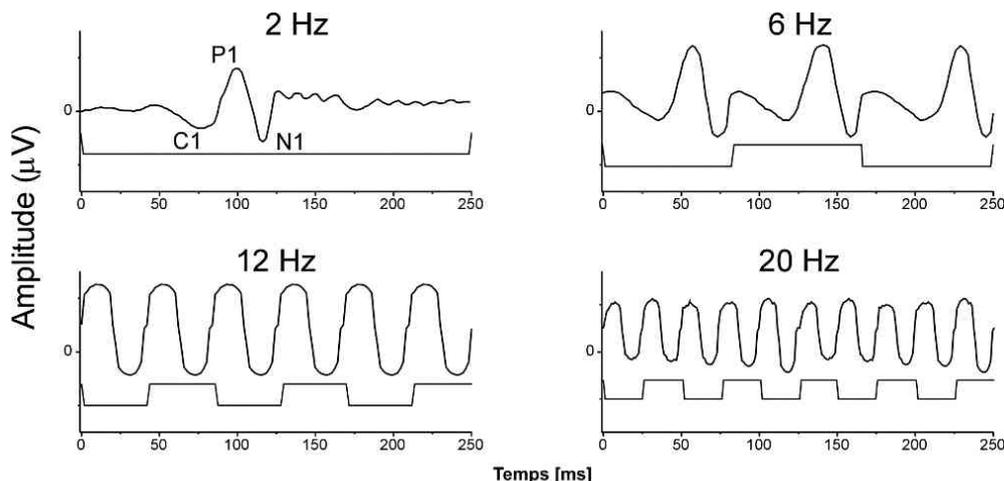


Figure II.6 D'après Zani et Proverbio (2002). Génération des potentiels évoqués stationnaires en fonction de la fréquence du stimulus visuel. En haut à gauche, le figure montre la réponse évoquée unitaire composée des pics appelés C1, P1 et N1. La répétition de la stimulation génère la réponse évoquée stationnaire selon un mécanisme qui est illustré dans les trois figures suivantes.

3.2.2 Somatosensoriels

Les potentiels évoqués stationnaires somatosensoriels reposent sur le même type de réponse cérébrale mais visent à utiliser une stimulation exogène sensorielle plutôt que visuelle. Dans le cadre des ICMs, l'idée est de présenter au sujet deux périphériques « vibrant » à des fréquences différentes. Le sujet, qui tient dans ses mains chacun des deux périphériques, peut se concentrer tour à tour sur l'un ou l'autre. On peut alors, de même que dans le cas des potentiels évoqués stationnaires visuels, détecter la fréquence « cible » et ainsi déduire l'intention du sujet. Les travaux concernant ce type d'interface sont assez récents et encore peu nombreux mais une démonstration de principe a été donnée dans Müller-Putz et al. (2006).

3.2.3 Auditifs

L'utilisation du même principe dans le cas de stimulations exogènes auditives concurrentes est également possible (Bidet-Caulet et al., 2007) mais nécessite encore des confirmations expérimentales. Dans un tel paradigme, le sujet entend deux sons concurrents (sinusoïdes à hautes fréquences) modulés par deux sinusoïdes à basses fréquences (un son dans chaque oreille). Lorsque le sujet se concentre sur l'un ou l'autre des sons, la réponse évoquée stationnaire auditive induit une augmentation de la puissance spectrale dans la fréquence du signal de modulation d'amplitude. On peut ainsi, de même que dans les deux cas précédents, construire un indice attentionnel destiné à rendre compte du côté vers lequel le sujet porte son attention. Un inconvénient de ce type de stimulus est la fatigue qu'il engendre (en raison du caractère assez désagréable des deux sons qu'il entend en continu).

3.3 Modulation des rythmes cérébraux

En plus des réponses décrites précédemment, on peut également observer dans le cerveau des réponses non-asservies en phase ou réponses induites qui sont des synchronisations ou des désynchronisations d'un ensemble de neurones en réponse à des événements

(en anglais on les nomme ERS ou ERD d'après Event-Related Synchronization et Event-Related Desynchronization). Ces réponses sont mises en évidence par le calcul de la puissance spectrale dans certaines bandes de fréquence particulières relativement à un état de repos. Ainsi, si la puissance spectrale dans la bande α (alpha) augmente (resp. diminue) après un événement, on parle d'ERS (resp. ERD). C'est Berger qui le premier a observé de tels phénomènes quand il s'aperçut que certains événements pouvaient « bloquer » le rythme α (Berger, 1929). Il est souvent admis que les ERSs correspondent à un état inactif d'une population cérébrale et, qu'inversement, les ERDs correspondent à une activité accrue d'une population neuronale : l'idée étant que lorsqu'une zone se met à être très active, les neurones ont une activité très erratique et s'écartent de leur fonctionnement en régime de repos, or ce régime de repos est une activité quasi-périodique à une fréquence qui dépend des interconnexions du groupe de neurones en jeu. En réalité, ceci n'est vrai que pour certaines fréquences et certaines zones cérébrales (Lopes da Silva, 2006). Les choses apparaissent en effet plus compliquées que cela : Lopes da Silva (2006) rappelle par exemple qu'une grosse part de la difficulté provient de l'hétérogénéité des types de neurones dans une zone cérébrale et de plus de la complexité des interconnexions entre les différents neurones (connexions excitatrices ou inhibitrices par exemple). L'interprétation de ce type de réponses dans le cas de l'EEG est de plus rendue très compliquée en raison de la faible résolution spatiale. Une description complète des différents types d'ERD-ERS qui peuvent être observés dans le cerveau est en-dehors de la portée de cette thèse, mais nous voulons cependant décrire succinctement les différentes fréquences qui peuvent être rencontrées dans les enregistrements EEG (Niedermeyer et Lopes da Silva, 2004, Chapitre 9) :

- δ (delta) : oscillations très lentes (0,1 à 3,5 Hz) d'amplitude importante principalement observées chez l'adulte pendant le sommeil profond ;
- θ (theta) : rythme légèrement plus rapide (4 à 7,5 Hz), il a été observé dans les processus émotionnels et cognitifs ;
- α (alpha) : oscillations entre 8 et 13 Hz souvent associées à des états de relaxation. Il est particulièrement visible dans les parties occipitales et a l'amplitude la plus importante des différents rythmes connus ;
- μ (mu) : ~ 10 Hz, ce rythme se distingue du rythme α par les zones cérébrales qu'il implique : le cortex sensori-moteur. Pour cette raison, on l'appelle aussi le rythme sensori-moteur. Il est bloqué lors de la préparation d'un mouvement ;
- β (beta) : oscillations relativement rapides (entre 13 et 30 Hz) observées chez l'adulte éveillé. Il est particulièrement impliqué dans l'exécution des mouvements ;
- γ (gamma) : oscillations rapides (> 30 Hz) souvent associées aux processus cognitifs supérieurs.

Le lecteur intéressé pourra trouver une description beaucoup plus complète dans Niedermeyer et Lopes da Silva (2004). La figure II.8 présente le spectre typique d'un EEG et résume les différentes bandes de fréquences les plus utilisées (le sujet était éveillé, en période de concentration juste avant une expérience d'ICM).

L'utilisation de ce type de réponse cérébrale dans le cadre des ICM repose sur une idée assez simple : il existe des tâches cognitives particulières pour lesquelles on sait qu'elles engendrent des ERDs-ERSs dans des zones cérébrales distinctes. Le sujet doit donc choisir différentes tâches mentales facilement « séparables »¹ et associer à chacune de ces tâches une commande (e.g. bouger un curseur de souris à droite ou à gauche).

¹ Ce terme sera précisé dans la suite de la thèse.

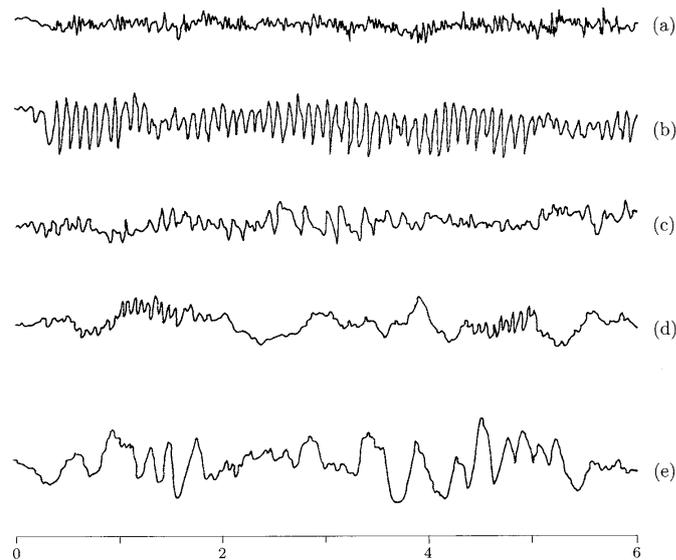


Figure II.7 Unité de temps en abscisse : seconde, en ordonnée : unité arbitraire. Electroencéphalogramme humain lors de différents états de conscience. (a) état d'excitation, (b) état relaxé, (c) état d'endormissement, (d) sommeil et (e) sommeil profond. Figure issue des travaux pionniers de H.H. Jasper sur l'électroencéphalogramme de l'homme (Jasper, 1941).

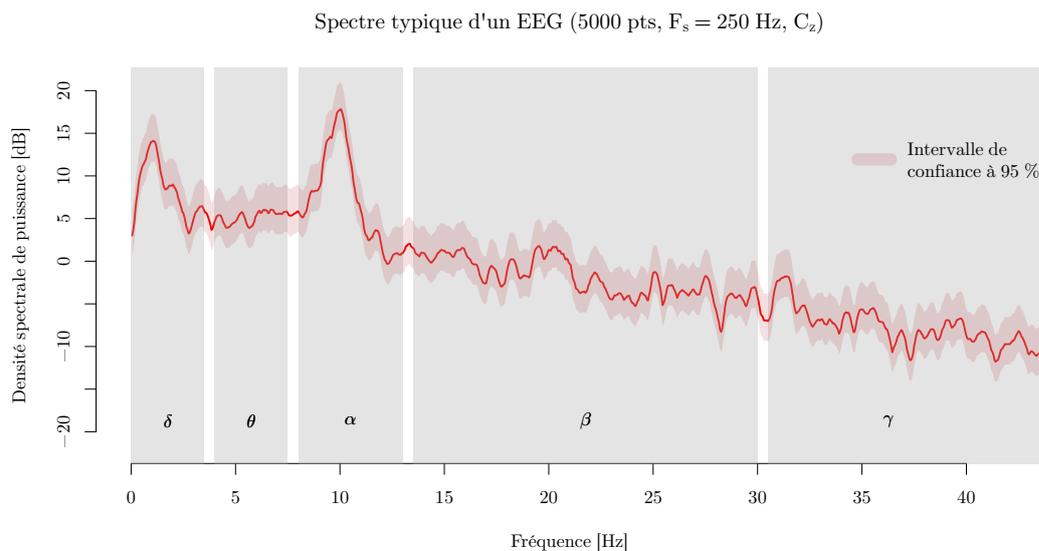


Figure II.8 Densité spectrale de puissance typique d'un EEG. Le spectre est estimé par la méthode du périodogramme et a été lissé à l'aide d'un noyau de Daniel.

3.3.1 Rythmes sensorimoteurs

On sait que les rythmes μ et β sont modulés au niveau des régions sensori-motrices lorsque les sujets réalisent ou imaginent un mouvement. Nous reviendrons plus en détail sur les principes neurophysiologiques sous-jacents à ce phénomène dans le [chapitre III](#). De plus, on sait que ces modifications sont observées dans des zones organisées de manière somatotopique (une zone du corps correspond de manière unique à une zone cérébrale). C'est principalement cette dernière propriété qui permet de discriminer entre différentes tâches d'imagerie motrice. Ainsi, on peut par exemple discriminer entre une imagerie du bras gauche par rapport à une imagerie du bras droit. Les systèmes ICMs basés sur l'imagerie motrice ont été largement répandus par l'équipe du Prof. Pfurtscheller (Graz, Autriche). Les paradigmes utilisés ont d'abord considéré deux tâches distinctes (Pfurtscheller et al.,

1996), mais les avancées technologiques ainsi que les progrès dans les méthodes de traitement mises en œuvre ont récemment permis d'étendre ces systèmes à des paradigmes dans lesquels les tâches mentales étaient plus nombreuses (Pfurtscheller et Neuper, 2001). Le groupe du Prof. Wolpaw aux États-Unis a également mis au point des systèmes basés sur ce type de réponse cérébrale pour le contrôle « rapide » d'un curseur en deux dimensions (Wolpaw et McFarland, 2004). De nombreux groupes travaillent aujourd'hui sur ce type de systèmes en utilisant différents systèmes de mesure : la SPIR (Coyle et al., 2004 et Sitaram et al., 2007), l'ECOG (Grimann et al., 2004 et Leuthardt et al., 2006) ou encore la MEG (Kauhanen et al., 2006 et Kübler et Neumann, 2005).

3.3.2 Rythmes « cognitifs »

On observe une modulation des rythmes cérébraux principaux (α et β) dans d'autres situations que dans l'imagerie motrice. Plusieurs exemples de telles tâches cognitives ont été utilisés dans des paradigmes ICMs. Dans la mesure où il suffit que les distributions spatiales de telles tâches soient facilement séparables, on peut envisager un grand nombre de tâches cognitives possibles. Des succès ont été rapportés (Curran et al., 2004) avec la rotation de lettres, le calcul mental et la navigation spatiale. Les méthodes de traitement utilisées dans ce type de systèmes sont semblables à celles qui ont été mises en œuvre pour l'imagerie motrice et les réponses sensori-motrices.

3.4 Potentiels corticaux lents

Les potentiels corticaux lents sont des variations de voltage dans l'EEG dans des fréquences inférieures à 1 Hz. On considère généralement (Birbaumer, 2006b) qu'une diminution de ces potentiels correspond à une réduction de la capacité d'excitabilité des neurones alors qu'une augmentation de ces potentiels correspond à une excitabilité accrue. Birbaumer et al. (2000) ont montré que les sujets pouvaient apprendre à contrôler volontairement ces potentiels à la suite de longues séances d'entraînement durant lesquelles un retour sensoriel était fourni au sujet. Ce type de paradigme se distingue des précédents par le fait qu'aucune tâche mentale précise n'est demandée au sujet pour accomplir ce contrôle volontaire. Cette raison est généralement avancée pour justifier l'apprentissage lent que nécessite ce type de systèmes. Ce type de réponse n'est, de même que pour la modulation des rythmes cérébraux, pas asservi en phase et nécessite donc d'utiliser des méthodes qui sont assez similaires à celles qui permettent l'analyse des ERDs-ERSs.

4 Différents types d'interfaces cerveau-machine

Après avoir présenté les différents types de réponses cérébrales qui peuvent être utilisés dans le cadre des ICMs, nous allons maintenant nous intéresser plus précisément aux paradigmes qui visent à les utiliser. La littérature du domaine foisonne d'adjectifs permettant de qualifier les différents types d'ICMs qui ont été développés par les équipes travaillant dans le domaine. Leur utilisation porte cependant souvent à confusion. Le terme **asynchrone** est ainsi apparu dans les années 2000 (Mason et Birch, 2000) pour décrire un nouveau type de système dans lequel le sujet n'était pas contraint à envoyer sa commande à des moments précis imposés par le système électronique. Or le terme **asynchrone** s'est vite avéré inadéquat pour distinguer certains systèmes : en effet, comment rendre compte de la différence entre une imagerie motrice réalisée sans aucune contrainte temporelle et un système basé sur les réponses évoquées stationnaires dans lequel le sujet peut changer à tout moment de tâche en se concentrant sur le stimulus « cible » de son choix ? Dans les deux cas, aucune contrainte temporelle n'est imposée au sujet mais il fallait rendre compte de la différence fondamentale entre un système basé sur un stimulus extérieur (bien que

n'imposant pas de contraintes temporelles) et un système dans lequel le sujet n'est soumis à aucun stimulus (ni pour lui indiquer le moment de sa tâche ni pour provoquer chez lui une réponse évoquée stationnaire). L'introduction des termes **cue-based** (Pfurtscheller et al., 2003) (une contrainte temporelle est imposée à l'utilisateur pour réaliser sa tâche) et **self-paced** (Mason et al., 2006) (le sujet décide du moment de sa commande) a alors tenté de lever cette ambiguïté. Mais, comme le montre en détail Mason et al. (2006), l'usage courant visant à employer indifféremment **asynchrone** et **self-paced** n'a pas rendu la situation plus claire. Cette section va tenter d'éclaircir la taxonomie des systèmes ICMs de façon à rendre compte de la multitude des paradigmes existants. Nous proposons de résumer une ICM par le diagramme de la figure II.9, qui a pour but de lever les différentes ambiguïtés lexicales. Ce schéma montre de manière précise l'ensemble des éléments nécessaires à la conception d'une ICM. Le principe de base apparaît sur ce schéma : construire un système qui établit une bijection entre l'ensemble des commandes disponibles de l'effecteur et l'ensemble des tâches mentales proposées au sujet. Ce schéma inclut des commutateurs $\mathbf{u}(t)$ et $\mathbf{v}(t)$. Ceux-ci ont pour objectif de rendre compte des différents états du système ainsi que de l'utilisateur dans l'hypothèse d'une utilisation en continu du dispositif (dans la vie de tous les jours). Ils permettent ainsi de rendre compte des périodes pendant lesquelles le sujet ne veut pas envoyer de commandes au système ou quand le système n'est pas en mesure d'analyser l'ensemble du flot de données.

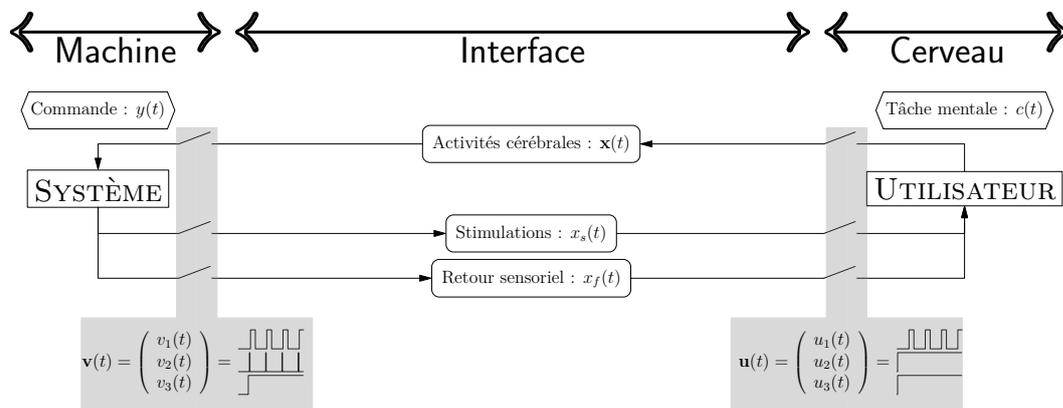


Figure II.9 Communication cerveau-machines. Ce schéma détaille le fonctionnement d'une interface cerveau-machine en séparant le rôle du sujet et du système. Les commentaires détaillés sont donnés dans les paragraphes ci-dessous.

4.1 Point de vue du sujet

4.1.1 Les tâches mentales

Nous avons d'ores et déjà décrit les phénomènes neurophysiologiques qui sont couramment utilisés. Ceux-ci ne doivent pas être confondus avec les tâches mentales qui permettent de les générer. Ainsi, pour générer une modulation d'un rythme moteur, on demande au sujet d'imaginer une tâche d'imagerie motrice. De la même façon, pour générer un P300, on demande au sujet de se concentrer sur une lettre et d'incrémenter un compteur mental. Il est difficile de détailler les différentes tâches qui ont permis de concevoir des ICMs. Certains exemples ont été donnés précédemment et nous en donnerons certains autres dans la suite.

4.1.2 « Interrupteurs utilisateurs »

La définition d'une ICM passe par la prise en compte de l'état de l'utilisateur au fil du temps. Sur le schéma de la figure II.9, trois variables temporelles permettent d'en rendre

compte : $\mathbf{u}(t) = [u_1(t), u_2(t), u_3(t)]^T$. La première rend compte de la possibilité pour le sujet de passer dans un état où il décide de ne plus contrôler le système ($u_1(t) = 0$), ces pensées sont alors libres et ne doivent pas interférer avec les commandes. De manière générale, on peut associer cet état à la volonté délibérée du sujet de ne pas choisir parmi les tâches mentales qui lui sont proposées ($c(t)$ est alors indéfinie). Cet état est ainsi connu dans la littérature du domaine comme l'état de repos. Les deux dernières variables sont liées et expriment le fait que le sujet peut décider sciemment de ne plus recevoir les informations fournies par le système (e.g. il éteint le système, $u_1(t) = u_2(t) = u_3(t) = 0$).

4.2 Point de vue du système

Nous allons à présent nous placer du point de vue du système et décrire les différentes variables qui décrivent son fonctionnement.

4.2.1 Les commandes

De manière symétrique, le système propose à l'utilisateur un ensemble de commandes activées lorsque la tâche mentale qui lui est associée est détectée. Comme mentionné dans [Vaughan et al. \(2003\)](#), il est souvent opportun d'initier la conception d'une ICM par l'analyse du besoin auquel on veut répondre. Ainsi, on ne propose pas les mêmes commandes suivant que l'on désire fournir à l'utilisateur une prothèse motrice, un moyen d'épeler des lettres ou encore un moyen d'allumer ou éteindre un poste de télévision. Il apparaît ainsi clairement deux critères permettant de caractériser les différentes commandes qui peuvent être proposées à l'utilisateur :

- densité temporelle : permet d'exprimer combien de commandes¹ on peut espérer envoyer par unité de temps . On peut le définir mathématiquement comme le nombre de commandes qu'on est capable d'envoyer toutes les secondes. On adopte la convention selon laquelle un système permettant d'avoir une réponse continue dans le temps a une densité temporelle infinie ;
- Nombre de commandes : de même on doit caractériser une ICM en fonction du nombre de commandes qu'elle met à la disposition de l'utilisateur. De manière analogue au critère précédent, on peut considérer qu'une ICM permettant de sélectionner une commande dans un espace continu ou dénombrable aurait un nombre de commandes infini.

On a représenté de manière schématique les cas correspondant aux extrêmes de chaque critère sur la [figure II.10](#). Cette figure montre quatre cas d'ICMs suivant les deux critères présentés dans cette section. Notons que cette proposition de classification d'après ces deux critères n'a pas, à notre connaissance, été proposée dans la littérature jusqu'à présent. On peut trouver des similarités entre la présentation des interfaces asynchrones qui a été faite dans [Mason et al. \(2006\)](#) et la présentation décrite ici. Cependant, ce rapport fait la distinction entre les ICMs en se basant uniquement sur un argument de commandes discrètes ou continues et passe sous silence la nuance importante visant faire la distinction entre la continuité dans le temps ou bien dans le nombre de commandes possibles.

4.2.2 « Interrupteurs systèmes »

Trois interrupteurs sont couramment utilisés en ICM pour réguler les périodes d'analyse de l'activité cérébrale et envoyer un retour sensoriel et des stimulations au sujet. Le premier, $v_1(t)$, permet au système de se concentrer sur les périodes pendant lesquelles il

¹ Par commande il faut entendre combien de commandes « nouvelles » générées par analyse de l'état mental du sujet. Une interpolation entre deux commandes ne constitue en effet pas une amélioration du nombre de commandes possibles par seconde.

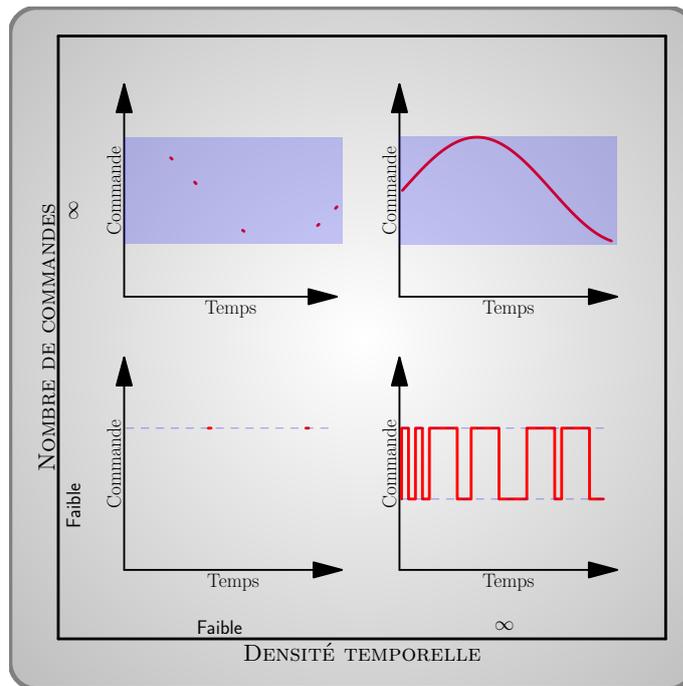


Figure II.10 Critères permettant de classer les différents types d'interfaces cerveau-machine en fonction des commandes du système. Dans chacun des quatre quadrants de la figure, on a représenté les cas les plus extrêmes du point de vue de la densité temporelle et du nombre de commandes. Les deux figures de gauche caractérisent des systèmes pour lesquels une commande est envoyée à des instants t relativement espacés dans le temps alors que celles de droite impliquent un contrôle continu du système. De plus, les deux figures du bas symbolisent des systèmes qui proposent à l'utilisateur un nombre de commandes fini (respectivement une et deux), tandis que les figures du haut montrent le fonctionnement d'un système proposant un nombre infini de commandes.

sait qu'une tâche est réalisée par le sujet. Tandis que $v_2(t)$ et $v_3(t)$ servent à bloquer les stimulations et/ou le retour sensoriel lorsque le système est indisponible.

4.2.3 Stimulations et retour sensoriel

Ces signaux (respectivement $x_s(t)$ et $x_f(t)$) sont envoyés par le système à destination de l'utilisateur. La variable de stimulation peut être constituée de deux types de signaux possibles dont les buts sont différents : elle peut être un signal destiné à provoquer une réponse évoquée chez le sujet (PE) ou bien un signal destiné seulement à indiquer au sujet une action à entreprendre (typiquement lui indiquer le début de sa tâche mentale). Il peut néanmoins être difficile de distinguer entre ces deux catégories et c'est pourquoi nous avons choisi de les rassembler au sein d'une même classe. En effet, il arrive que l'affichage d'un signal de déclenchement sur un écran (une croix par exemple) provoque chez le sujet un potentiel évoqué... les deux classes présentées comme distinctes perdent alors leurs frontières. Notons qu'un système basé sur un signal de stimulation s'avère relativement plus épuisant mentalement que les systèmes pour lesquels la stimulation est nulle (Curran et Stokes, 2003). Caractériser la stimulation de manière précise peut s'avérer important pour évaluer la charge cognitive que va nécessiter l'utilisation d'une ICM (Ryu et Myung, 2005).

Enfin, la figure II.9 inclut un signal émis par le système en direction du sujet que nous avons appelé retour sensoriel. Ce retour a pour but de fournir au sujet des informations importantes sur la tâche qu'il est en train de réaliser afin de l'aider à l'accomplir. Ce signal

n'est pas présent dans tous les systèmes, son influence reste pour le moment relativement peu étudiée dans les applications réelles.

4.3 Quelques exemples

Les paragraphes suivants nous permettent d'éclaircir les notions abordées précédemment à partir d'exemples tirés de systèmes réels.

4.3.1 « Thought Translation Device »

TTD : « Thought Translation Device » fut l'un des premiers systèmes ICMs. Développé par l'équipe du Prof. Birbaumer ([Birbaumer et al., 2000](#)), il est basé sur les potentiels corticaux lents. Il a été conçu pour aider des personnes atteints de SLA à épeler des lettres en sélectionnant successivement entre deux parties de l'alphabet. Ce système fonctionne en observant les augmentations et diminutions des potentiels corticaux lents et en associant à chacun de ces deux cas une sélection d'une partie de l'alphabet. Ce système particulier permet d'épeler des lettres à une vitesse d'environ une lettre toute les deux minutes ([Birbaumer et al., 2003](#) et [Birbaumer et al., 2000](#)). Si l'on considère la commande comme la sélection d'une des deux parties de l'alphabet, le nombre de commandes possibles est de deux et la densité temporelle est très faible. Quant aux variables caractéristiques qui régissent le fonctionnement du système, on peut observer que le sujet est contraint d'effectuer sa tâche mentale lorsque le système lui en a donné l'ordre ($x_s(t)$ est non nulle, cette variable doit indiquer à l'utilisateur le moment du départ de sa tâche). Enfin, comme le système sait à quel moment le sujet doit réaliser sa tâche, la variable $v_1(t)$ n'est pas constamment sur « ON » ($\exists t \in \mathbb{R} \mid v_1(t) \neq 1$). L'activité cérébrale du sujet n'est donc pas synchronisée avec un stimulus externe mais la réalisation de ces tâches mentales est contrainte à être synchronisée avec les signaux de déclenchement des tâches. L'entraînement au contrôle de ce type de périphérique, qui est basé sur les principes de conditionnement opérant, nécessite souvent une longue période d'apprentissage.

4.3.2 « P300 Speller »

Le « P300 Speller » est sans doute l'application la plus connue du domaine. Elle permet à l'utilisateur d'épeler des lettres en se concentrant sur la lettre de son choix. Comme son nom l'indique, ce système est basé sur le phénomène neurophysiologique du P300, qui est visible lors de l'apparition d'un stimulus rare parmi un nombre important de stimulus non-cible. Dans l'application initiale proposée dans [Farwell et Donchin \(1988\)](#), le sujet est placé devant une matrice de lettres de dimension 6×6 dont les lignes et les colonnes clignotent tour à tour (toutes les 175 ms). On lui demande de se concentrer sur la lettre qu'il veut épeler et d'incrémenter un compteur mental lorsque cette lettre clignote. Ce système a fait l'objet de multiples améliorations du point de vue du traitement du signal ([Rivet et al., 2009](#) et [Rivet et al., 2008](#)) et de la classification ([Hoffmann, 2007](#), [Hoffmann et al., 2007](#) et [Krusienski et al., 2006](#)) et on est aujourd'hui capable de choisir parmi les lettres de la matrice après seulement quelques (~ 5) clignotements de la lettre « cible ». Dans cette application, la stimulation est donc un signal périodique de période 125 ms constitué par les clignotements successifs. Ce signal sert à la fois de signal indiquant le moment auquel on s'attend à détecter une réponse P300 (en l'occurrence 300 ms après une stimulation) et également de signal de déclenchement de la réponse puisque le P300 (en tant que réponse neurophysiologique) est « évoqué » par la stimulation. Certains sujets seraient parvenus à épeler environ 8 lettres par minute ([Donchin et al., 2000](#)). Notons enfin qu'il a été démontré qu'il était possible de parvenir au contrôle de ce type d'interface par des sujets naïfs (aucun entraînement préalable). De plus, on peut noter depuis ces dernières

années un foisonnement des études basées sur le P300, celles-ci utilisant le fait qu'on peut remplacer les lettres de la matrice présentée au sujet par n'importe quel autre symbole.

4.3.3 Contrôle d'un curseur et rythmes sensorimoteurs

C'est à Albany (États-Unis) dans l'institut de recherche « Wadsworth Center » qu'a été développé dans les années 1990 un système basé sur le contrôle des rythmes moteurs (Wolpaw et al., 1991) dans le but de contrôler un curseur à l'écran. Basé sur les principes du conditionnement opérant, ce rythme associe aux augmentations et diminutions de la puissance dans les rythmes cérébraux μ et β des directions du mouvement de la souris sur l'écran. Dans l'application originelle, le curseur bouge à une vitesse constante de gauche à droite sur l'écran. Le déplacement vertical du curseur est quant à lui soumis aux variations d'amplitude dans les bandes de fréquences considérées (une augmentation est associée à un déplacement vers le haut et une diminution à un déplacement vers le bas). Dans ce système, aucune stimulation n'est présente ($\forall t \in \mathbb{R} \quad x_s(t) = 0$), le sujet peut tenter de faire varier l'amplitude de ces rythmes cérébraux à tout moment et le système mesure en continu les variables d'intérêt. Ce système a donc été considéré comme l'un des premiers systèmes asynchrones, au sens où le sujet peut envoyer une commande mentale au moment où il le désire. Malgré l'intérêt d'un tel paradigme, ce système n'est pas très répandu car l'entraînement nécessaire est assez important. Le nombre de commande est ici faible puisqu'on quantifie la puissance selon deux niveaux (augmentation ou diminution) mais les commandes pourraient en réalité être envoyées en continu.

4.3.4 Imagination de mouvement (Graz)

Contrairement au système précédent fondé sur des principes de conditionnement opérant, l'équipe du Prof. Pfurtscheller à Graz (Autriche) a développé un système fondé également sur la modulation des rythmes cérébraux mais dont l'apprentissage est simplifié par la définition de tâches cognitives précises. Les premiers systèmes utilisèrent l'imagination d'un mouvement de la main droite ou de la main gauche (Kalcher et al., 1996 et Pfurtscheller et Neuper, 2001). Cependant, la complexité de reconnaître la tâche d'imagination parmi l'ensemble de l'activité cérébrale a dû être compensée par l'introduction d'un signal de stimulation contraignant le sujet à réaliser sa tâche d'imagination à des moments précis.

4.3.5 Résumé

Nous n'avons présenté qu'une partie infime de l'ensemble des systèmes ICMs qui se sont développés depuis les années 1980. Les systèmes présentés ici sont historiquement les plus importants et ont engendré de multiples variantes par des modifications des tâches mentales utilisées ou encore des commandes mises à disposition de l'utilisateur. Cette courte revue va aussi nous permettre de définir de manière plus rigoureuse les différentes interfaces qui peuvent être rencontrées.

4.4 Définitions

Les variables et critères définis dans les figures II.9 et II.10 permettent en effet d'interpréter les différents systèmes ICMs utilisés jusqu'à présent selon une classification rigoureuse. Pour ce faire, il faut s'intéresser à la notion de causalité entre les variables définies sur la figure II.9. Et plus particulièrement aux relations de causalité existant entre le signal de stimulation fourni par le système $x_s(t)$, l'intention du sujet d'envoyer une commande $u_1(t)$, l'activité cérébrale du sujet $\mathbf{x}(t)$ et enfin la disponibilité du système pour l'analyse

de l'activité cérébrale $v_1(t)$. Nous noterons par la suite une relation de causalité entre deux de ces variables par le symbole \rightarrow . La [figure II.11](#) présente un résumé des principales ICMs qui peuvent être rencontrées.

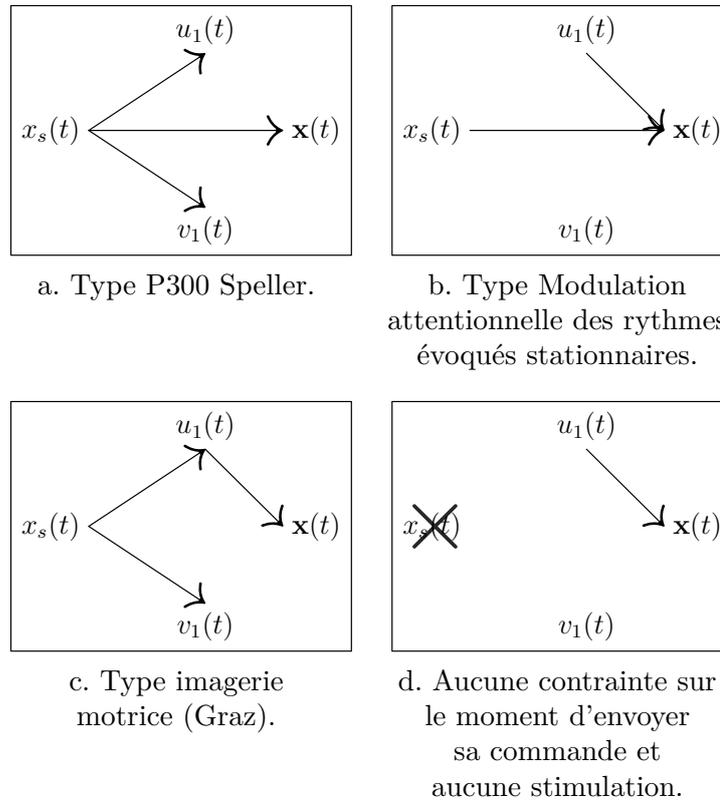


Figure II.11 Quatre différentes configurations des relations qui lient les variables propres au sujet ($u_1(t)$ indique les moments d'envoi d'une commande par le sujet et $x(t)$ représente l'activité cérébrale « cible » du sujet) et les variables propres aux systèmes ($x_s(t)$ la stimulation et $v_1(t)$ qui régit les périodes auxquelles le système analyse l'activité cérébrale).

Le concept d'interface asynchrone est historiquement né du fait que le système n'a la plupart du temps pas moyen de savoir l'état de la variable $u_1(t)$ et que les équipes travaillant sur le domaine se sont imposées la nécessité de transmettre cette information de manière uniquement cérébrale. De plus, le concept d'asynchrone s'est rapidement étendu aux interfaces pour lesquelles aucune stimulation n'est nécessaire. Le concept de « self-paced » renvoie quant à lui au fait que le sujet initie l'activité cérébrale par opposition aux activités cérébrales provoquées par un stimulus extérieur.

Souvent utilisés comme synonymes, ces deux termes ne nous semblent pas suffisamment précis pour qu'il soit possible de faire la distinction entre une tâche d'imagerie motrice libre (non contrainte d'être réalisée à des moments précis) et une tâche de modulation de l'attention. Le rapport [Mason et al. \(2006\)](#) et la définition fondamentale des interfaces asynchrones qui y est donnée est entièrement fondée sur l'existence d'une classe dans lequel le sujet peut être en période de repos (nous avons défini cet état comme correspondant au fait que $u_1(t) = 0$). Cette vision nous paraît réductrice car elle ne prend pas en compte le cas particulier des interfaces basées sur des tâches de modulation de l'attention, pour lesquelles le sujet peut effectuer sa tâche lorsqu'il le désire mais l'état de repos n'est pas forcément nécessaire (bien qu'on puisse utiliser l'importance de la modulation de l'attention pour détecter un potentiel état de repos).

La [figure II.11](#) montre que les termes utilisés jusqu'à présent ne peuvent pas rendre compte de la multiplicité des systèmes. Nous distinguerons à présent les interfaces dans lesquelles le sujet n'est pas contraint à envoyer une commande à un moment précis mais

dont l'utilisation requiert qu'il soit soumis à un stimulus, des interfaces dont la commande est libre d'être donnée à n'importe quel moment sans qu'aucun stimulus extérieur ne soit imposé au sujet. Le terme asynchrone englobe cependant ces deux types de systèmes.

5 Méthodologie

Après avoir décrit les différents systèmes qui sont en mesure de créer un moyen de communication entre le cerveau d'un sujet et un système électronique, nous allons à présent nous intéresser aux techniques qu'il faut mettre en œuvre pour rendre le système capable de reconnaître une activité cérébrale spécifique parmi un ensemble d'activités. Les traitements se font généralement en deux parties : tout d'abord, on extrait à partir de l'activité cérébrale du sujet un ensemble de caractéristiques qui sont aptes à rendre compte de l'état mental du sujet. Ensuite on compare ces caractéristiques à un ensemble de données connues afin de déduire l'état mental et ainsi la tâche visée (étape de classification). Cette procédure est très générale et vise à trouver une fonction mathématique qui permet d'inférer la tâche que réalise le sujet en fonction de son activité cérébrale. Ainsi, si on note $c(t) \in \mathcal{C}$ (où \mathcal{C} est l'ensemble des tâches mentales disponibles pour le sujet), l'état mental du sujet à l'instant t et $\mathbf{x}(t)$ l'activité cérébrale mesurée du sujet, on cherche à trouver la fonction f_t telle que :

$$\forall t \in \mathbb{R} \quad f_t(\mathbf{x}(t)) = c(t) \quad (\text{II.1})$$

Le nombre de méthodes utilisées dans le domaine est très important, nous donnerons donc seulement les principes généraux les plus importants qui seront nécessaires à la compréhension de la suite de ce manuscrit. Le lecteur intéressé par une revue relativement exhaustive des méthodes déjà utilisées pourra se tourner vers la multitude de tableaux présents dans [Bashashati et al. \(2007\)](#) et [Lotte et al. \(2007a\)](#). La plupart de ces méthodes a été développé pour des systèmes basés sur l'EEG, cependant elles peuvent sans aucun problème être appliquées à la MEG ou à d'autres systèmes d'acquisition.

5.1 Caractériser une activité cérébrale

Dans un premier temps, on cherche un espace de représentation de l'activité cérébrale qui soit adapté à la classification. En d'autres termes, on veut :

- enlever le bruit : le bruit est constitué des éléments extérieurs (parasites électriques...), des activités cérébrales non-spécifiques aux tâches considérées ainsi que les activités extra-cérébrales du sujet (mouvement des yeux, activités musculaires). Il est assez difficile de quantifier l'importance de ces différents bruits dans l'EEG mais on estime souvent que le rapport signal sur bruit est d'environ -20 dB ([Nuñez, 2005](#)) ;
- réduire la dimension : généralement, le nombre de capteurs en EEG est relativement élevé (entre 20 et 256), par conséquent, on veut également représenter l'essentiel de l'activité EEG qui permettra de reconnaître la tâche mentale réalisée à l'aide d'un nombre de variables le plus petit possible.

Or on a vu dans les sections précédentes qu'une tâche mentale pouvait la plupart du temps être caractérisée selon trois domaines : tout d'abord on sait qu'une tâche cognitive induit des changements dans la densité spectrale de puissance de certains signaux cérébraux (les bandes de puissance correspondant à certaines tâches sont souvent connues en partie) ; ensuite la réalisation d'une tâche implique des zones cérébrales restreintes (inutile donc d'observer l'activité du cerveau dans son ensemble si l'on attend une activation dans une zone précise du cerveau) ; enfin il arrive qu'une tâche soit composée de la succession

de plusieurs activations dans des zones cérébrales et/ou dans des fréquences distinctes. Ces trois domaines, respectivement appelés dans la suite de cette thèse les domaines **fréquentiel**, **spatial**, et **temporel**, doivent être utilisés de manière conjointe pour extraire des caractéristiques pertinentes de l'activité cérébrale.

5.1.1 Prétraitements

5.1.1.1 Spécificités fréquentielles des tâches

Afin de concentrer l'analyse sur les fréquences d'intérêt, on a souvent recours à une transformation temps-fréquence dans le but de réduire les bruits extérieurs. De cette transformation \mathcal{T} résultent les signaux filtrés spécifiques à des bandes de fréquences plus ou moins étroites

$$\tilde{\mathbf{x}}(t, f) = \mathcal{T} \{ \mathbf{x}(t) \}. \quad (\text{II.2})$$

Les systèmes d'acquisition fonctionnent généralement à des fréquences d'échantillonnage entre 100 et 1 KHz. Or les fréquences intéressantes (notamment pour l'imagerie motrice) de l'EEG se situent entre 5 et 35 Hz. On élimine donc la plupart du temps les composantes du signal en-dehors de cette bande (on veut notamment enlever les signaux dus aux alimentations électriques – 50 Hz en France). On peut pour ce faire utiliser des bancs de filtres à réponse impulsionnelle finie ou infinie¹ $\{h_{f_j}(t)\}_{j \in [1..N_f]}$ (N_f est le nombre de bandes de fréquence considéré), auquel cas on a

$$\forall i \in [1..N_s] \quad \forall j \in [1..N_f] \quad \tilde{x}_i(t, f_j) = (h_{f_j} * x_i)(t), \quad (\text{II.3})$$

où N_s est le nombre de capteurs. La conception du filtre $h_{f_j}(t)$ devra alors prendre en compte les contraintes liées à l'application visée ([Oppenheim et Schafer, 1989](#)).

D'autres transformées temps-fréquence ont également été considérées pour nettoyer les signaux, parmi les multiples approches testées, les plus couramment employées actuellement restent la transformée de Fourier à court terme et la transformée en ondelettes ([Mallat, 1999](#)). Ces transformées peuvent s'écrire sous la forme d'un produit scalaire avec un ensemble de fonctions que l'on appelle atomes (vérifiant un ensemble de propriétés comme décrit dans [Mallat \(1999\)](#))

$$\forall i \in [1..N_s] \quad \tilde{x}_i(t, f) = \langle x_i, \varphi_{t,f} \rangle = \int_{-\infty}^{\infty} x_i(u) \varphi_{t,f}(u-t)^* du, \quad (\text{II.4})$$

et qui sont constitués d'un ensemble de fenêtres réelles et symétriques à support fini $g(t) = g(-t)$ modulées à une fréquence spécifique dans le cas des atomes de Fourier

$$\varphi_{t,f}(\tau) = e^{2i\pi f\tau} g(\tau - t), \quad (\text{II.5})$$

ou bien d'une dilatation/translation d'une fonction mère $\psi \in L^2(\mathbb{R})$ dans le cas des ondelettes :

$$\varphi_{t,s}(\tau) = \frac{1}{\sqrt{s}} \psi\left(\frac{\tau - t}{s}\right). \quad (\text{II.6})$$

5.1.1.2 Spécificités spatiales des tâches

Alors que l'on mesure le plus souvent un ensemble de signaux caractérisant l'activité cérébrale globale du cerveau, on sait que les tâches demandées au sujet peuvent être

¹ Notons que l'utilisation d'un filtre à réponse impulsionnelle infinie nécessite moins de coefficients (donc moins de retard) mais risque de modifier la phase du signal, rendant celle-ci impossible à utiliser par la suite. Il est cependant possible de limiter l'effet de la modification en phase en utilisant un filtre à phase minimale.

discriminées en utilisant seulement l'activité de quelques zones cérébrales. Or en raison de la conduction du volume, les mesures acquises sont un mélange de des activités des sources cérébrales. Afin de passer d'une activité global et générale du cerveau à une activité locale et spécifique aux tâches « cible », on utilise des méthodes multivariées de filtrage spatial.

5.1.1.2.1 Sélection de capteurs

Le moyen le plus simple d'effectuer un filtrage spatial est de sélectionner les électrodes. On estime *a priori* l'emplacement des électrodes qui permettront le mieux d'observer une modification de l'activité cérébrale concomitante à la tâche, on observe donc uniquement celle-ci. Cette méthode a l'avantage de la simplicité, néanmoins, elle nécessite la connaissance d'un expert capable de sélectionner quelques électrodes d'intérêt. Cette approche est assez risquée en raison de la variabilité des réponses cérébrales entre individus. En effet, il se peut qu'une électrode intéressante pour un individu ne le soit plus pour un autre. Enfin, cette méthode ne tient absolument pas compte du bruit au niveau des capteurs (on fait l'hypothèse qu'il est homogène sur l'ensemble des capteurs), et ne tient pas compte du fait qu'il vaut quelquefois mieux regarder une activité cérébrale à partir d'un capteur plus éloigné si les signaux qu'il capte son moins bruités que ceux acquis à l'aide d'un capteur « proche » mais qui serait très bruité. La méthode de sélection d'une électrode a cependant été améliorée pour réaliser une sélection automatique des électrodes. Elle a ainsi été appliquée avec succès dans le cas du P300 ([Rakotomamonjy et Guigue, 2008](#)), ou encore dans le cas de l'utilisation des rythmes sensorimoteurs ([Lal et al., 2004](#)).

5.1.1.2.2 Laplacien

De manière plus générale, un filtrage spatial **linéaire** des signaux consiste à chercher un ensemble de L combinaisons linéaires des signaux qui permettent d'obtenir des taux de classification plus importants que dans le cas où on utilise directement les signaux provenant des capteurs. En raison de la linéarité de cette transformation, on peut l'écrire sous la forme

$$\mathbf{s}(t) = W^T \mathbf{x}(t), \quad (\text{II.7})$$

où $\mathbf{x}(t) \in \mathbb{R}^{N_s}$, $\mathbf{s}(t) \in \mathbb{R}^L$ et $W \in \mathbb{R}^{N_s \times L}$. Il existe plusieurs méthodes pour trouver W . L'utilisation d'un filtre laplacien approché (l'approximation de la dérivée seconde consiste à enlever la moyenne des électrodes voisines) a par exemple permis une augmentation des performances dans le cas de l'imagerie motrice ([McFarland et al., 1997](#)). Une telle méthode est à la frontière entre le prétraitement (nettoyer les données) et le filtrage spatial (extraire les activités spécifiques aux tâches mentales « cibles »).

5.1.1.2.3 « Common Spatial Patterns »

Dans le cas où l'on dispose d'un jeu de données d'entraînement pour lequel les tâches théoriques réalisées par le sujet sont connues, on peut se tourner vers les méthodes d'apprentissage supervisées des filtres spatiaux. La plus utilisée dans le contexte des ICM est l'algorithme du CSP (« Common Spatial Pattern ») ([Müller-Gerking et al., 1999](#) et [Blankertz et al., 2008](#)). L'idée de cet algorithme est d'exploiter de la meilleure manière possible les différences spatiales des activations liées à deux tâches distinctes (e.g. imagination de la main droite et imagination de la main gauche). Notons alors Ω l'ensemble des essais qui sont réalisés lors d'une expérience ICM. Ω , de cardinal $|\Omega|$, est composé d'un ensemble de tâches de la classe Ω_+ de cardinal $|\Omega_+|$ et d'un ensemble de tâches Ω_- de cardinal $|\Omega_-|$, tels que $|\Omega_+| + |\Omega_-| = |\Omega|$. Pour chaque réalisation d'une tâche $\omega \in \Omega$ (de

durée L_ω), on définit la matrice de covariance estimée $C_\omega \in \mathbb{R}^{N_s \times N_s}$ des signaux acquis pendant la réalisation de la tâche comme

$$C_\omega = \frac{1}{L_\omega} \sum_{t \in \omega} \mathbf{x}(t) \mathbf{x}(t)^T. \quad (\text{II.8})$$

On peut alors définir la matrice de covariance relative à chacune des tâches comme

$$\begin{aligned} C_+ &= \frac{1}{|\Omega_+|} \sum_{\omega \in \Omega_+} C_\omega, \\ C_- &= \frac{1}{|\Omega_-|} \sum_{\omega \in \Omega_-} C_\omega. \end{aligned} \quad (\text{II.9})$$

L'algorithme CSP cherche alors à diagonaliser conjointement C_+ et C_- , i.e. à trouver $W \in \mathbb{R}^{N_s \times N_s}$ ainsi que $\Lambda_+, \Lambda_- \in \mathbb{R}^{N_s \times N_s}$ diagonales telles que

$$\begin{cases} W^T C_+ W = \Lambda_+ \\ W^T C_- W = \Lambda_- \end{cases}, \quad (\text{II.10})$$

en contraignant de plus W (par normalisation) tel que $\Lambda_+ + \Lambda_- = I$. Mathématiquement, la [condition II.10](#) consiste en une décomposition propre généralisée, i.e. visant à trouver les vecteurs \mathbf{w} tels que

$$C_+ \mathbf{w} = \lambda C_- \mathbf{w}. \quad (\text{II.11})$$

Remarque 1 (DIAGONALISATION SIMULTANÉE DE DEUX MATRICES). *Comme C_- et C_+ sont des combinaisons linéaires de matrices de covariance, elles sont donc symétriques. Par conséquent, elles sont diagonalisables dans la même base W si et seulement si C_- et C_+ commutent (voir le théorème 4.17 dans [Schott \(2005\)](#)).*

L'équation [II.10](#) est satisfaite pour $W = [\mathbf{w}_1 \dots \mathbf{w}_{N_s}]$, matrice des vecteurs \mathbf{w} solutions de [II.11](#). De plus, les éléments diagonaux de Λ_+ et Λ_- sont respectivement $\lambda_{i,+} = \mathbf{w}_i^T C_+ \mathbf{w}_i$ et $\lambda_{i,-} = \mathbf{w}_i^T C_- \mathbf{w}_i$. Notons alors la correspondance $\lambda = \lambda_{i,+} / \lambda_{i,-}$. Afin de comprendre l'effet de cet algorithme, il faut interpréter les valeurs $\lambda_{i,+}$ (respectivement $\lambda_{i,-}$) comme la variance du capteur virtuel créé par projection $s_i(t) = \mathbf{w}_i^T x_i(t)$ lorsque le sujet réalise une tâche de la classe Ω_+ (respectivement Ω_-). Ainsi, lorsque $\lambda_{i,+}$ est proche de un, la variance du capteur virtuel associé est importante dans la condition positive et faible dans la condition négative. Enfin, il faut remarquer que la condition précédente fournit une base commune adaptée aux deux classes puisqu'elle permet une décorrélation des signaux. Le CSP peut aussi être vu en tant que modèle génératif. En effet, si l'on considère le modèle de mélange

$$\mathbf{x} = A\mathbf{s}, \quad \mathbf{s} \sim \mathcal{N}(\mathbf{0}, C_{+,-}), \quad (\text{II.12})$$

où \mathbf{s} suit une loi normale multivariée de moyenne nulle et de matrice de variance-covariance respectivement C_+ et C_- . On a alors un modèle de sources non-stationnaires, et comme montré dans [Pham et Cardoso \(2001\)](#), la diagonalisation simultanée de C_+ et C_- fournit l'estimateur au sens du maximum de vraisemblance de $W = (A^{-1})^T$. Nous reviendrons sur ce point dans la suite de cette thèse dans le cadre de l'extension du CSP au cas de paradigmes multi-classes ([chapitre III](#)). Le lecteur pourra constater que la formulation du CSP donnée ici est différente de la formulation originelle de [Müller-Gerking et al. \(1999\)](#). La vision initiale visait à interpréter le CSP du point de vue de la discrimination. L'algorithme consistait alors à former la matrice d'activités discriminatives $S_d = C_+ - C_-$ et la matrice d'activités communes $S_c = C_+ + C_-$. L'idée était alors de tenter de maximiser la variance discriminante $\mathbf{w}^T S_d \mathbf{w}$ tout en minimisant la variance commune $\mathbf{w}^T S_c \mathbf{w}$. Cette formulation mène alors à l'optimisation du quotient de Rayleigh

$$\max_{\mathbf{w} \in \mathbb{R}^{N_s}} \frac{\mathbf{w}^T S_d \mathbf{w}}{\mathbf{w}^T S_c \mathbf{w}}. \quad (\text{II.13})$$

L'optimisation de ce critère mène alors à nouveau à une décomposition en valeurs propres généralisée. Et on montre (Blankertz et al., 2008) que ces trois formulations sont équivalentes. Cet algorithme fournit un ensemble de N_s vecteurs classés par ordre d'importance. On garde généralement $L = 6$ ou $L = 8$ d'entre eux par paires (premier et dernier vecteurs, deuxième et pénultième...). Plusieurs extensions du CSP ont été développées depuis son introduction (Blankertz et al., 2008), afin notamment de prendre en compte les paradigmes à plus de deux classes, nous y reviendrons dans le chapitre III.

Afin d'illustrer l'effet du CSP, nous avons réalisé une simulation à partir de deux distributions normales de dimension deux et de matrices de covariance différentes. Les signaux originaux ainsi que les signaux après application de l'algorithme CSP sont représentés dans la figure II.12.

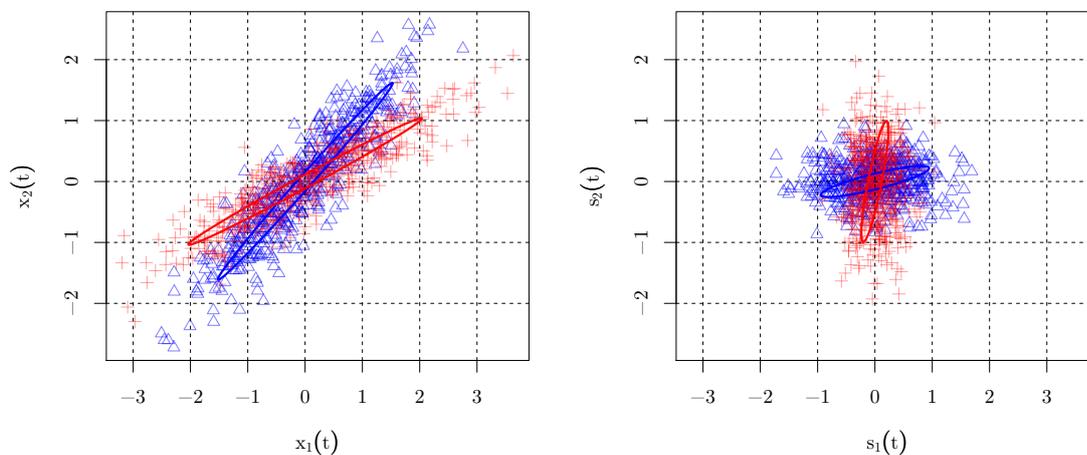


Figure II.12 Exemple simulé à partir de deux lois normales, chacune en dimension deux. A gauche, les tirages (500 points) représentant les valeurs mesurées au niveau des capteurs, et à droite, les activités des sources après application de l'algorithme CSP.

5.1.1.2.4 Séparation de sources

Il est également possible de trouver des pondérations des capteurs dans le cas où l'on n'a pas d'information à propos des tâches réalisées par le sujet (ou bien lorsqu'on ne préfère pas les utiliser). Le problème se pose alors dans les termes suivants : comment trouver des signaux utiles, les sources cérébrales, à partir de mélanges, les signaux mesurés au niveau des capteurs, de ceux-ci. Dans le cas où aucune information n'est connue *a priori*, ni sur les sources, ni sur le processus de mélange, cette situation est appelée Séparation Aveugle de Sources (SAS). La première solution à ce problème a été formalisée au début des années 80 (Hérault et Ans, 1984 et Hérault et al., 1985) en ne faisant qu'une seule hypothèse fondamentale : *l'indépendance mutuelle des sources*. Pour une vue complète et détaillée du problème de la séparation de source et des différentes méthodes qui permettent de la résoudre, on pourra se reporter aux livres récents (Jutten et Comon, 2007 et Comon et Jutten, 2007). Dans le cas de l'EEG, trois principes fondamentaux permettent de restreindre le problème de la SAS (Congedo et al., 2008a) :

- Dans l'hypothèse où les dipôles électriques restent suffisamment fixes dans l'espace, il n'y a pas de retard entre l'activité d'un neurone et la mesure au niveau du scalp ;

- L'approximation quasi-statique est valide, par conséquent on peut négliger les termes de couplage électromagnétiques dans les équations de Maxwell ;
- Pour les fréquences considérées (< 40 Hz), les effets capacitifs sont négligeables.

Ces trois considérations vont dans le sens du principe de superposition et permettent donc d'écrire le lien qui lie les sources cérébrales et les observations au niveau des capteurs sous la forme d'un *mélange linéaire instantané*¹

$$\mathbf{x}(t) = A\mathbf{s}(t). \quad (\text{II.14})$$

En utilisant alors le théorème de Darmois-Skitovic (Darmois, 1947), on montre qu'il est possible de retrouver $\mathbf{s}(t)$ en utilisant uniquement l'hypothèse d'indépendance mutuelle entre les sources uniquement si au plus une des sources suit une densité de probabilité gaussienne. Cependant, les sources estimées ne peuvent représenter les sources initiales qu'à une indétermination d'échelle (si $(A, \mathbf{s}(t))$ est solution, alors $\forall \gamma \in \mathbb{R}$ $(\frac{1}{\gamma}A, \gamma\mathbf{s}(t))$ est aussi solution) et une permutation près. Les méthodes basées sur l'indépendance pour résoudre le problème de la SAS sont des méthodes d'ACI (Analyse en Composantes Indépendantes). Leur principe commun est de proposer une mesure capable de rendre compte de l'indépendance entre les signaux (mesure directe de l'indépendance du type information mutuelle ou bien mesure approchée en utilisant les statistiques d'ordre supérieur par exemple) et de mettre en place une méthode d'optimisation capable de rendre les signaux $\mathbf{s}(t)$ mutuellement indépendants. Parmi les plus connues et les plus utilisées en ICM, on peut citer Infomax (Bell et Sejnowski, 1995) basée sur la maximisation de l'entropie, FastICA (Hyvärinen, 1999) qui vise à maximiser la non-gaussianité des signaux en utilisant une fonction auxiliaire non-linéaire, JADE (Souloumiac et Cardoso, 1991) qui est basée sur la minimisation des cumulants croisés d'ordre quatre.

Si les méthodes présentées jusqu' alors reposent sur l'hypothèse de sources iid (indépendantes et identiquement distribuées), c'est-à-dire qu'elles ne prennent pas en compte les relations temporelles entre les échantillons, plusieurs méthodes ont été mises au point (adressant alors le problème de la séparation semi-aveugles de sources) en levant l'hypothèse d'indépendance, i.e. le premier « i » (les sources sont alors supposées colorées ou temporellement corrélées) ou encore l'hypothèse « id » auquel cas les sources deviennent non-stationnaires. Dans le premier cas, la méthode la plus utilisée dans l'analyse des signaux cérébraux est SOBI (Belouchrani et al., 1997) qui repose sur la diagonalisation conjointe de matrices de corrélation retardées. Le deuxième cas sera traité plus en détail dans le chapitre III.

Les différentes méthodes de séparation de sources ont été utilisées de manière intensive dans l'analyse des signaux cérébraux et plus particulièrement dans le contexte des ICMs. Leur utilité a été montrée pour nettoyer les données (Vigário et al., 2000) ou bien pour augmenter les performances de classification (Naeem et al., 2006). L'un des principaux problèmes de cette méthode est lié à l'interprétation des composantes trouvées : en effet, si dans le cas du CSP, les composantes sont classées et peuvent être facilement sélectionnées, l'étape de sélection des composantes est délicate dans le cas de la séparation de sources et on doit faire appel à des experts ou des critères extérieurs.

¹ par la suite, nous supposons que le nombre de capteurs est égal au nombre de sources cérébrales recherchées afin de simplifier l'exposé. Ce cas est appelé *déterminé*. Notons qu'il est possible aussi de considérer les problèmes des mélanges sous- et sur-déterminés, i.e. respectivement les cas où l'on a plus (resp. moins) de sources que de capteurs. Ces deux cas imposent souvent d'utiliser des méthodes différentes en introduisant de nouveaux *a priori*.

5.1.1.2.5 Retour aux sources cérébrales « réelles »

Certains articles se sont intéressés à l'intérêt d'utiliser les méthodes de problème inverse afin d'effectuer un filtrage spatial. Ces approches sont fondées sur la connaissance des propriétés physiques du milieu de propagation et utilisent le fait que l'on connaît de manière approchée la fonction qui lie l'activité de chaque source neuronale à la mesure qui sera effectuée au niveau des capteurs. Par le principe de superposition exprimé plus haut, on sait que cette relation est linéaire et que, par conséquent, elle peut s'écrire sous la forme matricielle

$$\mathbf{x}(t) = G\mathbf{s}(t). \quad (\text{II.15})$$

On appelle G la matrice de « Lead field » ; elle est caractéristique de la géométrie du problème (position des sources, des capteurs, caractéristiques du milieu de propagation). Dans ce cas, $\mathbf{s}(t)$ est de très grande dimension et l'estimation de $\mathbf{s}(t)$ à partir de G et des mesures au niveau des capteurs $\mathbf{x}(t)$ ne peut être réalisée que par l'introduction de contraintes *a priori*. En effet, lorsque la dimension de l'espace des sources est largement plus grande que la dimension des mesures au niveau des capteurs, le problème est très mal posé et G est non-inversible. Il faut alors employer des méthodes de régularisation afin de calculer $\hat{\mathbf{s}}(t) = H\mathbf{x}(t)$ ou bien $H = f(G)$, estimations respectives des sources ou de la matrice inverse de la matrice de « leadfield ». Parmi les différentes techniques qui ont été utilisées, les plus courantes sont basées sur la minimisation de la norme L_2 qui conduit à la pseudo-inverse de Moore-Penrose ou encore la norme L_1 qui donne une solution plus parcimonieuse. Dans le contexte des ICMs, ces approches restent d'un emploi relativement marginal en raison de la lourdeur des méthodes numériques qu'elles nécessitent et du gain relativement mineur qu'elles engendrent (Congedo et al., 2006, Lotte, 2008, Besserve et al., 2008a et Congedo, 2006).

5.1.1.2.6 Autres méthodes de filtrage spatial

Outre les méthodes générales présentées jusqu'ici, il est également possible d'envisager des approches plus ciblées. Par exemple, xDAWN (Rivet et al., 2009 et Rivet et al., 2008) a été récemment proposée pour extraire les composantes les plus intéressantes (au sens de l'optimisation du rapport signal sur bruit) dans le cas des interfaces basées sur des réponses synchrones telles que le P300. D'autres critères ont également été utilisés, notamment pour débruiter les signaux (Boudet, 2008).

5.1.2 Domaine de représentation des caractéristiques

Une fois les signaux débruités et/ou filtrés, on a alors recours à l'étape qu'on appelle *extraction des caractéristiques*. Le but de cette étape est de trouver une représentation des signaux EEG qui permette de facilement discriminer entre deux tâches distinctes. De même que dans la section précédente, il est impossible de détailler l'ensemble des méthodes qui a été utilisé pour extraire des caractéristiques intéressantes des signaux EEG. Une description plus complète peut être trouvée dans Bashashati et al. (2007). Les principes généraux sont cependant les mêmes que ceux décrits précédemment : on cherche à utiliser en conjonction les caractéristiques temporelles, fréquentielles et spatiales des tâches réalisées en fonction des différents *a priori* dont on dispose. Nous distinguons dans la présentation qui suit les méthodes univariées qui calculent les caractéristiques composante par composante sans se préoccuper des interactions entre composantes ; des méthodes multivariées qui s'intéressent également aux interactions entre les différentes composantes issues du filtrage.

5.1.2.1 Méthodes univariées

Suivant la nature de la réponse considérée, on peut s'intéresser aux caractéristiques des signaux dans le domaine temporel ou bien dans le domaine fréquentiel. Par exemple, dans le cas de la réponse asservie en temps du P300, il suffit de s'intéresser à l'amplitude du signal environ 300 ms après la stimulation. Cependant, on ne peut pas considérer ce type de caractéristiques dans le cas de la modulation des rythmes cérébraux. On doit alors s'intéresser aux caractéristiques spectrales de chaque composante. Pour ce faire, on peut utiliser la variance des signaux filtrés à l'aide d'un filtre passe-bande de Butterworth (Grosse-Wentrup et Buss, 2008), ou encore l'estimation de la densité spectrale de puissance à partir de la transformée de Fourier ou de la transformée en ondelettes. Le logarithme est en général appliqué à l'ensemble de ces estimations afin que les lois résultantes soient plus proche d'une densité gaussienne.

D'autres approches univariées ont également été testées, l'utilisation des coefficients de réflexion d'un modèle autorégressifs (AR) s'est notamment révélée très robuste (Sykacek et al., 2004) dans le cas de l'imagerie motrice. De même que les coefficients d'un modèle AR adaptatif (Schlögl et al., 2005) ont également permis d'obtenir des bons résultats dans une expérience d'imagerie motrice à quatre classes.

Notons enfin l'utilisation marginale de caractéristiques issues de la littérature sur le chaos, par exemple la dimension fractale (Boostani et al., 2007).

5.1.2.2 Méthodes multivariées

Contrairement aux méthodes d'extraction de caractéristiques précédentes qui agissent composante par composante de manière indépendante, il est également possible de considérer des mesures capables de prendre en compte les interactions entre celles-ci. Dans Anderson et al. (1998), les auteurs ont comparé les performances obtenues à partir des coefficients ARs univariés, les matrices d'un modèle AR multivarié, les valeurs propres des matrices de corrélation ainsi que la transformée de Karhunen-Loève des coefficients ARs multivariés. Cette étude mettant en jeu quatre sujets n'a pas permis de montrer la supériorité d'une quelconque méthode comparée aux autres pour l'ensemble des sujets. Ils ont cependant noté des différences de performances significatives sujet par sujet.

Certains auteurs ont également cherché à s'intéresser à la phase des signaux, s'inscrivant dans la continuité des travaux de F. Varela (Rudrauf et al., 2003) qui considérait qu'une grande partie de l'échange d'information entre les signaux se faisait pas des phénomènes relatifs à la phase des signaux (synchronisation en phase). Dans le contexte des ICMs, des mesures de synchronie entre les différentes composantes ont été utilisées avec succès dans Besserve et al. (2008b). Beaucoup d'études ont également utilisé les travaux de J-P. Lachaux et la mesure de synchronie qu'il a appelée « Phase Locking Value » (PLV, mesure du verrouillage en phase). Parmi les différents travaux utilisant la PLV, on peut noter par exemple Brunner et al. (2006) dans le cas de l'imagerie motrice à trois classes. Enfin, il est possible de choisir un domaine de représentation adapté aux signaux EEG en utilisant un dictionnaire d'atomes de base qui vont servir à décomposer le signal. Dans Bénar et al. (2009), cette idée est même développée afin d'adapter la projection sur le dictionnaire essai par essai en utilisant une procédure de vote. Cette approche conduit à une représentation parcimonieuse des signaux.

La lourdeur des calculs que les méthodes multivariées impliquent restent cependant souvent prohibitive dans des systèmes temps réels.

5.1.3 Utilisation des successions d'activations

Pour certaines réponses neurophysiologiques, il peut s'avérer intéressant de considérer, en plus des caractéristiques décrites plus haut qui sont calculées de manière indépendante

d'une fenêtre à l'autre, la **trajectoire** de ces caractéristiques pendant une succession de fenêtres temporelles (e.g. de la préparation et l'intention du mouvement au rebond β dans le cas de l'imagerie motrice). Le modèle le plus simple destiné à prendre en compte de manière naturelle ce type d'enchaînement est le modèle AR qui a déjà été cité plus haut. Mais des modèles plus élaborés ont également été testés dans un but de classification. Les chaînes de Markov Cachées ont été utilisées pour l'analyse des signaux EEG (Flexer et al., 2000 et Zhong et Ghosh, 2002). Ces modèles ont même été introduits dans le contexte des ICMs pour la classification de signaux issus d'expériences d'imagerie motrice (Obermaier et al., 2001 et Obermaier et al., 2003). Ces deux articles sont basés sur un modèle temporel de changement dans les rythmes α et β . L'expérience était basée sur un paradigme d'imagerie motrice synchrone, après filtrage des signaux EEG, différentes caractéristiques étaient calculées (puissance spectrale, coefficients ARs) et fournies en entrée d'un classifieur basé sur une chaîne de Markov. Lors de la période d'apprentissage, un modèle de Markov était appris pour chaque classe. Lors de la période d'évaluation, pour chaque réalisation d'une tâche, on calculait les caractéristiques successives de la tâche et la classe prédite était celle dont le modèle donnait la vraisemblance la plus grande étant données les caractéristiques. Bien que cette méthode de classification se soit avérée légèrement meilleure qu'une méthode linéaire classique, l'utilisation de chaînes de Markov pour la classification des signaux EEG est restée relativement anecdotique. Une exception notable est la thèse de Silvia Chiappa (Chiappa, 2006), dans laquelle la supériorité des chaînes de Markov sur des méthodes plus classiques n'est pas flagrante.

Alors que les chaînes de Markov sont adaptées à des états latents discrets, d'autres auteurs ont également proposé d'utiliser des modèles temporels dans lesquels les variables cachées sont continues (filtres de Kalman). Ainsi, dans Sykacek et al. (2004), un modèle de Kalman est décrit dans un cadre variationnel pour réaliser une interface cerveau-machine adaptative. Bien que ce modèle ait donné d'excellents résultats, l'interprétation des différentes variables cachées et les différents paramètres à fixer ont certainement empêché sa diffusion dans la communauté. Une version plus simple de l'utilisation d'un filtre de Kalman a récemment été proposée (Tsui et al., 2009) pour une expérience sans stimulation ni contrainte temporelle ayant pour but le contrôle d'un robot à partir de commandes motrices.

5.2 Apprentissage et classification

Dans les parties précédentes, nous avons présenté les différentes techniques permettant de passer des signaux cérébraux (mesurés par exemple par EEG) enregistrés en continu à l'aide de N_s capteurs à un ensemble de caractéristiques permettant de résumer un intervalle de temps d'EEG à l'aide d'un nombre restreint de caractéristiques. Nous arrivons donc à l'étape de classification, *i.e* comment décider, à partir des caractéristiques et des connaissances acquises sur le fonctionnement cérébral du sujet, son état mental. Ce problème relève de l'apprentissage artificiel, il a été l'un des domaines scientifiques les plus féconds lors des deux dernières décennies. Notre but ne sera donc bien sûr pas de présenter ce domaine dans son ensemble mais plutôt les aspects importants qui permettront au lecteur de comprendre les techniques employées dans le contexte des ICMs. Pour une présentation plus complète du domaine, le lecteur pourra se diriger vers des ouvrages de référence récents, complets et instructifs Hastie et al. (2001) et Duda et al. (2000).

5.2.1 Présentation générale

Nous ne nous intéressons ici qu'au cas où l'on dispose d'une base d'exemples pour lesquels les caractéristiques ainsi que la tâche mentale associée sont supposées connues. Bien que cette hypothèse puisse apparaître forte dans le cas des expériences d'ICMs, la

plupart des articles du domaine l'ont adoptée. Ainsi, même si l'on n'est pas sûr que le sujet réalise bien la tâche qu'il doit accomplir en théorie, cette incertitude est négligée au risque d'aboutir à des systèmes dont l'apprentissage est très approximatif. Ce type de problème est appelé dans la littérature l'**apprentissage supervisé**.

Les caractéristiques fournies en entrée du système d'apprentissage sont appelées **entrées** ou encore **prédicteurs**, tandis que l'état mental du sujet est la **sortie** ou encore **réponse**. Comme nous l'avons vu dans la description des phénomènes neurophysiologiques, l'état mental du sujet peut être une variable **quantitative** (mesure continue) ou bien **qualitative** (par exemple imagerie du bras gauche ou du bras droit). Le problème de l'apprentissage dans le contexte des ICMs se résume à la formulation suivante qui comprend la phase d'apprentissage et la phase d'évaluation :

- phase d'apprentissage : soit $X \in \mathcal{X}$ vecteur aléatoire (représentant les caractéristiques résumant l'activité cérébrale mesurée sur un intervalle temporel) et $Y \in \mathcal{Y}$ variable aléatoire décrivant l'état mental du sujet sur l'intervalle considéré de distribution jointe $\mathbb{P}(X, Y)$. Durant la phase d'apprentissage, on cherche une fonction $f(X)$ qui permette de prédire Y étant données des valeurs de l'entrée X . D'une manière générale en théorie de l'apprentissage, ce problème s'exprime sous la forme d'un problème d'optimisation par rapport à $f \in \mathcal{F}$ dont la fonction de coût

$$\underbrace{L(Y, f(X))}_{\text{Erreur de prédiction}} + \underbrace{\lambda J(f)}_{\text{Régularisation}} \quad (\text{II.16})$$

doit être minimisée. Cette formulation met en valeur les différentes questions qui sont soulevées par le concept d'apprentissage : il faut tout d'abord choisir l'espace des fonctionnelles \mathcal{F} afin de limiter le domaine de recherche de la solution. Ce choix permet également d'inclure des informations *a priori* très importantes sur la connaissance que l'on peut avoir de la forme de f à attendre. Beaucoup de méthodes sont ainsi basées sur l'optimisation paramétrique dans un ensemble de fonctions f_θ indicées par le paramètre θ (qui peut être multidimensionnel). Cette approche facilite grandement la recherche d'une solution et permet d'obtenir d'excellents résultats lorsque la famille f_θ est adaptée au problème. Il faut également choisir comment quantifier l'erreur de prédiction $L(Y, f(X))$. Ce choix dépend de la nature du problème et notamment du caractère qualitatif ou quantitatif de Y (Vapnik, 2000). Enfin, le choix de la fonction de régularisation $J(f)$ revient en général au choix d'une norme qui visera à pénaliser certaines formes non désirées de f . Par exemple, la fonction $J(f) = \int [f''(x)]^2 dx$ revient à pénaliser les fonctions ayant des changements brutaux. λ permet de moduler l'influence du terme de régularisation par rapport à l'erreur de prédiction.

- phase d'évaluation : pour chaque nouvelle réalisation $\mathbf{x} \in \mathcal{X}$ dont l'état mental est inconnu, on veut prédire $\hat{y} = f(\mathbf{x})$ afin de connaître l'état mental du sujet. Lors de cette phase, on veut que les erreurs de prédiction réalisées par le système de classification soient les moins importantes possibles. La capacité du classifieur à traiter correctement de nouvelles données s'appelle la capacité de « généralisation ». Et les erreurs réalisées sont les « erreurs de généralisation ».

Dans le cas où Y est une variable qualitative (par exemple si l'on cherche à discriminer entre des imaginations motrices distinctes), alors on a un problème de **classification** figure II.13. Si Y est une variable quantitative (prenant donc des valeurs continues), alors on a à faire à un problème de **régression** figure II.13.

Afin d'illustrer ces notions à partir d'un cas simple, considérons le cas où aucune régularisation n'est appliquée, Y est une variable continue et l'erreur de prédiction est l'erreur quadratique $L(Y, f(X)) = (Y - f(X))^2$. La fonction à minimiser devient alors (Hastie et al., 2001) :

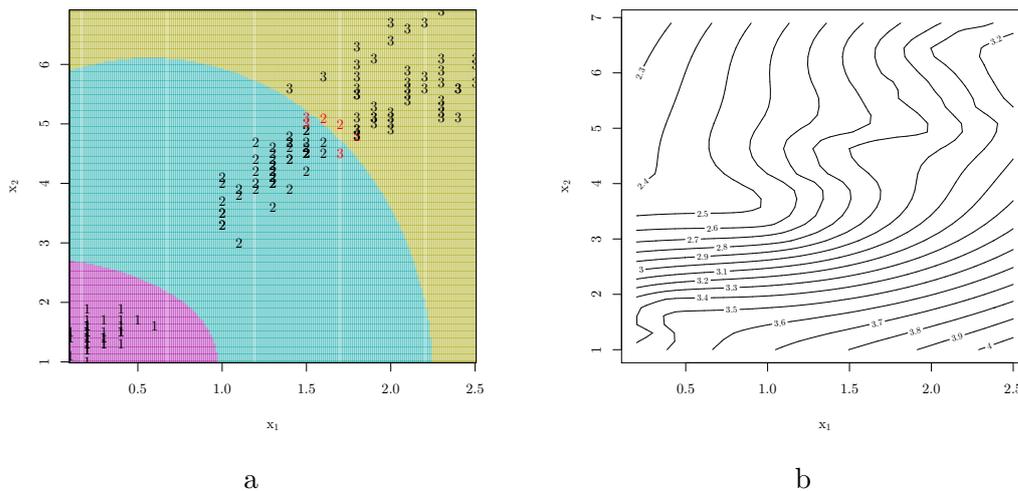


Figure II.13 a) Exemple de classification en dimension deux. Les points correctement classés sont dessinés en noir et les points mal classés sont en rouge. Les frontières de décision entre deux zones différentes sont matérialisées par un changement de couleur. On voit que sur cet exemple, seuls six points sont mal classés. Cet exemple utilise deux dimensions d'un jeu de données libre et largement répandu nommé iris. Pour le charger sous R par exemple, il suffit de taper `data(iris)`. On a classé l'espèce de la fleur (setosa -1-, versicolor -2- ou virginica -3-) en fonction de la largeur (x_1) et de la longueur (x_2) des pétales. b) Exemple de régression en dimension deux. En utilisant le même jeu de données, on a appris une fonction permettant de prédire la largeur des sépales en fonction de la largeur (x_1) et de la longueur (x_2) des pétales.

$$\Xi = \mathbb{E} \{ (Y - f(X))^2 \} = \int (y - f(x))^2 \mathbb{P}(dx, dy) \quad (\text{II.17})$$

qui, en conditionnant par rapport à X , devient

$$\Xi = \mathbb{E}_X \{ \mathbb{E}_{Y/X} \{ (Y - f(X))^2 \mid X \} \}.$$

On voit alors que cette fonction de coût peut être optimisée point par point selon X , d'où

$$f(x) = \arg \min_c \mathbb{E}_{Y/X} \{ (Y - c)^2 \mid X = x \}.$$

Dont la solution est

$$f(x) = \mathbb{E} \{ Y \mid X = x \}. \quad (\text{II.18})$$

Ainsi, $f(x)$ est définie comme l'espérance de Y sachant une réalisation x de X . La méthode des plus proches voisins est une approximation de ce résultat dans laquelle l'espérance est remplacée par une moyenne empirique sur un voisinage plus ou moins étendu de x .

Dans ce contexte, la régression linéaire revient à chercher $f(x)$ sous la forme $f(x) = x^T \beta$ et à limiter \mathcal{F} à ce type de fonctions. Alors en introduisant $f(x)$ dans le [critère II.17](#) et en dérivant par rapport à β , on obtient l'expression classique

$$\beta = (\mathbb{E} \{ XX^T \})^{-1} \mathbb{E} \{ XY \}. \quad (\text{II.19})$$

La suite de cette section est consacrée à la présentation d'un ensemble de méthodes classiques utilisées dans le contexte des ICMs. Pour une présentation détaillée des méthodes qui seront exposées, le lecteur pourra se reporter aux livres précédemment mentionnés.

5.2.2 Analyse discriminante linéaire

Dans le cas où Y est une variable qualitative pouvant prendre deux valeurs, on peut considérer la fonction qui vaut 0 lorsque la prédiction est égale à la valeur de la sortie et 1 lorsque la prédiction diffère de la valeur attendue. La fonction de coût s'écrit alors $\Xi_{\text{LDA}} = \mathbb{E}\{L(Y, f(X))\}$, qui en conditionnant à nouveau selon X et en simplifiant, donne la solution

$$f(x) = \arg \max_{y \in \mathcal{Y}} \mathbb{P}(y \mid X = x). \quad (\text{II.20})$$

Cette règle exprime le fait que la connaissance de $\mathbb{P}(Y \mid X)$ est suffisante pour obtenir une classification optimale. Or d'après la règle de Bayes, en notant $f_k(x)$ la densité conditionnelle de X dans la classe $Y = k$, et π_k la probabilité *a priori* de la classe k ($k \in \{1, 2\}$), on obtient

$$\mathbb{P}(G = k \mid X = x) = \frac{f_k(x)\pi_k}{f_1(x)\pi_1 + f_2(x)\pi_2},$$

ce qui signifie que la connaissance de $f_k(x)$ est suffisante pour une classification optimale. L'analyse discriminante linéaire est alors basée sur l'hypothèse de densités normales multivariées pour chaque $f_k(x)$ de moyennes respectives μ_k . De plus, les deux matrices de covariances sont supposées **identiques**, notées Σ . La règle de classification est alors entièrement déterminée en étudiant la valeur du log-ratio

$$\log \frac{\mathbb{P}(Y = 2 \mid X = x)}{\mathbb{P}(Y = 1 \mid X = x)} = \log \frac{\pi_2}{\pi_1} - \frac{1}{2}(\mu_2 + \mu_1)^T \Sigma^{-1}(\mu_2 - \mu_1) + x^T \Sigma^{-1}(\mu_2 - \mu_1), \quad (\text{II.21})$$

qui est linéaire en x et donne donc une frontière de séparation linéaire entre les deux classes.

Cette technique est très peu coûteuse en calculs et reste de loin la plus utilisée en ICM. Elle est par exemple utilisée dans le cas d'expériences basées sur l'imagerie motrice (Pfurtscheller et Lopes da Silva, 1999), pour extraire des P300 parmi l'activité cérébrale de fond (Bostanov, 2004), dans certaines conditions asynchrones (Scherer et al., 2004). Son principe a été étendu aux cas multi-classes en se ramenant à plusieurs problèmes bi-classes. Une extension Bayésienne très efficace a également été mise au point (Hoffmann et al., 2007, Rivet et al., 2009 et Rivet et al., 2008).

Notons enfin que l'analyse discriminante quadratique (Hastie et al., 2001) résulte du même modèle mais les frontières paraboliques résultent de l'hypothèse de deux matrices de variance-covariance différentes (la figure II.14 montre les frontières obtenues pour un problème identique en utilisant les méthodes de l'analyse discriminante linéaire et quadratique). Des versions incluant des régularisations ont également été utilisées en ICM (Müller et al., 2004).

5.2.3 Régression logistique

La régression logistique est une méthode de classification (Y est une variable qualitative prenant les valeurs $[1..K]$) basées sur la modélisation des probabilités *a posteriori* de chacune des K classes. On modélise celles-ci comme des fonctions linéaires en x . En imposant alors à chacune de ces probabilités de prendre une valeur entre 0 et 1 et de sommer à 1, on obtient le modèle

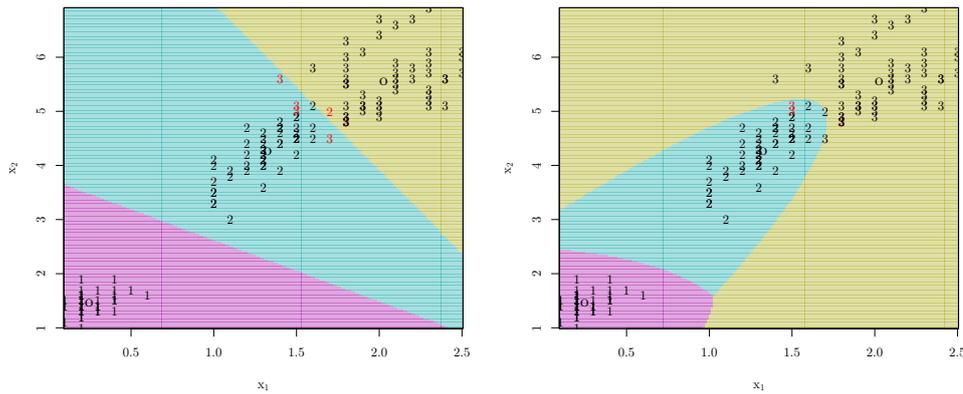


Figure II.14 Frontières obtenues pour le même problème que précédemment avec les méthodes de l'analyse discriminante linéaire et l'analyse discriminante quadratique. Les centroïdes de chaque classe sont représentés par un cercle. Les frontières entre les trois différentes classes sont matérialisées par un changement de couleur.

$$\begin{aligned}
 \log \frac{\mathbb{P}(Y = 1 | X = x)}{\mathbb{P}(Y = K | X = x)} &= \beta_{10} + \beta_1^T x \\
 \log \frac{\mathbb{P}(Y = 2 | X = x)}{\mathbb{P}(Y = K | X = x)} &= \beta_{20} + \beta_2^T x \\
 &\vdots \\
 \log \frac{\mathbb{P}(Y = K - 1 | X = x)}{\mathbb{P}(Y = K | X = x)} &= \beta_{(K-1)0} + \beta_{K-1}^T x
 \end{aligned} \tag{II.22}$$

qui revient de manière équivalente à fixer $\mathbb{P}(Y = K | X = x) = \frac{1}{1 + \sum_{l=1}^{K-1} \exp(\beta_{l0} + \beta_l^T x)}$ et pour tout k appartenant à l'intervalle $[1 .. K - 1]$, $\mathbb{P}(Y = k | X = x) = \frac{\exp(\beta_{k0} + \beta_k^T x)}{1 + \sum_{l=1}^{K-1} \exp(\beta_{l0} + \beta_l^T x)}$.

Nous ne détaillerons pas l'estimation des paramètres β ; celle-ci est généralement basée sur le principe du maximum de vraisemblance et repose sur une méthode d'optimisation numérique itérative (Newton-Raphson). La régression logistique a été pendant longtemps assez peu employée dans le contexte des ICMS ; elle s'est pourtant révélée assez robuste en présence de valeurs aberrantes et a donné des résultats au moins égaux à ceux obtenus avec l'analyse discriminante linéaire (Tomioka et al., 2007 et Grosse-Wentrup et Buss, 2008). Nous exposerons dans l'annexe A les avantages d'une régression logistique régularisée et nous étudierons plus en détail l'influence du paramètre de régularisation λ .

5.2.4 « Support Vector Machine »

On peut également s'intéresser au domaine d'optimisation \mathcal{F} de la fonction de prédiction f . La manière la plus efficace et la plus féconde d'aborder ce problème est d'utiliser un noyau symétrique défini positif et l'espace de fonctions correspondant appelé RKHS (reproducing kernel Hilbert Space – espace de Hilbert à noyau auto-reproduisant). Cette correspondance est donnée par le théorème de Moore-Aronszajn (Aronszajn, 1950).

Théorème 1 (THÉORÈME DE MOORE-ARONSZAJN). *Soit K un noyau symétrique, défini positif sur $\mathcal{E} \times \mathcal{E}$, alors il existe un unique espace de Hilbert de fonctions définies sur \mathcal{E} pour lequel K est un noyau reproduisant.*

Soit donc K un noyau symétrique défini-positif et \mathcal{H}_K le RKHS associé. Si l'on suppose que K admet une décomposition en fonctions propres

$$K(x_1, x_2) = \sum_{i=1}^{\infty} \gamma_i \varphi_i(x_1) \varphi_i(x_2),$$

les éléments f de \mathcal{H}_K s'écrivent alors

$$f(x) = \sum_{i=1}^{\infty} c_i \varphi_i(x).$$

En prenant dans II.16 $J(f) = \|f\|_{\mathcal{H}_K}^2 \stackrel{\text{def}}{=} \sum_{i=1}^{\infty} c_i^2 / \gamma_i$, et étant donné un ensemble d'apprentissage $\{(x_i, y_i)\}_{i \in [1..N]}$, alors le critère II.16 s'écrit

$$\min_{f \in \mathcal{H}_K} \left[\sum_{i=1}^N L(y_i, f(x_i)) + \lambda \|f\|_{\mathcal{H}_K}^2 \right] \quad (\text{II.23})$$

ou encore

$$\min_{\{c_j\}_1^{\infty}} \left[\sum_{i=1}^N L(y_i, \sum_{j=1}^{\infty} c_j \varphi_j(x_i)) + \lambda \sum_{j=1}^{\infty} c_j^2 / \gamma_j \right]. \quad (\text{II.24})$$

Toute la magie de cette représentation intervient ici puisqu'on peut alors montrer (Wahba, 1990) que ce critère d'optimisation en dimension infinie admet une solution de la forme

$$f(x) = \sum_{i=1}^N a_i K(x, x_i), \quad (\text{II.25})$$

et ainsi, d'après la propriété « auto-reproduisante » de K

$$J(f) = \sum_{i=1}^N \sum_{j=1}^N a_i a_j K(x_i, x_j). \quad (\text{II.26})$$

En écrivant alors K sous la forme d'une matrice $N \times N$, on obtient alors le critère dont l'optimisation peut se faire en dimension finie sous la forme matricielle

$$\min_{\mathbf{a}} [L(\mathbf{y}, K\mathbf{a}) + \lambda \mathbf{a}^T K \mathbf{a}], \quad (\text{II.27})$$

où $\mathbf{y} = [y_1, y_2, \dots, y_N]^T$ et $(K)_{(i,j)} = K(x_i, x_j)$. Des algorithmes d'optimisation simples peuvent alors être utilisés pour ce problème.

Les SVMs (Support Vector Machine) sont l'application directe de ce principe. Dans le cas d'un problème de classification à deux classes, la fonction $f(x)$ s'écrit $f(x) = a_0 + \sum_{i=1}^N a_i K(x, x_i)$ dont les paramètres sont choisis de sorte à minimiser

$$\min_{a_0, \mathbf{a}} \left\{ \sum_{i=1}^N [1 - y_i f(x_i)]_+ + \lambda \mathbf{a}^T K \mathbf{a} \right\} \quad (\text{II.28})$$

où $y_i \in \{-1, 1\}$ et $[\cdot]_+$ est la partie positive. Dans le cas d'un noyau linéaire $K(x_i, x_j) = x_i^T x_j$, cette méthode s'interprète comme un problème d'optimisation visant à maximiser l'écart entre les représentants des classes et un hyperplan. La dénomination « vecteur support » provient alors du fait que, en raison de la nature de la forme de l'erreur de prédiction (sorte de norme L_1), la méthode d'optimisation va choisir un certain nombre de coefficients a_i non nuls qui feront de certains représentants x_i des vecteurs « supportant » la frontière (« vecteurs supports »). Plus les classes seront disjointes, plus le nombre de

coefficients a_i nuls sera grand et plus le nombre de vecteurs supports des hyperplans séparant sera faible. La contrepartie de la puissance de la théorie présentée ici réside souvent dans la difficulté de choisir un noyau adapté aux données qu'on cherche à classifier. Certaines méthodes ont été proposées pour pallier ce problème (Rakotomamonjy et al., 2008).

Les SVMs sont couramment utilisés pour la classification dans les ICMs. Par exemple dans le cas d'ICMs synchrones basées sur des potentiels évoqués (Garrett et al., 2003 et Rakotomamonjy et Guigue, 2008), des SVMs linéaires ont permis d'obtenir d'excellents résultats. Des SVMs à noyaux Gaussiens ont également été utilisés avec succès (Kaper et al., 2004).

5.2.5 Évaluation et sélection de modèles

Parmi les différentes méthodes présentées précédemment, un des critères fondamental est la capacité de généralisation : comment va se comporter la méthode lors d'une session réalisée deux jours après l'entraînement des différents paramètres de la méthode ? Cette question est bien sûr valable pour la méthode de classification choisie mais également pour toutes les étapes de traitement fondées sur les données (et non sur un modèle). La méthode des k -plus proches voisins permet d'illustrer de manière simple le comportement de l'erreur de généralisation. Si on suppose le modèle bruité $Y = X + \varepsilon$ avec $\varepsilon \sim \mathcal{N}(0, \sigma^2)$, alors l'erreur de prédiction pour un nouveau point x_0 peut être décomposée en (Hastie et al., 2001)

$$\begin{aligned} \text{EPE}_k(x_0) &= \mathbb{E} \left\{ (Y - \hat{f}_k(x_0))^2 \mid X = x_0 \right\} \\ &= \sigma^2 + \underbrace{\left[f(x_0) - \frac{1}{k} \sum_{l=1}^k f(x_{(l)}) \right]^2}_{\text{Biais}^2(\hat{f}_k(x_0))} + \underbrace{\frac{\sigma^2}{k}}_{\text{Var}(\hat{f}_k(x_0))} \end{aligned} \quad (\text{II.29})$$

où $x_{(l)}$ indique la séquence des plus proches voisins de x_0 . Cette expression montre bien le compromis en jeu entre le biais du modèle et sa variance. Lorsque k augmente, la complexité du modèle diminue, on obtient une variance faible mais un biais important. A l'opposé, lorsque k diminue, la complexité du modèle devient grande, le biais est faible mais la variance est grande. Cette illustration à partir de la méthode des k -plus proches voisins est en fait générale lorsque l'on choisit parmi un ensemble de modèles de complexités variées. La figure II.15 résume de manière schématique le comportement de l'erreur de prédiction en fonction de la complexité des modèles appris.

Il existe plusieurs façons de sélectionner un modèle approprié parmi un ensemble de modèles ainsi que d'évaluer l'erreur de généralisation une fois celui-ci choisi. La faible quantité de données en ICM fait de la validation croisée l'une des seules utilisées pour sélectionner et évaluer l'erreur de prédiction (voir Hastie et al. (2001) pour une description de cette méthode). Mais notons également que l'évaluation d'un algorithme se fait quelquefois à partir de ce qu'on appelle le transfert de session à session : on utilise les données enregistrées un certain jour pour réaliser le calibrage des méthodes, on teste ensuite en utilisant les données d'un jour différent. Cette méthode s'avère très proche du fonctionnement réel d'une ICM, c'est pourquoi elle fait l'objet de beaucoup d'attention (Krauledat et al., 2008).

Notons enfin que l'une des principales difficultés de la classification dans le contexte des ICMs relève de ce qui est connu sous le terme de « malédiction de la dimension » (Bellman, 1961). Cette expression rend compte du fait que, dans un hypercube unitaire de dimension p , la portion r qu'occupe un hypercube et l'arête de cet hypercube $e_p(r)$ sont liées par la relation $e_p(r) = r^{1/p}$. Ainsi, pour occuper 10% du volume global de l'hypercube unitaire,

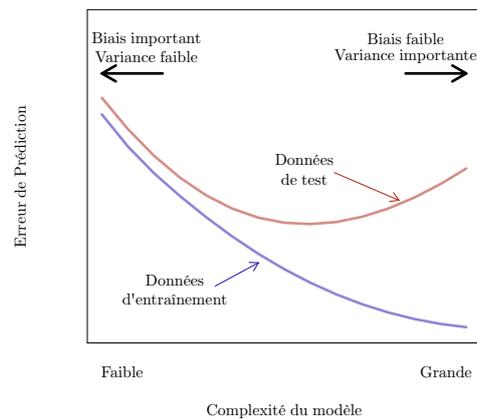


Figure II.15 Erreur de prédiction sur les données d'entraînement et de test en fonction de la complexité du modèle utilisé.

un petit hypercube d'arêtes de longueur 0.1 est suffisant en dimension $p = 2$ tandis qu'il faut une longueur 0.63 en dimension 10. Cet exemple simple montre que l'apprentissage devient un problème très rapidement compliqué lorsque les dimensions augmentent.

6 Enjeux actuels et problématiques de la thèse

Si de nombreuses évolutions ont été réalisées depuis leur apparition, il apparaît néanmoins que les ICMs sont assez loin d'être assez robustes pour que leur développement dans le grand public soit rendu possible. Le nombre de malades atteints de maladies motrices neuro-dégénératives graves ayant pu bénéficier d'un tel système pour améliorer leur niveau de vie reste extrêmement limité. Bien sûr le rapport signal sur bruit des systèmes d'acquisition EEG actuels, qui sont les plus répandus dans la conception des ICMs, est relativement faible et il est souvent difficile d'interpréter correctement les signaux. Cependant, imaginons que puisse être mis au point un système d'acquisition beaucoup plus performant, permettant de mesurer avec précision l'activité de chaque zone cérébrale avec une résolution temporelle proche de la milliseconde, pourrions-nous alors concevoir des ICMs parfaites, décryptant chacune de nos pensées sans délai ? Rien n'est moins sûr, le fonctionnement du cerveau nous est encore relativement opaque et l'analyse des signaux qu'il produit est aujourd'hui confrontée à plusieurs défis : tout d'abord la réalisation d'une tâche mentale n'apparaît pas toujours stable dans le temps et la connaissance des mécanismes neurophysiologiques à la base de certaines tâches d'imagerie motrice par exemple n'est pas suffisamment avancée pour que soit élucidé l'ensemble des mécanismes cérébraux qui lui sont associés ; de plus, le cerveau est un organe en constante évolution et la connaissance parfaite d'un mécanisme neuronal ne saura être suffisante face aux modifications inhérentes à l'apprentissage du sujet. L'analyse des méthodes ayant permis à leurs auteurs de remporter les différentes « BCI competitions » révèle une supériorité assez écrasante pour les méthodes les plus simples, la plupart du temps linéaires. Ces observations nous incitent à considérer le futur des ICMs dans deux sens distincts ; tout d'abord les collaborations étroites entre les neurophysiologistes et les personnes traitant les signaux doivent conduire à la connaissance plus approfondie des méthodes adaptées au fonctionnement cérébral, ensuite, les systèmes ICMs du futur devront prendre en compte les possibles phénomènes d'apprentissage pouvant s'opérer lors de l'interaction entre le sujet et le système d'analyse. Le système devra donc être capable de s'adapter aux changements du sujet.

La première piste est la raison d'être de cette thèse : tenter de comprendre les mécanismes neuronaux en jeu lors de la réalisation de tâches mentales déterminées, et proposer une méthode qui soit adaptée à cette réalisation. Cette approche passe par l'utilisation conjointe des trois domaines possibles d'analyse des tâches déjà cités : les domaines fréquentiel, spatial et temporel. Bien que la littérature foisonne d'une multitude de travaux concernant l'utilisation séparée d'un ou deux de ces trois domaines conjointement, l'utilisation des trois à la fois n'a pas été si courante. Le développement vers une interface adaptative sera également abordé à la fin de cette thèse.

7 Conclusion

Si l'objectif de ce chapitre n'était pas de passer en revue l'ensemble des systèmes existants, nous espérons cependant avoir donné au lecteur les connaissances suffisantes pour comprendre les enjeux de la conception des ICMs. Nous avons tout d'abord vu que les systèmes d'acquisition de l'activité cérébrale avaient des fonctionnements très distincts et parfois complémentaires. Le but de l'interface à concevoir apparaît donc de manière primordiale lors du choix du système d'acquisition. Nous avons ensuite expliqué les différents phénomènes neurophysiologiques actuellement utilisés pour la conception d'ICMs. Cette liste n'était pas exhaustive mais nous avons essayé de présenter de quelle manière le choix du phénomène neurophysiologique influait sur les possibilités que peut offrir une ICM. Nous insistons donc à nouveau sur la nécessité d'adapter le phénomène neurophysiologique au but de l'interface. Ensuite nous avons proposé une classification nouvelle des ICMs en différents groupes selon le type d'interaction qu'elles engendrent avec le sujet. A notre connaissance, cette classification est originale et nous semble lever un certain nombre de confusions sur les différents adjectifs employés jusqu'alors dans la littérature. La partie méthodologique nous a enfin permis de présenter les axes clés qui seront développés tout au long de cette thèse. Nous avons préféré donner les bases au lecteur lui permettant de comprendre la suite de la thèse plutôt que tenter une présentation exhaustive des différentes méthodes qui ont été utilisées dans le domaine (qui approche aujourd'hui le millier d'articles publiés dans des journaux internationaux).

8 Références du chapitre

- Anderson, C. W., Stolz, E. A. et Shamsunder, S. (1998). Multivariate autoregressive models for classification of spontaneous electroencephalographic signals during mental tasks. *IEEE Trans. Biomed. Eng.*, 45(3):277–286.
- Aronszajn, N. (1950). Theory of reproducing kernels. *Trans. Am. Math. Soc.*, 68(3):337–404.
- Bashashati, A., Fatourehchi, M., Ward, R. K. et Birch, G. E. (2007). A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals. *J. Neural Eng.*, 4(2):R32–R57.
- Bell, A. J. et Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159.
- Bellman, R. (1961). *Adaptive Control Processes*. Princeton University Press.
- Belouchrani, A., Abed-Meraim, K., Cardoso, J.-F. et Moulines, E. (1997). A blind source separation technique using second-order statistics. *IEEE Trans. Signal Process.*, 45(2):434–444.
- Benabid, A. L., Pollak, P., Gervason, C., Hoffmann, D. et Gao, D. M. et al. (1991). Long-term suppression of tremor by chronic stimulation of the ventral intermediate thalamic nucleus. *Lancet*, 337(8738):403–406.
- Berger, H. (1929). Über das elektroencephalogramm des menschen. *Archiv für Psychiatrie und Nervenkrankheiten*, 87(1):527–570.

- Besserve, M., Jerbi, K., Laurent, F., Baillet, S. et Martinerie, J. et al. (2007). Classification methods for ongoing EEG and MEG signals. *Biol. Res.*, 40(4):415–437.
- Besserve, M., Martinerie, J. et Garnero, L. (2008a). Non-invasive classification of cortical activities for brain computer interface: a variable selection approach. Dans *Proceedings of the 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*.
- Besserve, M., Philippe, M., Florence, G., Laurent, F. et Garnero, L. et al. (2008b). Prediction of performance level during a cognitive task from ongoing EEG oscillatory activities. *Clin. Neurophysiol.*, 119(4):897–908.
- Bidet-Caulet, A., Fischer, C., Besle, J., Aguera, P.-E. et Giard, M.-H. et al. (2007). Effects of selective attention on the electrophysiological representation of concurrent sounds in the human auditory cortex. *J. Neurosci.*, 27(35):9252–9261.
- Birbaumer, N. (2006b). Breaking the silence: Brain-computer interfaces (BCI) for communication and motor control. *Psychophysiology*, 43(6):517–532.
- Birbaumer, N. et Cohen, L. G. (2007). Brain-computer-interfaces (BCI): Communication and restoration of movement in paralysis. *J. Physiol.*, 579(3):621–636.
- Birbaumer, N., Hinterberger, T., Kübler, A. et Neumann, N. (2003). The thought-translation device (TTD): neurobehavioral mechanisms and clinical outcome. *IEEE Trans. Neural Syst. Rehabil. Eng.*, 11(2):120–123.
- Birbaumer, N., Kübler, A., Ghanayim, N., Hinterberger, T. et Perelmouter, J. et al. (2000). The thought translation device (TTD) for completely paralyzed patients. *IEEE Trans. Rehabil. Eng.*, 8(2):190–193.
- Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M. et Müller, K.-R. (2008). Optimizing spatial filters for robust EEG Single-Trial analysis. *IEEE Signal Process. Mag.*, 25(1):41–56.
- Boostani, R., Graimann, B., Moradi, M. et Pfurtscheller, G. (2007). A comparison approach toward finding the best feature and classifier in cue-based BCI. *Med. Biol. Eng. Comput.*, 45(4):403–412.
- Bostanov, V. (2004). BCI competition 2003–data sets Ib and IIb: feature extraction from event-related brain potentials with the continuous wavelet transform and the t-value scalogram. *IEEE Trans. Biomed. Eng.*, 51(6):1057–1061.
- Boudet, S. (2008). *Filtrage d’artefacts par analyse multicomposante de l’électroencéphalogramme de patients épileptiques*. Thèse de doctorat, Université de Lille 1.
- Brunner, C., Scherer, R., Graimann, B., Supp, G. et Pfurtscheller, G. (2006). Online control of a brain-computer interface using phase synchronization. *IEEE Trans. Biomed. Eng.*, 53(12 Pt 1):2501–2506.
- Bénar, C. G., Papadopoulo, T., Torrèsani, B. et Clerc, M. (2009). Consensus matching pursuit for multi-trial EEG signals.. *J Neurosci Methods*, 180(1):161–170.
- Chiappa, S. (2006). *Analysis and Classification of EEG Signals using Probabilistic Models for Brain Computer Interfaces*. Thèse de doctorat, École Polytechnique Fédérale de Lausanne.
- Congedo, M. (2006). Subspace projection filters for real-time brain electromagnetic imaging. *IEEE Trans. Biomed. Eng.*, 53(8):1624–1634.
- Congedo, M., Gouy-Pailler, C. et Jutten, C. (2008a). On the blind source separation of human electroencephalogram by approximate joint diagonalization of second order statistics. *Clin. Neurophysiol.*, 119(12):2677–2686.
- Congedo, M., Lotte, F. et Lécuyer, A. (2006). Classification of movement intention by spatially filtered electromagnetic inverse solutions. *Phys. Med. Biol.*, 51(8):1971–1989.
- Coyle, S., Ward, T., Markham, C. et McDarby, G. (2004). On the suitability of near-infrared (NIR) systems for next-generation brain-computer interfaces.. *Physiol. Meas.*, 25(4):815–822.

- Cunnington, R., Iansek, R., Bradshaw, J. L. et Phillips, J. G. (1996). Movement-related potentials associated with movement preparation and motor imagery. *Exp. Brain Res.*, 111(3):429–436.
- Curran, E. A. et Stokes, M. J. (2003). Learning to control brain activity: a review of the production and control of EEG components for driving brain-computer interface (BCI) systems. *Brain Cogn.*, 51(3):326–336.
- Curran, E., Sykacek, P., Stokes, M., Roberts, S. J. et Penny, W. et al. (2004). Cognitive tasks for driving a brain-computer interfacing system: a pilot study. *IEEE Trans. Neural Syst. Rehabil. Eng.*, 12(1):48–54.
- Darmois, G. (1947). Analyse des liaisons de probabilité. Dans *Proc. Internat. Statist. Conferences*, pages 231–240.
- Decety, J. (1996a). The neurophysiological basis of motor imagery. *Behav. Brain Res.*, 77(1-2):45–52.
- Donchin, E., Spencer, K. M. et Wijesinghe, R. (2000). The mental prosthesis: assessing the speed of a P300-based brain-computer interface. *IEEE Trans. Rehabil. Eng.*, 8(2):174–179.
- Duda, R. O., Hart, P. E. et Stork, D. G. (2000). *Pattern Classification*. Wiley-Interscience Publication.
- Farwell, L. A. et Donchin, E. (1988). Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalogr. Clin. Neurophysiol.*, 70(6):510–523.
- Flexer, A., Sykacek, P., Rezek, I. et Dorffner, G. (2000). Using hidden markov models to build an automatic, continuous and probabilistic sleep stager. Dans Sykacek, P., éditeur, *Proc. IEEE-INNS-ENNS International Joint Conference on Neural Networks IJCNN 2000*, pages 627–631 vol.3.
- Gao, X., Xu, D., Cheng, M. et Gao, S. (2003). A BCI-based environmental controller for the motion-disabled. *IEEE Trans. Neural Syst. Rehabil. Eng.*, 11(2):137–140.
- Garrett, D., Peterson, D. A., Anderson, C. W. et Thaut, M. H. (2003). Comparison of linear, nonlinear, and feature selection methods for EEG signal classification. *IEEE Trans. Neural Syst. Rehabil. Eng.*, 11(2):141–144.
- Graimann, B., Huggins, J. E., Levine, S. P. et Pfurtscheller, G. (2004). Toward a direct brain interface based on human subdural recordings and wavelet-packet analysis. *IEEE Trans. Biomed. Eng.*, 51(6):954–962.
- Grosse-Wentrup, M. et Buss, M. (2008). Multiclass common spatial patterns and information theoretic feature extraction. *IEEE Trans. Biomed. Eng.*, 55(8):1991–2000.
- Hastie, T., Tibshirani, R. et Friedman, J. H. (2001). *The Elements of Statistical Learning*. Springer.
- Herrmann, C. S. (2001). Human EEG responses to 1-100 Hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena. *Exp. Brain Res.*, 137(3-4):346–353.
- Hill, N. J., Lal, T. N., Schröder, M., Hinterberger, T. et Wilhelm, B. et al. (2006). Classifying EEG and ECoG signals without subject training for fast BCI implementation: comparison of nonparalyzed and completely paralyzed subjects. *IEEE Trans. Neural Syst. Rehabil. Eng.*, 14(2):183–186.
- Hinterberger, T., Veit, R., Wilhelm, B., Weiskopf, N. et Vatine, J.-J. et al. (2005). Neuronal mechanisms underlying control of a brain-computer interface. *Eur. J. Neurosci.*, 21(11):3169–3181.
- Hinterberger, T., Weiskopf, N., Veit, R., Wilhelm, B. et Betta, E. et al. (2004). An EEG-driven brain-computer interface combined with functional magnetic resonance imaging (fMRI). *IEEE Trans. Biomed. Eng.*, 51(6):971–974.
- Hochberg, L. R. et Donoghue, J. P. (2006). Sensors for brain-computer interfaces. *IEEE Eng. Med. Biol. Mag.*, 25(5):32–38.

- Hochberg, L. R., Serruya, M. D., Friehs, G. M., Mukand, J. A. et Saleh, M. et al. (2006). Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature*, 442(7099):164–171.
- Hoffmann, U. (2007). *Bayesian machine learning applied in a brain-computer interface for disabled users*. Thèse de doctorat, EPFL.
- Hoffmann, U., Vesin, J.-M., Ebrahimi, T. et Diserens, K. (2007). An efficient P300-based brain-computer interface for disabled subjects. *J. Neurosci. Methods*, 167(1):115–125.
- Hyvärinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neural Netw.*, 10(3):626–634.
- Hérault, J. et Ans, B. (1984). Réseau de neurones à synapses modifiables : décodage de messages sensoriels composites par apprentissage non supervisé et permanent. *C. R. Acad. Sc. Paris*, 299(40):525–528.
- Hérault, J., Jutten, C. et Ans, B. (1985). Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé. Dans *Proceedings of the 10th Workshop Traitement du signal et ses applications (GRETSI)*.
- Jasper, H. H. (1941). *Epilepsy and Cerebral Localization*, chapitre Electroencephalography. Springfield.
- Jeannerod, M. (1995). Mental imagery in the motor context. *Neuropsychologia*, 33(11):1419–1432.
- Jeannerod, M. (1994). The representing brain: Neural correlates of motor intention and imagery. *Behav. Brain Sci.*, 17(2):187–245.
- Kalcher, J., Flotzinger, D., Neuper, C., Gölly, S. et Pfurtscheller, G. (1996). Graz brain-computer interface II: towards communication between humans and computers based on online classification of three different EEG patterns. *Med. Biol. Eng. Comput.*, 34(5):382–388.
- Kaper, M., Meinicke, P., Grossekhoefer, U., Lingner, T. et Ritter, H. (2004). BCI competition 2003–data set IIb: support vector machines for the P300 speller paradigm. *IEEE Trans. Biomed. Eng.*, 51(6):1073–1076.
- Kauhanen, L., Nykopp, T., Lehtonen, J., Jylänki, P. et Heikkonen, J. et al. (2006). EEG and MEG brain-computer interface for tetraplegic patients. *IEEE Trans. Neural Syst. Rehabil. Eng.*, 14(2):190–193.
- Krauledat, M. (2008). *Analysis of Nonstationarities in EEG Signals for Improving Brain-Computer Interface Performance*. Thèse de doctorat, Technical University of Berlin.
- Krauledat, M., Tangermann, M., Blankertz, B. et Müller, K.-R. (2008). Towards zero training for brain-computer interfacing. *PLoS ONE*, 3(8):e2967.
- Krusienski, D. J., Sellers, E. W., Cabestaing, F., Bayouhdh, S. et McFarland, D. J. et al. (2006). A comparison of classification techniques for the p300 speller. *Journal of Neural Engineering*, 3(4):299–305.
- Kübler, A. et Neumann, N. (2005). Brain-computer interfaces—the key for the conscious brain locked into a paralyzed body. *Prog. Brain Res.*, 150(1):513–525.
- Lachaux, J.-P., Jerbi, K., Bertrand, O., Minotti, L. et Hoffmann, D. et al. (2007). Braintv: a novel approach for online mapping of human brain functions. *Biol Res*, 40(4):401–413.
- Lal, T. N., Schröder, M., Hinterberger, T., Weston, J. et Bogdan, M. et al. (2004). Support vector channel selection in BCI. *IEEE Trans. Biomed. Eng.*, 51(6):1003–1010.
- Lebedev, M. A. et Nicolelis, M. A. L. (2006). Brain-machine interfaces: past, present and future. *Trends Neurosci.*, 29(9):536–546.
- Leuthardt, E. C., Miller, K. J., Schalk, G., Rao, R. P. N. et Ojemann, J. G. (2006). Electroencephalography-based brain computer interface—the Seattle experience. *IEEE Trans. Neural Syst. Rehabil. Eng.*, 14(2):194–198.

- Leuthardt, E. C., Schalk, G., Wolpaw, J. R., Ojemann, J. G. et Moran, D. W. (2004). A brain-computer interface using electrocorticographic signals in humans. *J. Neural Eng.*, 1(2):63–71.
- Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature*, 453(7197):869–878.
- Logothetis, N. K., Pauls, J., Augath, M., Trinath, T. et Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412(7):150–157.
- Lopes da Silva, F. H. (2006). Event-related neural activities: what about phase?. *Prog. Brain Res.*, 159(1):3–17.
- Lotte, F. (2008). *Study of electroencephalographic signal processing and classification techniques towards the use of brain-computer interfaces in virtual reality applications*. Thèse de doctorat, INSA de Rennes.
- Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F. et Arnaldi, B. (2007a). A review of classification algorithms for EEG-based brain-computer interfaces. *J. Neural Eng.*, 4(2):R1–R13.
- Lotze, M. et Halsband, U. (2006). Motor imagery. *J. Physiol. Paris*, 99(4-6):386–395.
- Müller, K.-R., Krauledat, M., Dornhege, G., Curio, G. et Blankertz, B. (2004). Machine learning techniques for brain-computer interfaces. *Biomedizinische Technik*, 49(1):11–22.
- Mallat, S. (1999). *A Wavelet Tour of Signal Processing, Second Edition (Wavelet Analysis & Its Applications)*. Academic Press.
- Mason, S., Kronegg, J., Huggins, J., Fatourehchi, M. et Schlögl, A. (2006). Evaluating the performance of self-paced brain computer interface technology. Rapport technique, Neil Squire Society, Vancouver, Canada.
- Mason, S. G. et Birch, G. E. (2000). A brain-controlled switch for asynchronous control applications. *IEEE Trans. Biomed. Eng.*, 47(10):1297–1307.
- McFarland, D. J., McCane, L. M., David, S. V. et Wolpaw, J. R. (1997). Spatial filter selection for EEG-based communication. *Electroencephalogr. Clin. Neurophysiol.*, 103(3):386–394.
- Mellinger, J., Schalk, G., Braun, C., Preissl, H. et Rosenstiel, W. et al. (2007). An MEG-based brain-computer interface (BCI). *Neuroimage*, 36(3):581–593.
- Michelon, P., Vettel, J. M. et Zacks, J. M. (2006). Lateral somatotopic organization during imagined and prepared movements. *J. Neurophysiol.*, 95(2):811–822.
- Müller-Gerking, J., Pfurtscheller, G. et Flyvbjerg, H. (1999). Designing optimal spatial filters for single-trial EEG classification in a movement task. *Clin. Neurophysiol.*, 110(5):787–798.
- Müller-Putz, G. R., Scherer, R., Neuper, C. et Pfurtscheller, G. (2006). Steady-state somatosensory evoked potentials: suitable brain signals for brain-computer interfaces?. *IEEE Trans. Neural Syst. Rehabil. Eng.*, 14(1):30–37.
- Naeem, M., Brunner, C., Leeb, R., Graimann, B. et Pfurtscheller, G. (2006). Separability of four-class motor imagery data using independent components analysis. *J. Neural Eng.*, 3(3):208–216.
- Nicolelis, M. A. (2001). Actions from thoughts. *Nature*, 409(6818):403–407.
- Nuñez, P. L. (2005). *Electric Fields of the Brain: The Neurophysics of EEG*. Oxford Univ Pr (Txt).
- Obermaier, B., Müller, G. R. et Pfurtscheller, G. (2003). "Virtual keyboard" controlled by spontaneous EEG activity. *IEEE Trans. Neural Syst. Rehabil. Eng.*, 11(4):422–426.
- Obermaier, B., Neuper, C., Guger, C. et Pfurtscheller, G. (2001). Information transfer rate in a five-classes brain-computer interface. *IEEE Trans. Neural Syst. Rehabil. Eng.*, 9(3):283–288.
- Ogawa, S., Lee, T. M., Nayak, A. S. et Glynn, P. (1990). Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magn. Reson. Med.*, 14(1):68–78.

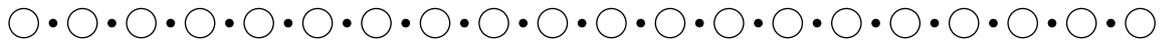
- Oppenheim, A. V. et Schafer, R. W. (1989). *Discrete-Time Signal Processing*. Prentice Hall.
- Pfurtscheller, G., Kalcher, J., Neuper, C., Flotzinger, D. et Pregenzer, M. (1996). On-line EEG classification during externally-paced hand movements using a neural network-based classifier. *Electroencephalogr. Clin. Neurophysiol.*, 99(5):416–425.
- Pfurtscheller, G. et Lopes da Silva, F. H. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clin. Neurophysiol.*, 110(11):1842–1857.
- Pfurtscheller, G. et Neuper, C. (2006). Chapter 28: Future prospects of ERD/ERS in the context of brain-computer interface (BCI) developments. *Prog. Brain Res.*, 159(1):433–437.
- Pfurtscheller, G. et Neuper, C. (2001). Motor imagery and direct brain-computer communication. *Proc. IEEE*, 89(7):1123–1134.
- Pfurtscheller, G., Neuper, C., Müller, G. R., Obermaier, B. et Krausz, G. et al. (2003). Graz-BCI: state of the art and clinical applications. *IEEE Trans. Neural Syst. Rehabil. Eng.*, 11(2):177–180.
- Pham, D.-T. et Cardoso, J.-F. (2001). Blind separation of instantaneous mixtures of nonstationary sources. *IEEE Trans. Signal Process.*, 49(9):1837–1848.
- Rakotomamonjy, A., Bach, F., Canu, S. et Grandvalet, Y. (2008). SimpleMKL. *J. Mach. Learn. Res.*, 9(1):2491–2521.
- Rakotomamonjy, A. et Guigue, V. (2008). BCI competition III: dataset II- ensemble of SVMs for BCI P300 speller. *IEEE Trans. Biomed. Eng.*, 55(3):1147–1154.
- Ramsey, N. F., van de Heuvel, M. P., Kho, K. H. et Leijten, F. S. S. (2006). Towards human BCI applications based on cognitive brain systems: an investigation of neural signals recorded from the dorsolateral prefrontal cortex. *IEEE Trans. Neural Syst. Rehabil. Eng.*, 14(2):214–217.
- Regan, D. (1966). Some characteristics of average steady-state and transient responses evoked by modulated light. *Electroencephalogr. Clin. Neurophysiol.*, 20(3):238–248.
- Rivet, B., Souloumiac, A., Attina, V. et Gibert, G. (2009). xDAWN algorithm to enhance evoked potentials: Application to brain computer interface. *IEEE Trans. Biomed. Eng.*, 56(8):2035–2043. A paraître.
- Rivet, B., Souloumiac, A., Gibert, G. et Attina, V. (2008). P300 speller brain-computer interface: Enhancement of P300 evoked potential by spatial filters. Dans *Proceedings of the 16th European Signal Processing Conference (EUSIPCO-2008), EURASIP*.
- Rovati, L., Salvatori, G., Bulf, L. et Fonda, S. (2007). Optical and electrical recording of the neural activity evoked by graded contrast visual stimulus. *Biomed. Eng. Online*, 6(1):28.
- Rudrauf, D., Lutz, A., Cosmelli, D., Lachaux, J.-P. et Quyen, M. L. V. (2003). From autopoiesis to neurophenomenology: Francisco Varela’s exploration of the biophysics of being. *Biol. Res.*, 36(1):27–65.
- Ryu, K. et Myung, R. (2005). Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic. *Int. J. Ind. Ergon.*, 2005(35):991–1009.
- Scherer, R., Müller, G. R., Neuper, C., Graimann, B. et Pfurtscheller, G. (2004). An asynchronously controlled EEG-based virtual keyboard: improvement of the spelling rate. *IEEE Trans. Biomed. Eng.*, 51(6):979–984.
- Schlögl, A., Lee, F., Bischof, H. et Pfurtscheller, G. (2005). Characterization of four-class motor imagery EEG data for the BCI-competition 2005. *J. Neural Eng.*, 2(4):L14–L22.
- Schott, J. R. (2005). *Matrix Analysis for Statistics*. Wiley-Interscience, Hoboken, USA, Deuxième édition.
- Sitaram, R., Zhang, H., Guan, C., Thulasidas, M. et Hoshi, Y. et al. (2007). Temporal classification of multichannel near-infrared spectroscopy signals of motor imagery for developing a brain-computer interface. *Neuroimage*, 34(4):1416–1427.

- Souloumiac, A. et Cardoso, J.-F. (1991). Comparaisons de methodes de separation de sources. Dans *Proceedings of the GRETSI (conférence sur le traitement des signaux et des images)*.
- Sykacek, P., Roberts, S. J. et Stokes, M. (2004). Adaptive BCI based on variational bayesian Kalman filtering: an empirical evaluation. *IEEE Trans. Biomed. Eng.*, 51(5):719–727.
- Tomioka, R., Aihara, K. et Müller, K.-R. (2007). Logistic regression for single trial EEG classification. Dans Schölkopf, B., Platt, J. et Hoffman, T., éditeurs, *Advances in Neural Information Processing Systems 19*, pages 1377–1384. MIT Press, Cambridge, MA.
- Tsui, C., Gan, J. et Roberts, S. (2009). A self-paced brain-computer interface for controlling a robot simulator: an online event labelling paradigm and an extended Kalman filter based algorithm for online training. *Med. Biol. Eng. Comput.*, 47(3):257–265.
- Vapnik, V. N. (2000). *The Nature of Statistical Learning Theory*. Springer.
- Vaughan, T. M., Heetderks, W. J., Trejo, L. J., Rymer, W. Z. et Weinrich, M. et al. (2003). Brain-computer interface technology: a review of the second international meeting. *IEEE Trans. Neural Syst. Rehabil. Eng.*, 11(2):94–109.
- Vaughan, T. M. et Wolpaw, J. R. (2006). The third international meeting on brain-computer interface technology: making a difference. *IEEE Trans. Neural Syst. Rehabil. Eng.*, 14(2):126–127.
- Vaughan, T. M., Wolpaw, J. R. et Donchin, E. (1996). EEG-based communication: prospects and problems. *IEEE Trans. Rehabil. Eng.*, 4(4):425–430.
- Vigário, R., Särelä, J., Jousmäki, V., Hämäläinen, M. et Oja, E. (2000). Independent component approach to the analysis of EEG and MEG recordings. *IEEE Trans. Biomed. Eng.*, 47(5):589–593.
- Wahba, G. (1990). *Spline Models for Observational Data*. SIAM.
- Watson, J. (1913). Psychology as the behaviorist views it. *Psychological Review*, 20(1):158–177.
- Weiskopf, N., Mathiak, K., Bock, S. W., Scharnowski, F. et Veit, R. et al. (2004). Principles of a brain-computer interface (BCI) based on real-time functional magnetic resonance imaging (fMRI). *IEEE Trans. Biomed. Eng.*, 51(6):966–970.
- Wilson, J., Felton, E., Garell, P., Schalk, G. et Williams, J. (2006). ECoG factors underlying multimodal control of a brain-computer interface. *IEEE Trans. Neural Syst. Rehabil. Eng.*, 14(2):246–250.
- Wolpaw, J. R., Birbaumer, N., Heetderks, W. J., McFarland, D. J. et Peckham, P. H. et al. (2000). Brain-computer interface technology: a review of the first international meeting. *IEEE Trans. Rehabil. Eng.*, 8(2):164–173.
- Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G. et Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clin. Neurophysiol.*, 113(6):767–791.
- Wolpaw, J. R., Loeb, G. E., Allison, B. Z., Donchin, E. et do Nascimento, O. F. et al. (2006). BCI meeting 2005–workshop on signals and recording methods. *IEEE Trans. Neural Syst. Rehabil. Eng.*, 14(2):138–141.
- Wolpaw, J. R. et McFarland, D. J. (2004). Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *Proc. Natl. Acad. Sci. U. S. A.*, 101(51):17849–17854.
- Wolpaw, J. R., McFarland, D. J., Neat, G. W. et Forneris, C. A. (1991). An EEG-based brain-computer interface for cursor control. *Electroencephalogr. Clin. Neurophysiol.*, 78(3):252–259.
- Zhong, S. et Ghosh, J. (2002). HMMs and coupled HMMs for multi-channel EEG classification. Dans *Proc. IEEE Int. Joint Conf. on Neural Networks*, pages 1154–1159, Honolulu, Hawaii.

- Comon, P. et Jutten, C., éditeurs (2007). *Séparation de sources - Volume 1 : Concepts de base et analyse en composantes indépendantes* Hermès - Lavoisier.
- Jutten, C. et Comon, P., éditeurs (2007). *Séparation de sources - Volume 2 : Au-delà de l'aveugle et applications* Hermès - Lavoisier.
- Niedermeyer, E. et Lopes da Silva, F. H., éditeurs (2004). *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields* Lippincott Williams & Wilkins.
- Zani, A. et Proverbio, A., éditeurs (2002). *The Cognitive Electrophysiology of Mind and Brain* Academic Press.

Chapitre III

FILTRAGE SPATIAL : EXTRACTION DE SOURCES NON-STATIONNAIRES



1	Introduction	56
2	Non-stationnarité et imagerie motrice	58
3	Séparation de sources non-stationnaires	63
3.1	Formulation théorique	63
3.2	« Common Spatial Pattern »	64
3.3	Utilisation de la diversité entre classes : DCA	65
3.4	Utilisation des variations d'énergie dans une tâche : DCA-MS	66
4	Paradigme expérimental	67
5	Méthode	68
5.1	Phase d'entraînement	68
5.2	Phase de test	69
5.3	Validation croisée	69
5.4	Transfert inter-session	69
6	Résultats	70
6.1	Validation croisée	70
6.2	Transfert session à session	72
6.3	Comparaison avec la « BCI Competition 2008 »	74
6.4	Artefact numérique ou information réelle	75
7	Utilisation du domaine fréquentiel	75
8	Discussion	77
9	Conclusion	79
10	Références du chapitre	79

CE CHAPITRE traite des méthodes de filtrage spatial pour l'extraction de composantes intéressantes dans les Interfaces Cerveau-Machine (ICMs). Bien que nous traitions en particulier dans ce chapitre d'un exemple d'application consacré à l'imagerie motrice et à un protocole synchrone basé sur l'électroencéphalographie (EEG), la méthode présentée ici est applicable à beaucoup d'autres paradigmes et se généralise également parfaitement à d'autres systèmes d'acquisition. L'importance du filtrage spatial dans les ICMs a été entrevue dans le [chapitre II](#). Nous avons présenté deux méthodes principales fondées respectivement sur un algorithme appelé « Common Spatial Pattern » (CSP) et sur les algorithmes de résolution du problème de séparation aveugle de sources (SAS). Nous montrons dans ce chapitre que ces deux méthodes peuvent être formulées selon le même cadre théorique en utilisant le concept d'extraction de sources non-stationnaires. Ce cadre théorique basé sur le principe du maximum de vraisemblance nous conduit à proposer un cadre commun aux algorithmes CSPs ainsi qu'à la séparation de sources non-stationnaires ; et

également à proposer une extension basée sur des principes neurophysiologiques adaptés aux successions d'activations et de désactivations des sources cérébrales présentes dans la réalisation de tâches cérébrales cognitives telles que l'imagerie motrice.

Les performances des méthodes présentées dans ce chapitre sont évaluées à partir d'un jeu de données dans lequel neuf sujets ont pris part à une expérience ICM basée sur l'imagerie motrice. Outre le nombre de sujets qui est relativement important pour une étude quantitative des performances, ce jeu de données a également l'avantage d'être très homogène entre les différents sujets et contient des enregistrements qui ont été réalisés à plusieurs jours d'intervalle. Ainsi la performance des méthodes du point de vue du transfert de session à session peut être évaluée, en addition de la méthode plus classique fondée sur la validation croisée qui permet d'avoir une estimation intéressante de l'erreur de généralisation.

L'originalité de ce chapitre repose sur trois points distincts : tout d'abord la formulation dans un cadre commun des algorithmes CSPs ainsi que des algorithmes de séparation de sources non-stationnaires permet de mieux comprendre leur fonctionnement respectif et permet d'envisager rigoureusement l'application des méthodes de diagonalisation conjointe comme l'extension naturelle du principe du CSP au cas de paradigmes basés sur strictement plus de deux classes ; de plus cette formulation nous permet de proposer une extension naturelle des méthodes utilisées jusqu'à présent afin de mieux prendre en compte les phénomènes neurophysiologiques en jeu lors des tâches cognitives telles que l'imagerie motrice ; enfin si l'évaluation quantitative de l'extension du CSP du point de vue du transfert session à session dans le cas où le nombre de classes est strictement supérieur à deux avait déjà été étudiée, il n'en est pas de même pour la séparation basée sur la diagonalisation conjointe ni bien sûr pour l'extension proposée.

1 Introduction

Le but des Interfaces Cerveau-Machines (ICMs) est d'établir un moyen de communication direct entre les intentions du sujet et un système électronique (Wolpaw et al., 2002). Dans un contexte médical, les ICMs sont conçus pour fournir aux personnes souffrant de graves handicaps moteurs (souvent conséquences de maladies neuro-dégénératives graves) un moyen de restaurer la communication et la mobilité (Birbaumer et Cohen, 2007 et Nijholt et al., 2008). Leur conception nécessite de mesurer l'activité du système nerveux central, généralement à l'aide de l'électroencéphalographie (EEG), pour ensuite convertir cette activité en une commande pour le système électronique. Bien que chaque pensée de l'utilisateur ne puisse être décodée en raison de la complexité du cerveau, il est possible de proposer au sujet un petit sous-ensemble de tâches mentales, suffisamment bien connues d'un point de vue neurophysiologique, qui peut être reconnu parmi l'ensemble de l'activité cérébrale du sujet et d'associer à chacune de ces tâches mentales une commande spécifique (Pfurtscheller et al., 2006). L'un des principes le plus connu et le plus utilisé consiste à utiliser l'organisation somatotopique¹ des aires corticales motrices (Pfurtscheller et Neuper, 2001). Lorsqu'un sujet imagine un mouvement d'une certaine partie de son corps, une activité spatialement localisée dans le cerveau peut être détectée et ainsi associée à une commande pour le système électronique (Pfurtscheller et al., 2006). Mais le faible Rapport Signal sur Bruit (RSB) et la faible résolution spatiale des données recueillies rend relativement difficile l'analyse directe et l'interprétation de l'activité électrique du cerveau. Ceci peut rendre le contrôle des systèmes électroniques quelque peu approximatif. De plus, les ICMs doivent tenir compte de l'importante variabilité qui existe entre les différents sujets et qui oblige les méthodes à être adaptées à chaque sujet.

¹ « Somatotopique » signifie que chaque partie du corps est représentée de manière unique dans une certaine zone du cerveau. On peut donc dire qu'il existe une bijection entre la représentation corticale des zones du corps et ces zones proprement dites.

Afin de réduire l'impact de tels obstacles, un point crucial de l'analyse des signaux EEG consiste à transformer les signaux acquis à partir d'un grand nombre de capteurs mesurant une activité globale et ambiguë en un petit nombre de **composantes** caractérisant uniquement l'activité cérébrale relative à la tâche que l'on cherche à mesurer. Cette étape porte le nom de **filtrage spatial**. L'algorithme le plus répandu pour réaliser cette étape s'appelle le « Common Spatial Pattern » (CSP). Il a été introduit dans la communauté dans le contexte de la discrimination entre deux tâches mentales d'imagerie motrice (Müller-Gerking et al., 1999) et a montré au fil des années des performances très compétitives (et souvent supérieures) par rapport à d'autres algorithmes utilisés dans le même but (Lotte et al., 2007a). Étant données deux matrices de covariance estimées correspondant à deux tâches distinctes (par exemple imagerie motrice de la main gauche ou de la main droite), l'algorithme CSP cherche les filtres spatiaux linéaires qui maximisent le rapport de variance entre les deux classes. Depuis son introduction, plusieurs améliorations du CSP ont été proposées. On peut citer par exemple les versions locale, parcimonieuse et spectrale qui ont permis des améliorations substantielles des taux de classification dans certains cas précis (Dornhege et al., 2006, Lemm et al., 2005 et Wang et Zheng, 2008). Le lecteur pourra trouver une description complète des différents types de CSPs dans Blankertz et al. (2008).

Mais parallèlement au développement des algorithmes CSPs, une autre approche prometteuse s'est développée pour extraire des composantes spécifiques aux tâches mentales utilisées en ICM. Cette approche est basée sur l'interprétation du problème du filtrage spatial en termes de Séparation Aveugle de Sources (SAS) (Jutten et Hérault, 1991). De manière générale, le problème de la SAS fait référence aux situations où un ensemble de capteurs mesure un mélange de signaux issus de l'activité de sources inobservables. Le problème de l'extraction de l'activité des sources à partir des signaux mesurés a été abordé de différentes manières en faisant de multiples hypothèses concernant les propriétés des sources ou le processus de mixage. Par exemple dans Jutten et Hérault (1991), l'analyse en composantes indépendantes (ACI), a été introduite en utilisant deux hypothèses fondamentales : tout d'abord il existe une relation linéaire liant les signaux générés par les sources et les signaux observés, ensuite, les sources sont supposées statistiquement indépendantes et non-Gaussiennes (en réalité il faut qu'au plus une seule source ait une densité de probabilité Gaussienne (Darmois, 1947)). Si l'hypothèse non-Gaussienne s'est avérée très efficace pour enlever les artefacts oculaires des signaux EEG (Vigário et al., 2000), d'autres approches basées sur la décorrélation avec des matrices de corrélation décalées (Belouchrani et al., 1997) ou sur la non-stationnarité se sont révélées plus adaptées à certains types de signaux naturels (Congedo et al., 2008a, Parra et Spence, 2001 et Matsuoka et al., 1995).

Comme résumé dans Congedo et al. (2008a), la non-stationnarité fournit un cadre théorique idéal pour analyser les signaux EEG provenant d'expériences d'ICMs. Ce cadre s'appuie sur les travaux théoriques de Pham (2001) et sur des algorithmes pratiques extrêmement efficaces. Comme montré dans Pham (2001) et Pham et Cardoso (2001), une hypothèse générique fondée sur la non-stationnarité est suffisante pour séparer un mélange linéaire instantané de sources non-stationnaires. Les auteurs ont montré, en utilisant le principe du maximum de vraisemblance, que les statistiques du second ordre étaient suffisantes pour extraire des sources non-stationnaires Gaussiennes, tandis que les statistiques d'ordre supérieur devaient être utilisées dans le cas de l'hypothèse non-Gaussienne (cf Congedo et al. (2008a) pour plus de détails concernant les hypothèses statistiques associées à différents algorithmes de séparation). Comme initialement montré dans Pham (2001), le problème de la séparation de sources non-stationnaires peut être résolu à partir des statistiques du second ordre en utilisant des algorithmes de Diagonalisation Conjointe Approchée (DCA). La DCA a d'abord été utilisée dans le contexte des ICMs multiclassées par Dornhege et al. (2004) afin de trouver des filtres spatiaux efficaces. C'est seulement dans Grosse-Wentrup et Buss (2008) que la connexion entre le problème de la SAS et cette

approche a été exhibée. Cette méthode a permis d'obtenir des performances supérieures à une méthode basée sur une extension multi-classes du CSP.

Dans ce chapitre, nous abordons en détail l'utilisation du cadre de la séparation de sources non-stationnaires dans le contexte des ICMs. Tout d'abord, nous mettons en exergue les connexions existant entre les algorithmes CSPs, la séparation de sources non-stationnaires et la DCA en utilisant le principe du maximum de vraisemblance. Nous proposons alors une nouvelle méthode fondée sur la DCA et adaptée aux connaissances neurophysiologiques de la réalisation des tâches mentales utilisées en ICM. Les performances du CSP, de l'approche classique fondée sur la DCA ainsi que de notre extension sont alors comparées en utilisant à la fois la validation croisée et également le transfert session-à-session (Krauledat et al., 2008).

La section 2 présente les arguments neurophysiologiques en faveur de l'utilisation de la non-stationnarité dans le cas particulier de l'imagerie motrice. Nous formulons alors différentes hypothèses concernant les sources cérébrales impliquées dans l'imagerie motrice et en déduisons différents modèles dans la section 3. Nous présentons ensuite le paradigme expérimental pour l'évaluation des performances de chacune des méthodes dans la section 4 et nous rentrerons dans les détails de l'application du cadre de ce chapitre aux ICMs dans la section 5. Les résultats sont présentés dans la section 6 et nous nous intéressons à une extension de la méthode visant à utiliser de manière plus précise les diversités du domaine fréquentiel dans la section 7.

2 Non-stationnarité et imagerie motrice

Si l'on connaît relativement bien les caractéristiques spatiales et fréquentielles ainsi que la trajectoire temporelle des activités cérébrales liées à l'exécution d'un mouvement (Kandel et Schwartz, 1985 et Jankelowitz et Colebatch, 2002), les connaissances concernant l'imagination des mouvements s'avèrent plus controversées. Dans une optique de discrimination, il est important de distinguer entre trois types d'implications des structures cérébrales (Michelon et al., 2006):

- implication : la structure cérébrale est impliquée dans la tâche motrice,
- implication somatotopique latérale : la structure cérébrale est impliquée et une demi-moitié du corps se projette de manière préférentielle dans l'hémisphère correspondant au même côté du corps (latéralisation somatotopique ipsilatérale) ou bien dans l'hémisphère opposé (latéralisation somatotopique controlatérale),
- implication somatotopique homunculaire : on parle d'implication somatotopique homunculaire lorsque des parties voisines d'une moitié du corps se projettent dans des structures cérébrales proches (voir figure III.1 pour un exemple de structures présentant une telle organisation, le cortex moteur primaire (M1) et le cortex sensoriel primaire (S1)).

Ces trois types de structures sont liés de la manière suivante : l'implication « somatotopique homunculaire » implique « l'implication somatotopique latérale » qui implique « l'implication ». Sans rentrer dans les différents débats qui sont encore vifs afin de déterminer les différences d'implication de multiples zones cérébrales dans le cas de la réalisation d'un mouvement, de la préparation d'un mouvement ou encore de l'imagination d'un mouvement (Michelon et al., 2006, Ehrsson et al., 2003, Lotze et Halsband, 2006, Decety, 1996a et Decety, 1996b), résumons les découvertes les plus importantes pour la discrimination des tâches d'imagerie motrice. Tout d'abord, on connaît les structures impliquées de manière primordiale dans les fonctions motrices, il s'agit du cortex moteur primaire (M1), de l'aire motrice supplémentaire (SMA), du cortex pré-moteur (PMC), des zones motrices

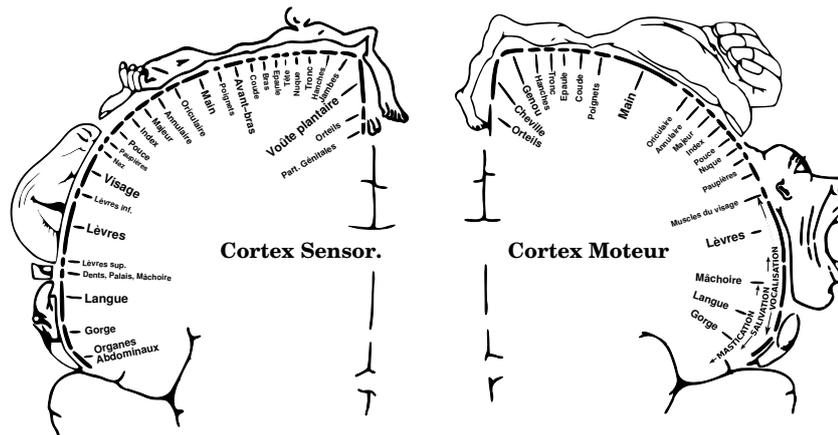


Figure III.1 Homunculus de Penfield. La partie motrice, à droite, montre l'organisation de la commande motrice des muscles du corps humain. La partie sensorielle, à gauche, montre l'organisation des zones corticales somesthésiques.

cingulaires (CZ) et du cervelet. La plupart de ces zones est impliquée de manière somatotopique homunculaire dans le cas de la réalisation de mouvements. L'exécution d'un mouvement volontaire peut être divisée en deux phases : la phase de préparation et la phase d'exécution. Les BPs (Bereitschaft Potentials, potentiels d'alerte), qui caractérisent le niveau d'alerte du sujet, peuvent être observés très tôt dans le SMA, quelquefois jusqu'à 2 s avant l'exécution du mouvement (Jankelowitz et Colebatch, 2002 et Jahanshahi et Hallett, 2003). Ensuite on peut les observer au niveau du PMC. Environ 400 ms avant l'exécution du mouvement, les BPs commencent à apparaître dans le cortex moteur primaire, principalement controlatéralement à la zone d'exécution du mouvement. La déviation liée aux BPs atteint alors son pic au moment de l'exécution du mouvement. Cependant, les implications des structures cérébrales motrices dans les cas de l'imagination ou la préparation de mouvements a donné lieu à des résultats souvent contradictoires. Les études les plus récentes (employant des techniques pour lesquelles la résolution spatiale était meilleure) montrent que les implications dans le cas de l'imagination de mouvements sont analogues à celles observées dans le cas de mouvements réels (l'amplitude des signaux observés seraient néanmoins beaucoup moins importante) (Michelon et al., 2006 et Ehrsson et al., 2003). De plus, comme initialement supposé dans Decety (1996b), la différence entre l'imagerie motrice et la préparation de mouvements serait plus une différence de degré d'implication et de variabilité que relevant de différences entre les structures impliquées. Notons enfin l'implication courante du cortex visuel dans le cas de l'imagerie motrice, il est en effet difficile de contrôler comment le sujet réalise sa tâche. En particulier, les zones impliquées seront assez différentes entre une imagerie subjective (le sujet est impliqué directement dans la tâche imaginée) et une imagerie visuelle (le sujet a visualisé la réalisation de la tâche par une personne extérieure).

Dans l'objectif de discriminer entre différents types d'imagerie motrice, on s'intéresse habituellement aux variations de potentiels électriques liées au mouvement (Wang et al., 2004b). Les phénomènes les plus utilisés sont les ERD/ERS (qui signifient respectivement Event-Related Desynchronization et Event-Related Synchronization) dans les bandes de fréquence μ et β , ainsi que les Bereitschaft Potentials (BP), qui caractérisent une diminution lente et progressive des potentiels qui est généralement associée à une plus grande propension des cellules neuronales à émettre un potentiel d'action. Ces différents phénomènes sont générés par ce que nous appellerons des « sources cérébrales », qui sont positionnées dans les différentes structures citées précédemment (SMA, PMC, M1). Par conséquent, si l'on avait accès à l'activité de ces sources de manière directe, on pourrait par exemple observer les positions des sources actives pendant une imagination de mouvement dans une structure homunculaire afin de savoir si le sujet est en train d'imaginer un

mouvement du pied, de la main ou de n'importe quelle autre partie du corps. Cependant, l'activité de ces sources ne peut être directement observée à l'aide de l'EEG, qui enregistre un mélange (dû à la propagation des signaux électriques des sources jusqu'aux capteurs) de l'ensemble de l'activité du cerveau.

Afin de tenter de retrouver l'activité des sources liées uniquement à la tâche à partir des observations au niveau des capteurs, on utilise un filtrage spatial afin de trouver les poids adéquats à appliquer aux capteurs. Illustrons son principe en considérant une tâche simple d'imagerie motrice impliquant l'imagination d'un mouvement de la main gauche et l'imagination d'un mouvement d'un pied. Les tâches sont réalisées par le sujet après qu'un stimulus lui a été présenté pour lui indiquer la tâche à réaliser. A partir des données EEG, $\mathbf{x}(t) \in \mathbb{R}^{N_s}$, enregistrées en continu à une fréquence de 100 Hz, on applique différents poids à chacun des capteurs afin d'obtenir un ensemble de sources (voir II.7) :

$$\mathbf{s}(t) = W^T \mathbf{x}(t). \quad (\text{III.1})$$

La manière qui a permis de trouver $W \in \mathbb{R}^{N_s \times L}$ sera détaillée plus tard. En plus d'être positionnées à un endroit précis du cerveau qui dépend de la tâche, chaque source cérébrale est également impliquée avec une fréquence caractéristique (par exemple μ et β). C'est pourquoi nous allons observer les signaux dans des bandes de fréquence spécifiques. Pour ce faire, calculons ensuite la transformée de $\mathbf{s}(t)$ par un banc de filtres de Butterworth de largeur 2 Hz, disjoints et recouvrant la bande de fréquence entre 7 et 29 Hz (soit un total de onze filtres). On obtient alors $\tilde{s}_i(t, f_j)$ l'activité de la source i dans la bande de fréquence j . On est alors intéressé par l'évolution de la variance de ces sources au cours de la réalisation d'une tâche d'imagerie motrice et accessoirement par l'évolution comparée, entre deux tâches distinctes, de ces variances. Comme chaque tâche dure environ 5 secondes, considérons $f_S^{\{L, F\}}(i, j, t)$ la densité de probabilité de l'énergie de la source numéro i à la fréquence j et à l'instant t durant la réalisation d'une tâche L (imagerie main gauche) et F (imagerie pied) où $t \in [-1 s, 7 s]$. Observons quelques profils possibles sur la figure III.2. Parmi les trois cas qui sont présentés sur cette figure, correspondant à trois sources différentes dans trois bandes de fréquence distinctes, on distingue trois situations opposées. Le graphique du haut montrant la variance d'une source dans la bande entre 11 et 13 Hz montre que cette source peut être très utile pour discriminer entre les deux tâches considérées. En effet, alors que la variance de la source est constante pendant toute la réalisation de la tâche d'imagination de la main gauche (en rouge), la variance de cette même source diminue significativement lorsque le sujet réalise une tâche d'imagination d'un mouvement du pied. Cet effet est relativement constant pendant toute la durée de la tâche. Le deuxième graphique (milieu) présente un cas de figure opposé : alors que la variance de la source ne varie relativement pas lorsque le sujet réalise une tâche d'imagination du pied, celle-ci augmente significativement lors d'une imagination de la main gauche. Cet effet est ici beaucoup plus marqué en début de tâche et s'amenuise par la suite. Enfin, le troisième graphique montre un cas de figure intéressant mais inutile dans un but de discrimination. En effet, on a une source dont la variance semble augmenter de manière relativement importante lorsqu'on approche de la fin de la tâche autant dans le cas de l'imagination de la main gauche que du pied. Cette source ne permet donc pas de discriminer entre les deux tâches considérées. Les caractéristiques des filtres qui ont permis d'obtenir les deux premiers graphes de la figure III.2 sont présentées dans la figure III.3. Celle-ci présente respectivement les coefficients appliqués au niveau des capteurs (à gauche de chacun des deux groupes a et b) ainsi que la distribution spatiale de l'activité cérébrale correspondant à ces filtres (à droite de chacun des deux groupes a et b). On peut interpréter la distribution de l'activité correspondante comme la mesure qui serait réalisée au niveau des capteurs dans le cas où uniquement cette source serait active et les autres seraient silencieuses.

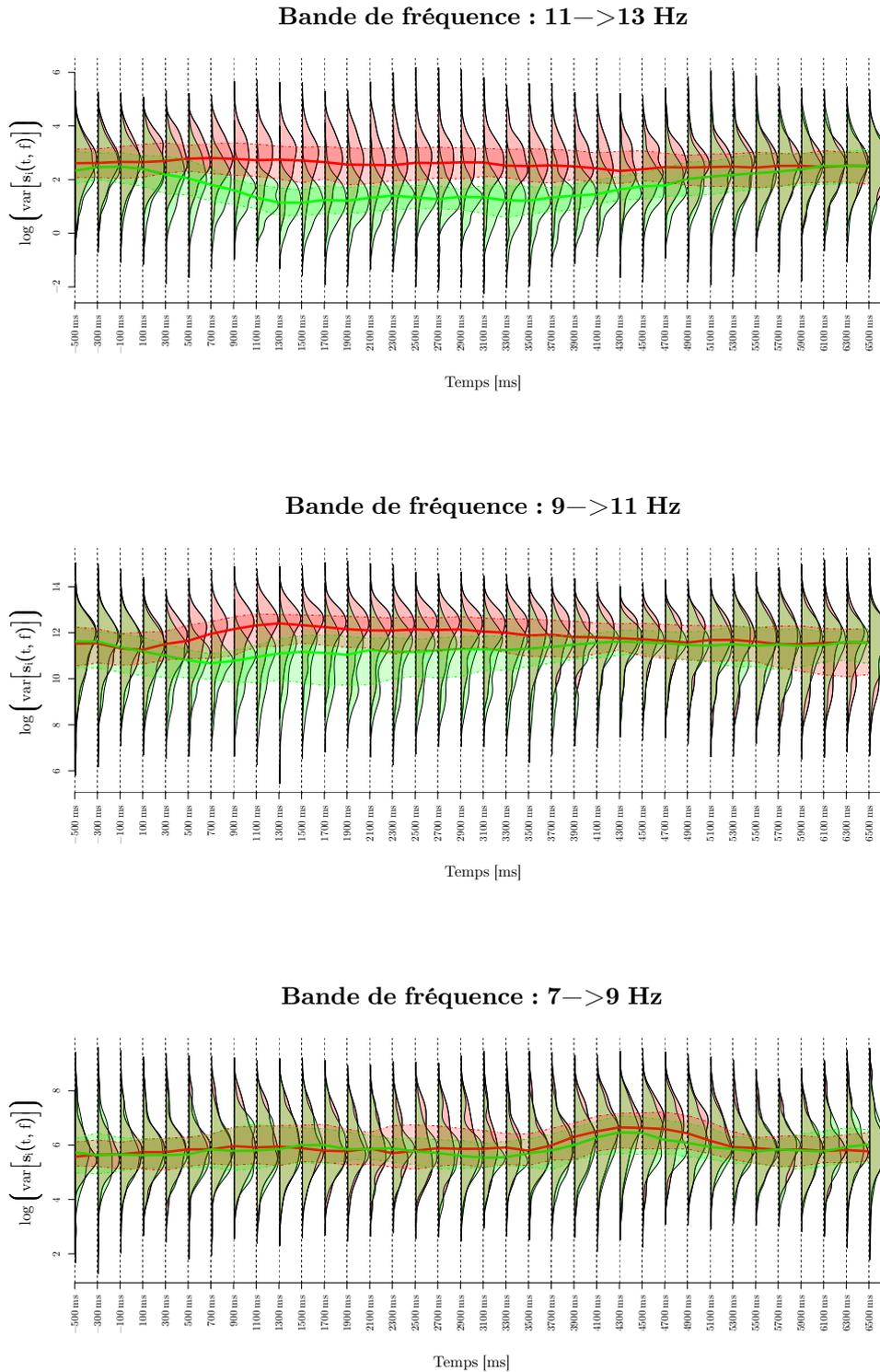


Figure III.2 Quelques profils possibles des variances pour des fréquences différentes et des sources distinctes. A partir de chaque moment indiqué sur l'axe des abscisses, on a évalué les densités de probabilité $f_S^{\{L,F\}}(i, j, t)$ pour chacune des deux tâches L (main gauche, rouge) et F (pied, vert). Sur chaque graphique, à chaque moment t , sont représentés les densités estimées à partir de 100 réalisations pour chaque tâche, la médiane en trait plein épais et les quantile 25-75% en pointillés.

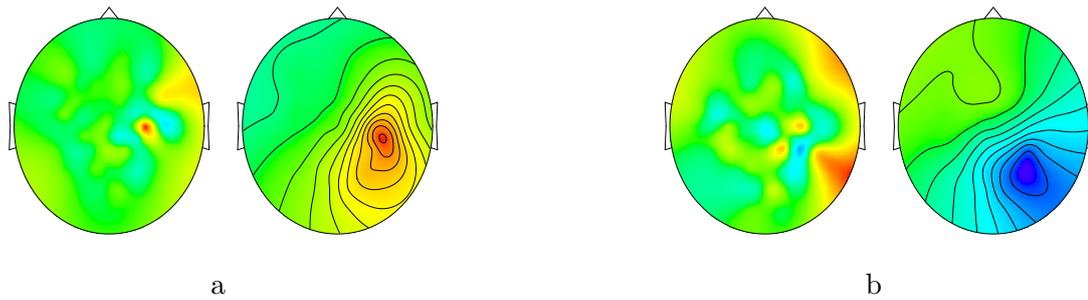


Figure III.3 a) Coefficients des filtres (gauche) correspondant à la première source de la figure précédente ainsi que distribution spatiale de l'activité associée (droite avec les lignes de niveau). b) Coefficients des filtres (gauche) correspondant à la deuxième source de la figure précédente ainsi que distribution spatiale de l'activité associée (droite avec les lignes de niveau).

Comment alors utiliser ces observations et les connaissances *a priori* sur la réalisation des tâches d'imagination de mouvements afin de calculer les filtres spatiaux W ?

Considérons qu'une expérience d'ICM est caractérisée par un ensemble de processus temporels relatifs aux sources cérébrales. Ces processus peuvent alors être observés selon différents points de vue (évolution de l'amplitude des signaux, évolution du spectre, ...) et sur différents intervalles temporels. Supposons que, pour un point de vue donné, la matrice de covariance D_k des processus sur un intervalle T_k soit diagonale. Alors d'après [Afsari \(2008\)](#), on sait qu'il est possible de retrouver les sources m et n à partir d'une observation mélangée de celles-ci si leurs variances ne varient pas de manière proportionnelle. Cela signifie qu'il suffit qu'il existe une quelconque diversité entre les sources m et n sur au moins un intervalle pour qu'il soit possible de trouver les sources à partir des observations mélangées.

Ce principe nous permet donc de pouvoir envisager l'utilisation de différentes diversités pour le contexte particulier de l'imagination de mouvements. En effet, à partir des observations faites sur l'exemple précédent ainsi que les connaissances *a priori* concernant les tâches d'imagerie motrice, on sait qu'il existe plusieurs types de diversité potentiels entre les différentes sources impliquées dans l'imagerie motrice. On peut par exemple citer :

- diversité entre tâches : dans le cas de structures cérébrales qui présentent une implication somatotopique homunculaire, une source cérébrale impliquée dans une tâche d'imagerie motrice d'une partie du corps ne sera pas impliquée dans l'imagerie d'une partie distincte du corps. Ainsi, on doit observer des sources à différentes positions spatiales pour chaque tâche d'une expérience d'imagerie motrice. La matrice de covariance des sources sur les intervalles correspondant à la réalisation d'une tâche sera donc différente de la matrice de covariance des sources sur les intervalles correspondant à la réalisation d'une tâche distincte,
- variation de l'énergie à l'intérieur d'une tâche : dans la mesure où l'imagerie motrice consiste en une succession d'activations et de désactivations de sources distinctes localisées dans différentes zones motrices (e.g. SMA et PMC), on peut aussi attendre des différences de variance entre différentes sources impliquées dans une même tâche d'imagerie motrice.

Ces variations peuvent être observées selon le point de vue temporel ou fréquentiel comme nous allons le voir dans ce chapitre. La suite de ce chapitre repose sur une hypothèse fondamentale capable d'englober ces deux types de diversité

Hypothèse 1 (VARIANCE CONSTANTE PAR MORCEAUX). Soient K sous-intervalles T_1, \dots, T_K définis sur la période d'observation considérée. On définit alors $\widetilde{\mathbf{s}}(t)^{(2)}$ les processus de moyennes et variances identiques à $\mathbf{s}(t)$ mais dont les statistiques à l'ordre strictement plus grand que deux s'annulent¹. Alors on dit que les sources $\mathbf{s}(t)$ ont une variance constante par morceaux lorsque, sur chaque sous-intervalle T_k , les processus $\widetilde{\mathbf{s}}(t)^{(2)}$ suivent une loi de probabilité Gaussienne multivariée de matrice de variance-covariance diagonale. Mathématiquement, on note

$$\forall k \in [1..K] \quad \widetilde{\mathbf{s}}(t)^{(2)} \sim \mathcal{N}(0, \Sigma_{\mathbf{s},k}) \quad (\text{III.2})$$

où pour tout k , on note $\Sigma_{\mathbf{s},k} = \text{diag}(\sigma_{1,k}^2, \dots, \sigma_{L,k}^2)$. et $\mathcal{N}(0, \Sigma_{\mathbf{s},k})$ est une distribution normale multivariée de moyenne nulle et de matrice de variance-covariance $\Sigma_{\mathbf{s},k}$. Dans le cadre qui nous intéresse, on impose également que les variances des sources $\Sigma_{\mathbf{s},k}$ ne suivent pas une variation proportionnelle.

Cette hypothèse signifie seulement que les variances des sources doivent être différentes entre deux intervalles distincts.

Nous verrons par la suite que le choix des sous-intervalles considérés dépendra du type de non-stationnarité utilisé. Ainsi, l'hypothèse précédente pourra être utilisée à la fois pour modéliser la diversité entre les tâches mais également pour modéliser la diversité à l'intérieur d'une même tâche. Soulignons que l'hypothèse de Gaussianité précédente, formulée dans [Pham \(2001\)](#) et [Pham et Cardoso \(2001\)](#), n'est qu'une hypothèse de travail afin de conduire de manière simple à un algorithme. Cette hypothèse signifie que les méthodes d'extraction des sources à partir des mesures au niveau des capteurs n'utiliseront que les statistiques de second ordre même si les statistiques d'ordre supérieur ne s'annulent pas. Le fait que les sources ne soient pas Gaussiennes n'empêche en rien la méthode de fonctionner.

3 Séparation de sources non-stationnaires

3.1 Formulation théorique

Considérons le processus aléatoire $\mathbf{x}(t) \in \mathbb{R}^{N_s}$ représentant les données EEG enregistrées à l'aide de N_s capteurs. Nous voulons trouver une transformation linéaire de $\mathbf{x}(t)$ (s'écrivant sous forme matricielle selon l'équation II.7) de façon à inférer l'intention c de l'utilisateur parmi un ensemble de M tâches prédéfinies $c \in \mathcal{C} = \{c_1, \dots, c_M\}$. La classification à partir des signaux $\mathbf{s}(t)$ doit conduire à une amélioration du score de classification comparée à l'utilisation directe de $\mathbf{x}(t)$. Nous nous plaçons dans le contexte d'une expérience d'ICM et considérons l'ensemble des données d'une expérience définies sur un intervalle de temps $[0, T]$.

Étant donné l'hypothèse formulée précédemment, en notant $\hat{\Sigma}_{\mathbf{x},k} = \frac{1}{|T_k|} \sum_{t \in T_k} \mathbf{x}(t)\mathbf{x}(t)^T$ où $|T_k|$ est le nombre d'échantillons temporels dans le sous-intervalle T_k , et $p_k = \frac{|T_k|}{\sum_j |T_j|}$ est la proportion d'échantillons temporels dans l'intervalle T_k par rapport au nombre total de points dans l'ensemble des sous-intervalles, alors on montre ([Pham, 2001](#)) que le principe du maximum de vraisemblance pour retrouver $\mathbf{s}(t)$ à partir de $\mathbf{x}(t)$ conduit à l'optimisation du critère

$$C_{\text{MV}} = \sum_{k=1}^K p_k \text{KL}(W^T \hat{\Sigma}_{\mathbf{x},k} W \parallel \text{diag}(W^T \hat{\Sigma}_{\mathbf{x},k} W)). \quad (\text{III.3})$$

¹ On définit ainsi une sorte de restriction à l'ordre deux des processus réels.

La mesure $\text{KL}(R \parallel \text{diag}(R))$ est la divergence de Kullback-Leibler entre R et la matrice diagonale ayant la même diagonale que R . Cette mesure constitue donc une mesure de déviation de R par rapport à la diagonalité. Par conséquent en définissant $\text{off}(R) = \text{KL}(R \parallel \text{diag}(R))$, le critère III.3 peut être interprété comme un critère visant à diagonaliser de manière conjointe l'ensemble des matrices $\left\{ W^T \hat{\Sigma}_{\mathbf{x},k} W \right\}_{k \in [1..K]}$ et s'écrit

$$C_{\text{MV}} = \sum_{k=1}^K p_k \text{off}(W^T \hat{\Sigma}_{\mathbf{x},k} W). \quad (\text{III.4})$$

Par conséquent le problème consistant à retrouver les sources $\mathbf{s}(t)$ responsables des mesures $\mathbf{x}(t)$ étant donnée l'hypothèse de la variance constante par morceaux se résume à la diagonalisation conjointe approchée d'un ensemble de matrices de covariance estimées $\left\{ \hat{\Sigma}_{\mathbf{x},k} \right\}_{k \in [1..K]}$, c'est-à-dire trouver une matrice W^T diagonalisant cet ensemble de matrices. Par conséquent, pour tout k , on veut que $W^T \hat{\Sigma}_{\mathbf{x},k} W$ s'exprime

$$\forall k \in [1..K] \quad W^T \hat{\Sigma}_{\mathbf{x},k} W = D_k, \quad (\text{III.5})$$

où D_k est aussi proche que possible d'une matrice diagonale¹.

3.2 « Common Spatial Pattern »

3.2.1 Deux classes

Bien que l'algorithme du CSP (Müller-Gerking et al., 1999 et Fukunaga, 1990) n'ait pas été initialement formulé de la sorte, il peut néanmoins être interprété comme l'application directe du cadre de l'extraction de sources non-stationnaires. Considérons en effet un paradigme comportant deux classes distinctes $\mathcal{C} = \{c_1, c_2\}$. Notons alors les deux ensembles en fonction de l'état mental du sujet $c(t)$

$$\begin{aligned} T_1 &= \{t \in [0, T] \mid c(t) = c_1\}, \\ T_2 &= \{t \in [0, T] \mid c(t) = c_2\}. \end{aligned} \quad (\text{III.6})$$

D'un point de vue génératif, si l'on note $\Sigma_{\mathbf{s}, T_1}$ (resp. $\Sigma_{\mathbf{s}, T_2}$) la matrice de variance-covariance des sources responsables des données EEG sur le sous-intervalle T_1 (resp. T_2), alors l'hypothèse de linéarité entre les sources et les capteurs implique

$$\begin{aligned} \mathbf{s}_{T_1}(t) \sim \mathcal{N}(0, \Sigma_{\mathbf{s}, T_1}) &\Rightarrow \mathbf{x}_{T_1}(t) \sim \mathcal{N}(0, W^{-T} \Sigma_{\mathbf{s}, T_1} W^{-1}), \\ \mathbf{s}_{T_2}(t) \sim \mathcal{N}(0, \Sigma_{\mathbf{s}, T_2}) &\Rightarrow \mathbf{x}_{T_2}(t) \sim \mathcal{N}(0, W^{-T} \Sigma_{\mathbf{s}, T_2} W^{-1}). \end{aligned} \quad (\text{III.7})$$

Or l'algorithme CSP fait l'hypothèse de lois normales multivariées de matrices de variance-covariance diagonales pour chacune des deux sources (Blankertz et al., 2008). Ainsi, on retrouve bien le cadre théorique de la séparation de sources non-stationnaires. L'algorithme CSP rajoute néanmoins une contrainte dans la résolution de ce problème ; il faut $\Sigma_{\mathbf{s}, T_1} + \Sigma_{\mathbf{s}, T_2} = I$. En notant $\hat{\Sigma}_{\mathbf{x}, T_1}$ (resp. $\hat{\Sigma}_{\mathbf{x}, T_2}$) la matrice de covariance estimée de $\mathbf{x}(t)$ sur T_1 (resp. T_2), l'algorithme CSP cherche alors un ensemble de vecteurs \mathbf{w} qui maximisent le critère

¹ A partir du travail de plusieurs auteurs (Pham, 2001, Tichavsky et Yeredor, 2009, Souloumiac, 2009, Ziehe et al., 2004 et Vollgraf et Obermayer, 2006), nous fournissons une bibliothèque fonctionnant avec R afin de résoudre efficacement la minimisation III.4 (voir l'annexe B).

$$\max_{\mathbf{w}} \frac{\mathbf{w}^T \hat{\Sigma}_{\mathbf{x}, T_1} \mathbf{w}}{\mathbf{w}^T \hat{\Sigma}_{\mathbf{x}, T_2} \mathbf{w} + \mathbf{w}^T \hat{\Sigma}_{\mathbf{x}, T_2} \mathbf{w}}. \quad (\text{III.8})$$

Ce quotient est un quotient de Rayleigh (Fukunaga, 1990 et Gray, 2006) dont la solution est obtenue par une décomposition en valeurs propres généralisée. Cette méthode conduit donc à la diagonalisation exacte de deux matrices de covariance. L'avantage de cette méthode est de fournir un ensemble de solutions \mathbf{w} qui sont classées par ordre d'importance selon la valeur des valeurs propres (voir l'équation II.11 et la discussion qui suit). L'utilisation d'une décomposition en valeurs propres généralisée peut également s'avérer un inconvénient. En effet, comme montré dans Afsari (2008), la diagonalisation exacte est très sensible aux erreurs d'estimation sur les matrices de covariance estimées car le bruit engendré par l'estimation va être lui aussi diagonalisé. Les erreurs sur les matrices de covariance se retrouvent ainsi directement sur la solution. Ceci plaide en faveur de l'utilisation de méthodes de diagonalisation approchées qui permettent l'inclusion d'un plus grand nombre de matrices dans l'ensemble à diagonaliser.

3.2.2 Extension à plus de deux classes

L'extension au cas de plus de deux classes est assez peu satisfaisante et repose sur la décomposition du problème en un ensemble de problèmes binaires (Dornhege et al., 2004 et Wu et al., 2005). Le cadre du CSP pour deux classes s'applique mais en considérant un ensemble de M problèmes parmi lesquels le m -ième s'exprime en posant

$$\begin{aligned} T_1 &= \{t \in [0, T] \mid c(t) = c_m\} & , \\ T_2 &= \{t \in [0, T] \mid \exists i \in [1..M] \setminus \{m\} \quad c(t) = c_i\}. \end{aligned} \quad (\text{III.9})$$

Les vecteurs intéressants \mathbf{w} sont alors sélectionnés de manière indépendante entre les différents problèmes.

3.3 Utilisation de la diversité entre classes : DCA

La diversité entre les classes (variance des sources constante lors de la réalisation d'une tâche et différente entre deux essais de classes distinctes) dans le cadre d'une expérience d'ICM s'applique aussi bien au cas de deux comme de plusieurs classes. Bien que le lien n'ait pas été explicité entre l'utilisation de la DCA et le cadre présenté ici, notons que l'utilisation de la DCA pour une expérience d'imagerie motrice multi-classes a d'abord été proposée dans Dornhege et al. (2004) puis récemment étendue dans Grosse-Wentrup et Buss (2008). Ces deux travaux reposent sur la définition des sous-intervalles

$$T_m = \{t \in [0, T] \mid c(t) = c_m\}. \quad (\text{III.10})$$

Les matrices de variance-covariance relatives à chaque sous-intervalle correspondent alors aux matrices de variance de chaque classe. Le cadre théorique exposé au début de cette partie montre alors qu'il est possible d'extraire les sources de chaque tâche en diagonalisant de manière conjointe l'ensemble des estimations des matrices de covariance de $\mathbf{x}(t)$ sur les sous-intervalles T_m

$$\aleph = \left\{ \hat{\Sigma}_{\mathbf{x}, T_m} \right\}_{m \in [1..M]} \quad (\text{III.11})$$

et le critère d'optimisation correspondant, dans l'hypothèse d'un nombre de réalisations de tâches égal pour chaque classe, s'écrit

$$C_{\text{DCA}} = \sum_{m=1}^M \text{off}(W^T \hat{\Sigma}_{\mathbf{x}, T_m} W). \quad (\text{III.12})$$

Ce modèle est relativement bien adapté à l'imagerie motrice puisqu'il suppose une diversité des sources cérébrales liées à chaque type de tâche. Or cette hypothèse va dans le sens de l'utilisation de l'organisation somatotopique homunculaire des aires motrices impliquées dans l'imagination de mouvement. Cependant, ce modèle se limite à considérer les sources cérébrales d'une certaine tâche actives du début à la fin de la réalisation. Le modèle que nous proposons tente donc de pallier ce point en tenant compte des possibilités d'activations et de désactivations de sources au cours d'une réalisation qui dure pendant quelques secondes.

3.4 Utilisation des variations d'énergie dans une tâche : DCA-MS

Afin de prendre en compte les activations et désactivations successives des sources cérébrales pendant la réalisation d'une tâche mentale, on définit les sous-intervalles T_k à partir d'une **subdivision régulière** d'ordre n de chacun des essais. Mathématiquement, si on note Ω l'ensemble des périodes de réalisation d'une tâche mentale et $\omega = [a, b] \in \Omega$ un intervalle correspondant à la réalisation d'une tâche, la subdivision considérée est définie en fonction de $h = (b - a)/n$ (pas de la subdivision) par les points de coupure de la subdivision

$$\omega_0 = a < \omega_1 = a + h < \omega_2 = a + 2h < \dots < \omega_{n-1} = a + (n - 1)h < \omega_n = b. \quad (\text{III.13})$$

Soit alors Ω_i les parties des intervalles de Ω appartenant à la i -ème subdivision, où $i \in [1 .. n]$. Et on définit alors les sous-intervalles $T_{m,i}$ relativement à la tâche m et la subdivision par

$$T_{m,i} = \{t \in \Omega_i \mid c(t) = c_m\}. \quad (\text{III.14})$$

On obtient donc un total de $M \times n$ intervalles sur lesquels on définit la matrice de covariance des données

$$\hat{\Sigma}_{\mathbf{x}, T_{m,i}} = \frac{1}{|T_{m,i}|} \sum_{t \in T_{m,i}} \mathbf{x}(t) \mathbf{x}(t)^T \quad (\text{III.15})$$

Le cadre de l'extraction de sources non-stationnaires présenté ci-dessus montre alors qu'en diagonalisant conjointement ces différentes matrices de covariance, cette méthode permettra d'extraire différentes sources impliquées dans la réalisation des tâches d'imagerie motrice. La fonction de coût associée à cette diagonalisation est

$$C_{\text{DCA-MS}} = \sum_{m=1}^M \sum_{i=1}^n \text{off}(W^T \hat{\Sigma}_{\mathbf{x}, T_{m,i}} W). \quad (\text{III.16})$$

Cette méthode vise donc à utiliser à la fois la diversité entre différentes tâches mais également la diversité à l'intérieur d'une même tâche. Notons que le choix de n est influencé par deux contraintes antagonistes, d'une part on aimerait choisir n suffisamment grand pour capturer l'ensemble des activations et désactivations de chacune des sources impliquées dans la tâche d'imagerie motrice, mais d'autre part ce nombre est contraint par des considérations sur l'estimation de la variance des signaux sur chaque partie de la subdivision. En effet, des subdivisions trop petites rendraient l'estimation des matrices de covariance très sensible aux variations dues au bruit. Dans la suite de cette présentation, la durée des essais est d'environ 4 s, n a donc été fixé à quatre d'après notre expérience de la nécessité

de conserver des intervalles d'analyse d'au moins une seconde. Des expériences avec trois subdivisions n'ont pas révélé des différences de performance flagrantes.

Dans la suite, cette méthode sera dénommée DCA-MS d'après Diagonalisation Conjointe Approchée (version multi-segments).

4 Paradigme expérimental

Considérons les données enregistrées par l'équipe du Prof. Pfurtscheller (Graz, Autriche) durant une expérience d'ICM basée sur quatre tâches d'imagerie motrice. Neuf sujets ont pris part à l'expérience, chacun d'entre eux a réalisé deux sessions à des dates différentes. Chaque session était composée de six parties séparées par une pause de quelques minutes. Pendant chacune des six parties, le sujet réalisait 48 tâches d'imagerie motrice (12 pour chacune des quatre tâches possibles). Au total, le sujet réalisait donc un total de 288 essais par session (72 pour chaque classe).

Les sujets étaient assis confortablement dans un siège en face d'un écran d'ordinateur. Les quatre tâches d'imagerie proposées au sujet étaient **main gauche**, **main droite**, **pied** et **langue**. Au début de chaque essai ($t = 0$ s), une croix apparaissait sur l'écran en plus d'un bref signal sonore. Après deux secondes, une flèche apparaissait pointant vers l'un des quatre points cardinaux, indiquant au sujet la tâche qu'il avait à réaliser. Cette indication restait affichée à l'écran pendant 1,25 s. Le sujet devait alors réaliser sa tâche mentale jusqu'à $t = 6$ s sans qu'aucun retour ne lui soit fourni pendant la réalisation de sa tâche. Un court instant de repos d'une durée moyenne de 1,5 s suivait la fin de la réalisation de la tâche, jusqu'à ce qu'un nouvel essai commence. La [figure III.4\(a\)](#) illustre le déroulement temporel du paradigme.

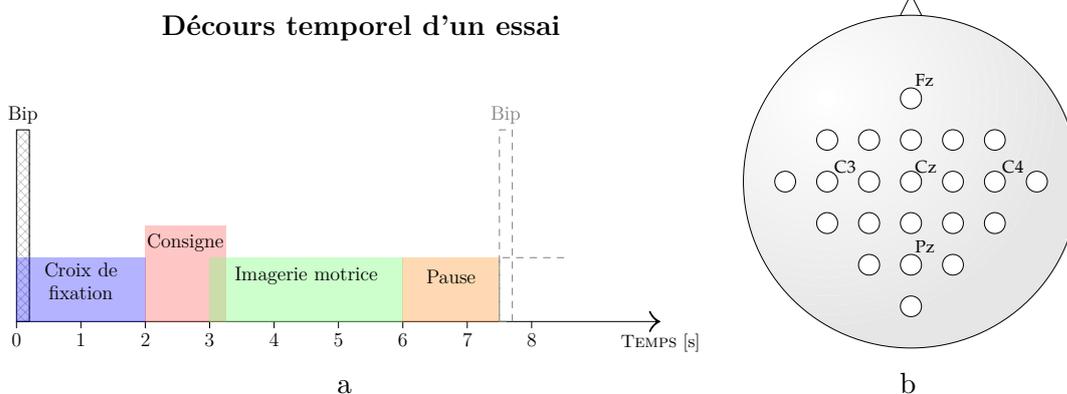


Figure III.4 a) Schéma représentant le déroulement temporel de la réalisation d'une tâche d'imagerie motrice. b) Montage EEG des 22 capteurs. La distance entre deux électrodes est de 3,5 cm. Certains emplacements correspondent aux positions standards du système 10/20, ces positions sont indiquées le cas échéant.

Les signaux EEG des sujets étaient enregistrés à l'aide de 22 capteurs (électrodes Ag/AgCl) disposés sur le scalp selon le schéma de la [figure III.4\(b\)](#). La distance entre deux électrodes était de l'ordre de 3,5 cm. Un montage monopolaire a été utilisé pour tous les enregistrements. Le mastoïde gauche était utilisé comme référence et celui de droite comme masse. Les signaux EEG étaient échantillonnés à 250 Hz et ont été filtrés analogiquement par un filtre passe-bande dont la bande passante s'étendait entre 0,5 Hz et 100 Hz. Un filtre à encoche a été également utilisé pour enlever la composante liée au signal électrique d'alimentation (50 Hz).

Bien qu'une inspection visuelle des signaux EEG ait été réalisée par un expert, aucun essai n'a été enlevé du jeu de donnée afin d'évaluer la robustesse et la sensibilité des

méthodes aux artefacts et aux valeurs aberrantes. Cependant, on peut noter qu'environ 7,5 % des essais contenait des artefacts. Notons enfin que l'équipe autrichienne a décidé, après nous avoir fourni les données, de les utiliser pour la « BCI Competition 2008 ». Les résultats obtenus avec nos méthodes seront donc comparés aux résultats obtenus par les participants à la compétition (qui nous était interdite pour des raisons évidentes).

5 Méthode

Afin de comparer les performances des trois méthodes décrites précédemment (CSP, DCA et DCA-MS), une **validation croisée** ainsi qu'une évaluation des performances pour le **transfert inter-session** sont mises en place, nous décrivons tout d'abord les procédures dans ces deux cas. Dans la suite de cette section, nous appellerons Ω^m l'ensemble des intervalles d'imagerie motrice de la classe $m \in \{L, R, F, T\}$ (respectivement d'après main gauche, main droite, pied et langue). Le i -ème élément de Ω^m , noté ω_i^m , correspond au i -ème essai de la tâche m ($i \in [1 .. N_T]$, où N_T vaut 72 dans cette expérience si l'on considère l'ensemble des données d'une session). L'ensemble des données d'entraînement sera noté \mathcal{T} et l'ensemble des données de test sera noté \mathcal{E} ($\mathcal{E} \cap \mathcal{T} = \emptyset$). Dans tous les cas suivants, les données seront supposées filtrées préalablement entre 5 et 35 Hz par un filtre passe-bande de Butterworth d'ordre 5. Par abus de notation, on pourra noter $\mathcal{T} \cap \Omega^m$ l'ensemble des essais de la classe m dans les données d'entraînement.

5.1 Phase d'entraînement

A partir des données d'entraînement, on calcule les filtres spatiaux $W_{\{\text{CSP}, \text{DCA}, \text{DCA-MS}\}}$ pour les trois méthodes présentées précédemment. Dans le cas du CSP, on conserve deux vecteurs (premier et dernier) pour chacun des quatre problèmes binaires, soit un total de $L = 8$ vecteurs. Dans les deux autres cas, on utilise la procédure mise en place dans [Grosse-Wentrup et Buss \(2008\)](#) pour sélectionner un ensemble de $L = 8$ vecteurs parmi les 22 obtenus suite à la diagonalisation jointe des matrices de covariance. Cette méthode utilise un critère basé sur la théorie de l'information pour classer les différents vecteurs en calculant une approximation de l'information mutuelle entre la classe et les données projetées. Cette information mutuelle est donnée, dans l'hypothèse de classes équiprobables par la formule

$$I(c(t), \mathbf{w}^T \mathbf{x}(t)) \approx - \sum_{m=1}^M \frac{1}{M} \log \sqrt{\mathbf{w}^T \hat{\Sigma}_{\mathbf{x}, T_m} \mathbf{w}} - \frac{3}{16} \left(\sum_{m=1}^M \frac{1}{M} ((\mathbf{w}^T \hat{\Sigma}_{\mathbf{x}, T_m} \mathbf{w})^2 - 1) \right)^2. \quad (\text{III.17})$$

Les filtres spatiaux les meilleurs sont alors ceux qui obtiennent l'information mutuelle la plus grande. Ce critère fournit donc une mesure quantitative permettant de mesurer la qualité des filtres obtenus.

Dans les trois méthodes, la concaténation des différents vecteurs permet d'obtenir une matrice $W \in \mathbb{R}^{N_s \times L}$ qui est utilisée pour projeter les données et obtenir ainsi les sources estimées correspondant aux données d'entraînement $\hat{\mathbf{s}}(t) = W^T \mathbf{x}(t)$.

Chaque source est ensuite filtrée par un banc de $N_f = 15$ filtres recouvrant la bande de fréquence entre 5 et 35 Hz, de largeurs 2 Hz et de bandes passantes disjointes, résultant en $\hat{s}_i(t, f_j)$, $i \in [1 .. L]$ et $j \in [1 .. N_f]$. Ce traitement donne donc lieu à $N_f \times L$ signaux unidimensionnels dont les caractéristiques vont être utilisées pour entraîner un classifieur basé sur la régression logistique (régularisation ℓ_1/ℓ_2 de paramètre 0,3 fixé par validation croisée, voir [annexe A](#) pour les détails concernant le classifieur). Le système de classification que nous mettons en place peut s'écrire comme une application

$$\begin{aligned} \mathcal{A} &: F(\mathcal{I}, \mathbb{R}^{Ns \times L}) \longmapsto F(\mathcal{I}, [0, 1]^4) \\ \chi(t) &\longrightarrow [p_L(t), p_R(t), p_F(t), p_T(t)]^T \end{aligned} \quad (\text{III.18})$$

où $F(\mathcal{I}, \mathbb{R}^{Ns \times L})$ est l'ensemble des fonctions de l'intervalle de temps correspondant à la réalisation d'une tâche dans l'espace des caractéristiques (trajectoire d'évolution des caractéristiques au cours d'une tâche) et $F(\mathcal{I}, [0, 1]^4)$ est l'ensemble des fonctions du même intervalle de temps \mathcal{I} dans un espace représentant les probabilités respectives de chacune des classes. On a également pour tout t dans \mathcal{I} , $p_L(t) + p_R(t) + p_F(t) + p_T(t) = 1$. Afin de construire cette application \mathcal{A} , on échantillonne \mathcal{I} en utilisant un ensemble de fenêtres glissantes de longueur une seconde et de taux de recouvrement 80 %. Pour chaque fenêtre glissante centrée autour de $t \in \mathcal{I}$ de longueur w , l'application \mathcal{A} est alors évaluée comme la régression logistique entre les caractéristiques des données d'entraînement définies comme

$$\chi_{i,j}(t) = \log \left(\int_{-\frac{w}{2}}^{\frac{w}{2}} |\hat{s}_i(t+u, f_j)|^2 du \right) \quad (\text{III.19})$$

pour chaque $\omega_i^m \in \mathcal{T} \cap \Omega^m$ et les classes théoriques m représentant la tâche réalisée par le sujet. Les caractéristiques de \mathcal{A} ainsi que W sont alors gardées en mémoire afin de passer à la phase de test.

5.2 Phase de test

La première étape consiste à appliquer le filtrage spatial linéaire W obtenu durant l'étape d'entraînement. La même étape de transformation en signaux temps-fréquence que pendant l'étape d'entraînement est appliquée. Ensuite les caractéristiques de chaque essai de l'ensemble \mathcal{E} sont calculées selon l'équation III.19 pour chacune des fenêtres temporelles utilisées dans la phase d'entraînement. A partir de l'évolution des caractéristiques temporelles pour chaque essai de la base de test $\chi_\omega(t)$, on applique \mathcal{A} afin d'obtenir une estimation des trajectoires des probabilités de chaque classe $p_m(t)$. Pour chaque essai ω , on attribue la classe m définie par

$$\hat{m} = c(t) = \arg \max_m \left(\arg \max_t p_m(t) \right).$$

C'est-à-dire que l'on choisit la classe qui a atteint la maximum global des probabilités durant l'essai. Ces deux phases sont résumées sur la figure III.5.

5.3 Validation croisée

Afin de tester les performances des méthodes, on utilise tout d'abord la validation croisée. Pour ce faire, on considère les données session par session. Pour chaque session, on définit un ensemble d'entraînement au hasard, de cardinal $M \times T_s$, contenant un nombre égal d'essais de chaque classe. On applique la phase d'entraînement puis la phase de test et on compte le nombre de fausses classifications. Cette procédure est répétée 20 fois (1×20 validations croisées pour lesquelles les données d'entraînement sont choisies indépendamment et de manière aléatoire) afin d'obtenir le taux moyen de classification pour chaque méthode. Dans l'étude qui suit, T_s varie par pas de 10 entre 10 et 60.

5.4 Transfert inter-session

L'évaluation des performances par cette méthode consiste à utiliser l'une des sessions réalisées par le sujet comme données d'entraînement et l'autre session constitue la base de

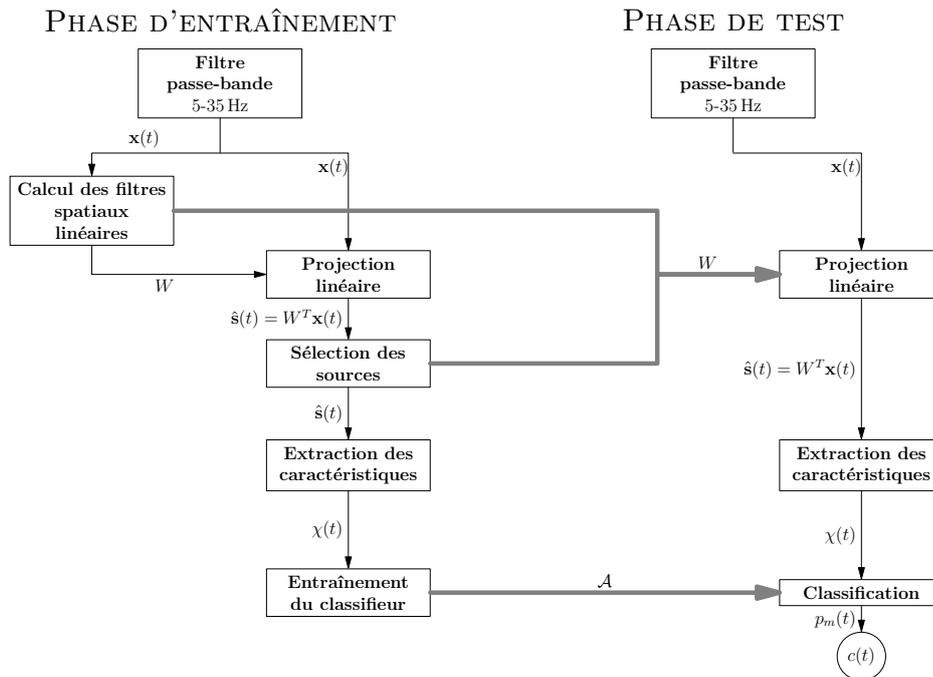


Figure III.5 Schéma bloc résumant les différentes phases de traitement dans les étapes d'entraînement et de test.

données de test. On peut alors également définir le taux de bonnes classifications comme le nombre d'essais bien classés parmi les données de test par rapport au nombre total d'essais que contient la base de données de test.

6 Résultats

6.1 Validation croisée

6.1.1 Analyse de variance dans le cas $T_s = 60$

Afin de comparer les trois méthodes de filtrage spatial dans le cas de la validation croisée, nous considérons les performances obtenues avec chacune des trois méthodes dans le cas où la taille du jeu de données d'entraînement est maximum ($T_s = 60$ sur les 72 essais disponibles). L'application de la procédure décrite dans la [section 5](#) pour chacun des sujets, chacune des sessions et chacune des trois méthodes de filtrage spatial constitue l'ensemble des données. L'analyse de variance ([Everitt, 2005](#)) nécessite tout d'abord d'identifier le type de chaque variable. Les données rentrent dans le cadre de la théorie sur les mesures répétées. Les variables SESSION (deux niveaux, session 1 ou session 2) et MÉTHODE (trois niveaux, CSP, DCA ou DCA-MS) sont des facteurs intra-sujet. La conception de l'expérience implique aussi un facteur aléatoire SUJET (voir [Baron et Li \(2003\)](#) concernant la raison pour laquelle ce facteur doit être aléatoire). La variable dépendante PERF est définie comme le pourcentage de bonnes classifications moyen obtenu lors des 20 validations croisées. Le [tableau III.1](#) rapporte les résultats détaillés de l'analyse de variance.

Celui-ci fait apparaître un effet principal du facteur MÉTHODE ($F(2; 16) = 5,99$, $p = 0,011$), tandis qu'aucun effet principal de la variable SESSION ni de l'interaction MÉTHODE \times SESSION n'est observé. Ces résultats montrent qu'il n'y a pas de différences significatives entre les deux sessions d'un même sujet, par conséquent aucun effet d'entraînement

	Deg. de liberté	stat. F	p-value
SESSION	(1;8)	0,94	0,360
MÉTHODE	(2;16)	5,99	0,011
MÉTHODE × SESSION	(2;16)	1,16	0,339

Tableau III.1 Résumé de l'analyse de variance pour les performances en validation croisée. L'influence de chacun des facteurs de la première colonne est évaluée en calculant la statistique F de degrés de liberté indiqué dans la deuxième colonne. La troisième colonne indique la p -valeur correspondante. En est indiqué un effet significatif ($p < 0.05$).

n'est observé dans ces données. Il est important de noter que la variable MÉTHODE × SESSION ne donne aucun effet significatif non plus. Cela montre que l'effet observé pour la variable MÉTHODE est consistant entre les deux sessions.

Comme le résultat de l'effet de la variable MÉTHODE montre qu'au moins une paire de méthodes est significativement différente, il reste à comparer les trois paires possibles (CSP vs DCA, CSP vs DCA-MS et DCA vs DCA-MS). Pour réaliser cette analyse post hoc¹, nous calculons la p -valeur quantifiant la différence entre chaque paire et utilisons une fonction de contraste de Tukey pour corriger les p -valeurs en raison des comparaisons multiples. Les résultats de cette étape montrent une différence significative entre DCA-MS et CSP ($t(17) = 2,867$, $p = 0,027$) alors qu'aucune différence significative n'est observée entre DCA et CSP ou entre DCA-MS et DCA ($p = 0,120$ et $p = 0,726$ respectivement).

6.1.2 Résultats détaillés pour $T_s = 60$

Nous présentons dans le [tableau III.2](#) les performances en validation croisée obtenues par la meilleure des trois méthodes pour chaque sujet et chaque session. Ces performances représentent la moyenne des pourcentages de bonne classification pour les 20 répétitions. Seules les performances correspondant à une taille des données d'entraînement de $T_s = 60$ sont considérées.

Les performances sont très hétérogènes entre les sujets, allant de 48,1 % pour le sujet 6 lors de la première session à 87,5 % pour le sujet 3 lors de la seconde session. Notons enfin que les performances moyennes obtenues avec les trois méthodes sont de 68,7 % ($\sigma = 13,0$) pour la méthode DCA, 65,9 % pour le CSP ($\sigma = 9,9$) et 69,8 % pour DCA-MS ($\sigma = 10,6$).

6.1.3 Influence de T_s

Au cours de cette étude, la variable T_s a pris toutes les valeurs des dizaines entre 10 et 60. Les résultats pour $T_s = 60$ ont déjà été largement explorés ci-dessus ; nous nous intéressons maintenant à l'influence de cette variable sur les performances en validation croisée. La [figure III.6](#) montre l'évolution des performances de classification en fonction de la taille du jeu de données d'entraînement pour deux sujets (sujet 3 session 2 et sujet 7 session 2). Ces deux échantillons sont relativement représentatifs de l'ensemble des sujets. Cette figure montre d'une part que les différences entre les méthodes sont minimales pour certains d'entre eux (e.g. sujet 7) mais peuvent s'avérer plus importantes pour d'autres (e.g. sujet 3). D'autre part, La [figure III.6](#) montre que le rang de chaque méthode est

¹ Expression latine signifiant littéralement « après cela ». Dans l'expression romaine originelle *post hoc, ergo propter hoc* (c'est arrivé après ceci, donc ceci en est la cause), les romains voulaient dénoncer la tendance de l'esprit humain à vouloir trouver des liens là où il n'en existe pas. En statistique, on parle d'une analyse post hoc lorsqu'on veut préciser, en utilisant des comparaisons paires par paires, une différence qui est apparue à l'aide d'une analyse de variance.

		Performance [%]	Meilleure méthode
Sujet 1	Session 1	72,2	DCA
	Session 2	77,2	DCA-MS
Sujet 2	Session 1	62,7	CSP
	Session 2	59,3	DCA
Sujet 3	Session 1	85,4	DCA
	Session 2	87,5	DCA
Sujet 4	Session 1	76,0	DCA-MS
	Session 2	70,5	DCA
Sujet 5	Session 1	75,0	DCA-MS
	Session 2	80,6	DCA-MS
Sujet 6	Session 1	48,1	CSP
	Session 2	55,3	DCA-MS
Sujet 7	Session 1	58,1	DCA-MS
	Session 2	55,7	DCA-MS
Sujet 8	Session 1	81,7	DCA
	Session 2	77,9	DCA-MS
Sujet 9	Session 1	72,6	DCA
	Session 2	79,1	DCA

Tableau III.2 Performances détaillées obtenues par la meilleure méthode (troisième colonne) pour chacun des sujets et chacune des sessions.

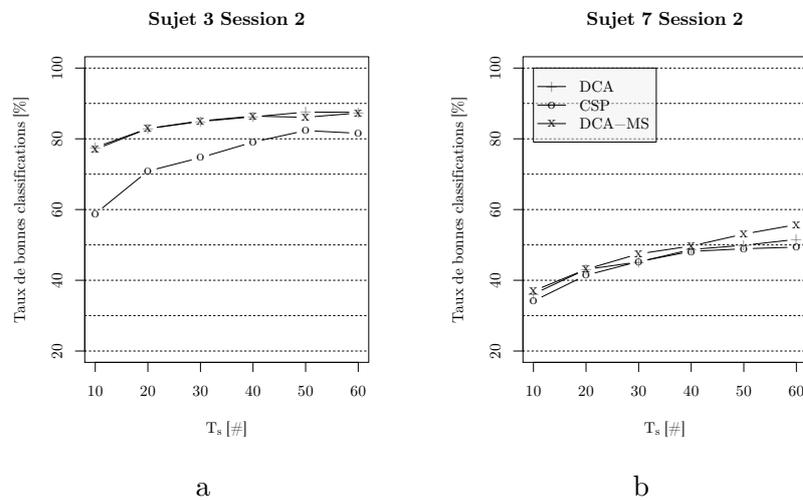


Figure III.6 Figure montrant l'influence du nombre d'essais dans le jeu de données d'entraînement. Seuls les sujets 3 et 7, session 2 sont présentés ici.

relativement bien respecté pour l'ensemble des T_s considérés ici. Ceci justifie *a posteriori* que nous n'ayons considéré que les résultats pour $T_s = 60$ dans l'analyse de variance.

6.2 Transfert session à session

L'évaluation des performances dans le cas du transfert entre sessions est possible sur ce jeu de données qui comporte deux sessions réalisées sur deux jours différents. Les performances qui sont présentées dans la suite de cette section sont représentées par deux

valeurs pour chacun des sujets. On notera $1 \rightarrow 2$ dans le cas où la session 1 est utilisée pour entraîner les données et la session 2 en tant que session d'évaluation ; $2 \rightarrow 1$ dans le cas contraire.

6.2.1 Analyse de variance

Le même modèle que dans le cas de la validation croisée est utilisé. Les résultats de l'analyse de variance sont rapportés dans le [tableau III.3](#). Les résultats sont similaires au cas de la validation croisée dans la mesure où aucun effet principal du facteur SESSION ni de l'interaction MÉTHODE \times SESSION. Tandis qu'un effet significatif principal de la variable MÉTHODE est à nouveau observé ($F(2; 16) = 5,80, p = 0,013$). Une analyse post hoc, destinée à comparer les méthodes deux à deux, est à nouveau réalisée. Elle permet d'observer à nouveau une différence significative entre DCA-MS et CSP ($t(17) = 2,981, p = 0,022$). Aucun effet significatif n'est observé entre DCA et DCA-MS tandis qu'un effet marginalement significatif apparaît entre DCA et CSP ($t(17) = 2,545, p = 0,052$). Ces effets confirment donc les résultats obtenus lors de la validation croisée.

	Deg. de liberté	stat. F	p-value
SESSION	(1;8)	2,45	0,156
MÉTHODE	(2;16)	5,80	0,013
MÉTHODE \times SESSION	(2;16)	0,11	0,89

Tableau III.3 Résumé de l'analyse de variance pour les performances en transfert session à session. L'influence de chacun des facteurs de la première colonne est évaluée en calculant la statistique F de degrés de liberté indiqué dans la deuxième colonne. La troisième colonne indique la p -valeur correspondante. En est indiqué un effet significatif ($p < 0.05$).

6.2.2 Résultats détaillés

Nous présentons les résultats détaillés pour chaque sujet, chaque session et chaque méthode dans la [figure III.7](#). Les performances d'une méthode pour un sujet donné dans les cas $2 \rightarrow 1$ (Sess. 1) et $1 \rightarrow 2$ (Sess. 2) sont reliées par une ligne droite. Cette figure montre bien l'hétérogénéité des résultats d'un sujet à un autre mais la relative constance d'un sujet d'une session à l'autre.

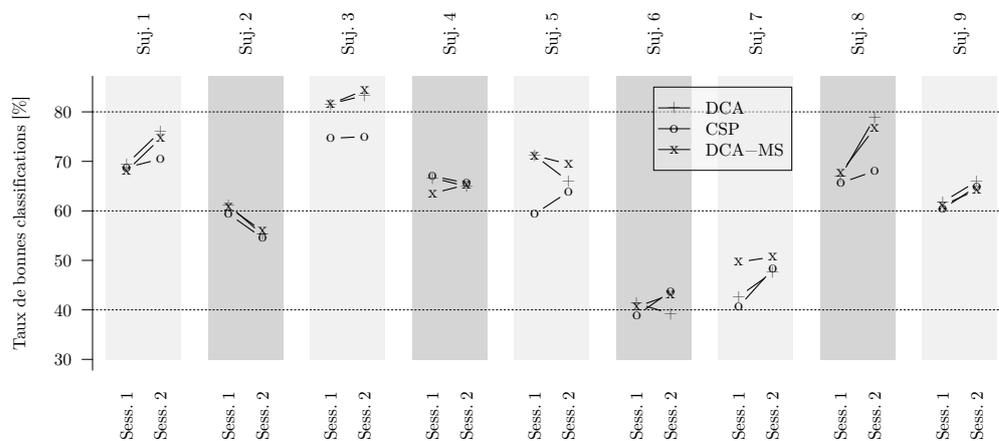


Figure III.7 Résumé des performances obtenues par les trois méthodes considérées pour chacun des sujets dans les cas $2 \rightarrow 1$ (Sess. 1) et $1 \rightarrow 2$ (Sess. 2).

Si l'on moyenne l'ensemble des 18 sessions pour chacune des méthodes, on obtient un score moyen de 63,3 % pour DCA ($\sigma = 13, 5$), 60,5 % pour CSP ($\sigma = 11, 0$) et enfin 63,8 % pour DCA-MS ($\sigma = 12, 3$). Ces résultats montrent également qu'il vaut mieux réentraîner les filtres spatiaux à chaque session afin d'obtenir les meilleures performances.

6.3 Comparaison avec la « BCI Competition 2008 »

Le jeu de données utilisé dans cette analyse a été mis à la disposition du public lors de la quatrième « BCI Competition ». La compétition a été proposée de la manière suivante : les données de la session 1 étaient fournies avec les tâches réalisées par les sujets tandis que seules les données EEG de la session 2 étaient fournies (sans les étiquettes). Les participants à la compétition devaient alors étiqueter chacun des 288 essais de la deuxième session parmi les quatre tâches disponibles. Leurs méthodes étaient donc entraînées sur le jeu de données de la session 1 et testées sur la session 2. Cette procédure correspond donc aux résultats que nous avons appelés 1 \rightarrow 2. De plus la méthode d'évaluation était quelque peu différente de celle qui a été employée jusqu'à présent dans notre analyse. La mesure employée était la mesure appelée Kappa (κ) (Cohen, 1960) définie à partir de la matrice de confusion. Notons la matrice de confusion des essais prédits par rapport aux essais réels (en probabilités)

$$C_{\text{Préd., Réel}} = \begin{pmatrix} p_{L,L} & p_{L,R} & p_{L,F} & p_{L,T} \\ p_{R,L} & p_{R,R} & p_{R,F} & p_{R,T} \\ p_{F,L} & p_{F,R} & p_{F,F} & p_{F,T} \\ p_{T,L} & p_{T,R} & p_{T,F} & p_{T,T} \end{pmatrix} \quad (\text{III.20})$$

où $p_{a,b}$ représente la probabilité de prédire la tâche a lorsque le sujet réalisait la tâche b. Alors κ est défini par

$$\kappa = \frac{\sum_{a=\{L,R,F,T\}} p_{a,a} - \sum_{a=\{L,R,F,T\}} \left[\left(\sum_{b=\{L,R,F,T\}} p_{a,b} \right) \left(\sum_{b=\{L,R,F,T\}} p_{b,a} \right) \right]}{1 - \sum_{a=\{L,R,F,T\}} \left[\left(\sum_{b=\{L,R,F,T\}} p_{a,b} \right) \left(\sum_{b=\{L,R,F,T\}} p_{b,a} \right) \right]} \quad (\text{III.21})$$

Les scores obtenus par les trois méthodes comparées dans notre analyse sont présentés dans le [tableau III.4](#) ainsi que les scores obtenus par les trois meilleurs compétiteurs. Les résultats confirment la supériorité de DCA-MS et placent cette méthode de façon relativement honorable comparativement aux meilleurs résultats obtenus.

Ce tableau confirme une nette supériorité des méthodes basées sur la diagonalisation conjointe approchée par rapport au CSP. Ces résultats confirment donc la validité de l'approche proposée dans ce chapitre afin d'améliorer les performances des systèmes ICMs. La différence de performance entre la méthode DCA et DCA-MS apparaît comme relativement négligeable en observant la moyenne. Cependant, il faut néanmoins observer que la méthode DCA-MS est meilleure que la méthode DCA pour 7 sujets sur 9. Or il est statistiquement relativement improbable que l'on observe une telle hégémonie au hasard ($p < 0,05$ en utilisant un modèle multinomial équiprobable). Il faut noter que les différences observées avec les deux meilleurs participants de la compétition sont assez négligeables pour la plupart des sujets. Néanmoins, pour le sujet 7, on observe un écart très important. Bien que la raison exacte de ce résultat n'ait pas été exhibée, il se peut que, pour ce sujet, le choix des bandes de fréquence soit plus important que pour les autres sujets et que les participants à la compétition aient largement optimisé ce paramètre. Les performances de notre méthode sont également sensiblement inférieures à celles obtenues par le meilleur compétiteur pour le sujet 9. En conclusion, les deux méthodes explorées dans ce chapitre, DCA et DCA-MS permettent d'obtenir d'excellents résultats et sont compétitives par rapport aux meilleurs algorithmes actuels. En particulier, l'extension

	1 ^{er}	2 ^{ème}	3 ^{ème}	CSP	DCA	DCA-MS
Sujet 1	0,68	0,69	0,38	0,52	0,65	0,66
Sujet 2	0,42	0,34	0,18	0,39	0,40	0,42
Sujet 3	0,75	0,71	0,48	0,67	0,77	0,77
Sujet 4	0,48	0,44	0,33	0,50	0,50	0,51
Sujet 5	0,40	0,16	0,07	0,49	0,44	0,50
Sujet 6	0,27	0,21	0,14	0,19	0,18	0,21
Sujet 7	0,77	0,66	0,29	0,26	0,25	0,30
Sujet 8	0,75	0,73	0,49	0,57	0,72	0,69
Sujet 9	0,61	0,69	0,44	0,40	0,50	0,46
Moyenne	0,57	0,52	0,31	0,41	0,49	0,50

Tableau III.4 Mesures Kappa obtenues par les trois meilleurs compétiteurs ainsi que par les trois méthodes évaluées ici (CSP, DCA et DCA-MS). En est indiqué le meilleur score pour chacun des sujets.

proposée dans ce chapitre (DCA-MS) permet d'obtenir des résultats au moins équivalents à la méthode DCA (et la plupart du temps meilleurs), validant par là même la démarche.

6.4 Artefact numérique ou information réelle

Comme mentionné par [Afsari \(2008\)](#), l'inclusion de nouvelles matrices dans l'ensemble à diagonaliser peut quelquefois être suffisante pour améliorer la qualité des résultats. Ainsi, afin d'être sûrs que les bonnes performances de la méthode DCA-MS étaient dues à l'apport d'information concernant les activations et les désactivations des sources pendant une tâche d'imagerie motrice, nous avons appliqué la méthode présentée pour DCA-MS en rendant aléatoire la position des sous-intervalles de chaque essai. Ce faisant, nous rendons aléatoire la structure temporelle de chaque essai et éliminons toute information que l'application de DCA-MS pourrait apporter. Les performances deviennent alors comparables à celles obtenues avec la méthode DCA pour la validation croisée aussi bien que pour le cas du transfert entre sessions (66,8% et 62,1% respectivement). Cette procédure montre que la prise en considération de la diversité à l'intérieur de chaque essai est potentiellement intéressante pour augmenter les performances obtenues.

7 Utilisation du domaine fréquentiel

Jusqu'à présent, la séparation était basée sur le modèle simple liant les sources aux observations

$$\mathbf{s}(t) = W^T \mathbf{x}(t).$$

On peut étendre ce modèle en considérant que la matrice de séparation W^T dépend de chacune des fréquences considérées. Formellement, dans la littérature de la séparation aveugle de sources, ceci revient à considérer un modèle convolutif qui peut s'écrire dans le domaine temps-fréquence

$$\mathbf{s}(t, f) = W_f^T \mathbf{x}(t, f). \quad (\text{III.22})$$

Dans ce cas, si la transformée temps-fréquence choisie est un ensemble de $N_f = 15$ filtres recouvrant la bande de fréquence entre 5 et 35 Hz, de largeur de bande de 2 Hz et disjoints,

alors le modèle III.22 appliqué à la méthode DCA implique de considérer N_f problèmes disjoints de diagonalisation conjointe. Les sous-intervalles de stationnarité de la matrice de variance-covariance sont définis de la même manière que ceux de l'équation III.10. Ainsi, en posant $\hat{\Sigma}_{\mathbf{x}, T_m, f_j}$ la covariance estimée des signaux $\mathbf{x}(t)$ filtrés par le filtre autour de la fréquence f_j , sur le sous-intervalle T_m , alors le cadre exposé au début de ce chapitre montre qu'on peut extraire chacune des sources $\hat{\mathbf{s}}(t, f_j)$ en considérant le problème autour de la fréquence f_j comme la diagonalisation jointe de l'ensemble

$$\mathfrak{N}_{f_j} = \left\{ \hat{\Sigma}_{\mathbf{x}, T_m, f_j} \right\}_{m \in [1 \dots M]} . \quad (\text{III.23})$$

Notons qu'à la différence de la méthode DCA qui donne N_s sources potentielles, cette méthode fréquentielle résulte en $N_s \times N_f$ sources potentielles. Cependant, on considère le même nombre de caractéristiques dans les deux cas en prenant en compte le fait que chacune des matrices de séparation W_f^T est calculée de manière privilégiée pour une seule fréquence. On obtient ainsi $L \times N_f$ caractéristiques dans ce cas également, où rappelons que L est le nombre de filtres spatiaux retenus. Le principe de la méthode est résumé sur la figure III.8.

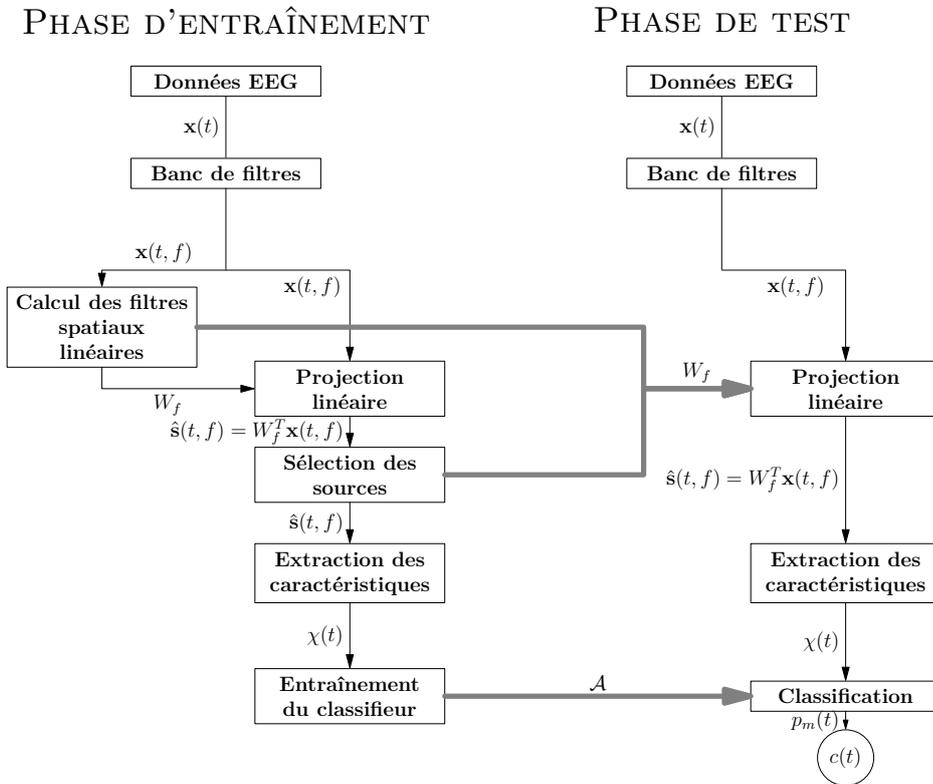


Figure III.8 Schéma bloc résumant les différentes phases de traitement dans les étapes d'entraînement et de test dans le cas de l'utilisation de l'information fréquentielle.

Seul le cas du transfert entre sessions a été considéré pour comparer la méthode DCA à la méthode fréquentielle présentée ici que nous appellerons DCA-FR (diagonalisation conjointe approchée dans le domaine des fréquences). Nous ne détaillons pas les résultats qui sont relativement similaires à ceux présentés dans la section précédente. En moyenne, DCA-FR conduit à une amélioration d'environ 1 % des résultats obtenus par DCA. DCA-FR obtient environ 64 %. Cette amélioration est relativement minime si l'on considère l'augmentation de complexité que cette méthode engendre. Cette méthode peut néanmoins s'avérer intéressante par rapport à DCA-MS lorsque le nombre d'échantillons temporels d'un essai ne permet pas de considérer les successions d'activations et de désactivations.

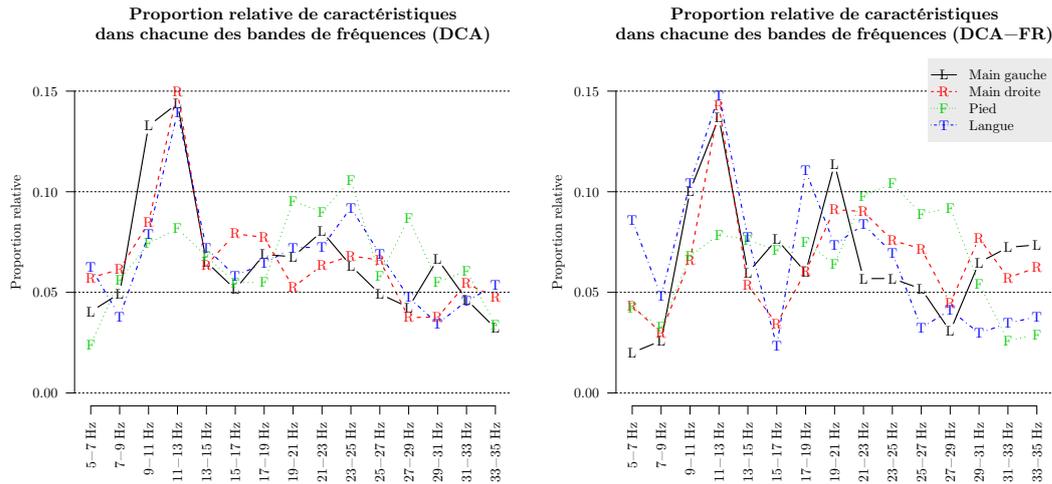


Figure III.9 Analyse de l'utilisation des bandes de fréquences par le classifieur. Pour chacune des courbes représentant l'utilisation des bandes pour classifier chaque tâche, la somme de tous les points d'une courbe est de un, de telle sorte que nous pouvons observer la proportion relative de l'utilisation de chacune des bandes.

Elle nous donne cependant l'occasion de nous intéresser de manière plus approfondie à l'utilisation des différentes caractéristiques par le classifieur. Comme décrit dans l'[annexe A](#), le classifieur réalise la sélection des composantes en même temps que la classification à l'aide d'une contrainte basée sur une combinaison des normes ℓ_1/ℓ_2 . Il est donc intéressant d'observer quelles sont les bandes de fréquence les plus utilisées pour classifier chacune des tâches. Pour ce faire, on relève chacun des coefficients non nuls de chacun des quatre modèles linéaires entraînés par le classifieur (une combinaison linéaire par classe). Chaque coefficient linéaire correspond alors à une caractéristique d'une source dans une bande de fréquence déterminée. Soit donc r_{f_i} la proportion de caractéristiques utilisées appartenant à la bande f_i pour un sujet et une session. On a donc

$$\sum_{f_i} r_{f_i} = 1. \quad (\text{III.24})$$

On moyenne alors les r_{f_i} pour l'ensemble des sujets et des sessions. En appliquant alors cette méthode pour DCA et DCA-FR, on obtient la [figure III.9](#) qui résume l'utilisation fréquentielle des bandes de fréquence par notre classifieur. Cette figure montre que la bande prédominante μ est utilisée de manière privilégiée pour les deux méthodes pour trois tâches sur quatre (exceptée la tâche « pied »). Il est également intéressant de constater que des caractéristiques dans l'ensemble de la bande considérée sont utilisées. Enfin, on observe bien la différence entre DCA et DCA-FR en constatant que les caractéristiques utilisées dans les bandes β et γ sont beaucoup plus importantes dans le cas de DCA-FR. Il semble donc que la diagonalisation jointe des matrices de covariance (méthode DCA) a donné lieu à une sur-représentation de la bande prépondérante tandis que le modèle fréquentiel utilisé dans DCA-FR laisse une place plus importante aux bandes « annexes ». Ce résultat montre que le modèle linéaire instantané s'avère approximativement correct pour les fréquences basses alors qu'un modèle convolutif pourrait s'avérer plus adapté pour l'analyse des fréquences plus hautes.

8 Discussion

Dans ce chapitre, nous avons présenté un cadre théorique très général pour résoudre le problème de l'extraction de sources dans la cas d'expériences d'ICM. Ce cadre repose sur le concept de non-stationnarité qui a été explicité à l'aide de l'[hypothèse 1](#). Dans

ce cadre, la solution obtenue à l'aide du principe de maximum de vraisemblance (Pham, 2001) repose sur la diagonalisation conjointe d'un ensemble de matrices. La définition de cet ensemble de matrices permet d'influencer le type de sources considérées et donc de réaliser une extraction de sources fondée en partie sur un modèle. Nous avons montré que ce cadre représente l'extension naturelle du CSP au cas où plus de deux classes sont considérées.

Les résultats présentés concernant ce jeu de données en particulier peuvent être comparés à différents autres travaux ayant considéré ces mêmes données. Ainsi, deux articles avaient été précédemment publiés sur ce jeu de données (Naeem et al., 2006 et Brunner et al., 2007) (bien que le huitième sujet ne soit pas présent dans ces analyses). Dans Naeem et al. (2006), différents algorithmes classiques d'ACI (SOBI, Infomax) étaient comparés au CSP. Les meilleurs résultats étaient obtenus par le CSP avec un taux de bonnes classifications moyen en validation croisée de l'ordre de 65 % et de 61,5 % en transfert entre sessions. Les résultats présentés dans cette partie peuvent être comparés aux résultats de ces articles en enlevant le sujet 8. DCA-MS obtient un score de bonnes classifications de l'ordre de 68,5 % en validation croisée et de 62,8 % en transfert entre sessions. Les procédures présentées ici permettent donc d'améliorer celles présentées dans Naeem et al. (2006). Dans Brunner et al. (2007) les auteurs ont utilisé Infomax et obtenu des scores de classification en validation croisée supérieurs à ceux présentés ici (de l'ordre de 75 %). Ces excellents résultats ont cependant été obtenus au prix d'une lourde procédure d'extraction et de sélection des caractéristiques. En effet, au lieu de considérer quelques bandes de fréquences disjointes, les auteurs ont utilisé une multitude de filtres recouvrant la bande de fréquence entre 5 et 25 Hz et de largeurs variables (9 à 11 Hz, 10 à 12 Hz, ..., 23 à 25 Hz, 9 à 13 Hz, ...). De cette procédure résultaient alors 1364 caractéristiques à classifier. C'est pourquoi une lourde étape de sélection des caractéristiques était employée (Sequential Floating Forward Selection - SFFS). Néanmoins cette procédure, bien que permettant d'obtenir d'excellents résultats dans le cas de la validation croisée, ne permettait pas d'accroître significativement les performances de transfert entre sessions (de l'ordre de 61,5 %). Ceci montre que cette procédure a donné lieu à un effet de sur-apprentissage des données de chaque session, utilisant des caractéristiques peu fiables à long terme. C'est-à-dire que des caractéristiques propres au jour de réalisation étaient apprises, et par conséquent ne permettait pas d'être généralisées à des jours différents. Le travail présenté dans cette partie est également relativement proche de Grosse-Wentrup et Buss (2008), qui avait d'ores et déjà comparé CSP et DCA. Cependant l'analyse ne présentait que trois sujets et seul le cas de la validation croisée pour des essais réalisés lors du même jour était considérée. Dans cet article, les auteurs ont observé une nette supériorité de DCA par rapport au CSP.

Les résultats obtenus sont très satisfaisants si l'on considère que l'ensemble de ces sujets utilisaient une ICM pour la première fois (Blankertz et al., 2007). On a néanmoins constaté une grande hétérogénéité entre les différents sujets, les scores de classification allant de 40 % à presque 90 %. Ceci suggère que l'habileté intrinsèque d'un sujet à contrôler une ICM est primordiale, essentiellement lors des premières utilisations. Les conditions d'évaluation de performances présentées ici sont proches des conditions réelles. Les performances obtenues dans le cas présent sont donc très bonnes compte tenu du nombre de classes proposées. Le paradigme présenté ici est relativement contraint puisque le sujet est forcé à envoyer sa commande à des moments précis. La procédure mise en œuvre dans cette partie est néanmoins relativement générale et peut être étendue à des tâches autres que l'imagerie motrice et à des paradigmes moins contraignant (voir les parties suivantes). Enfin, l'hétérogénéité des résultats de chaque méthode pour chacun des sujets suggère qu'il est important de proposer un ensemble varié de méthodes de filtrage spatial afin que chacun d'eux puisse trouver celle qui lui « convient ». Les coûts de calcul engendrés par les méthodes d'ACI sont généralement considérés comme relativement importants.

Or la méthode présentée ici ne souffre pas d'un tel inconvénient puisque la diagonalisation conjointe est réalisée de manière presque instantanée pour des tailles de matrices de covariance jusqu'à $N_s = 200$. Il est même possible d'envisager un réapprentissage en temps réel des filtres spatiaux sans que le sujet n'ait à faire de pause. La bibliothèque mise à disposition `jointDiag` (fonctionnant avec R) a contribué à rendre les méthodes de diagonalisation jointe efficaces (voir l'[annexe B](#)).

Outre les diverses méthodes présentées dans cette partie visant à améliorer les performances actuelles des ICM, il nous paraît important de souligner l'aspect modulaire et extensible du cadre présenté. Les différentes méthodes de filtrage qui ont ainsi pu être mises en place ne sont que l'extension naturelle du cadre à différents modèles adaptés à l'ensemble de tâches considérées. Si la plupart des systèmes utilisés jusqu'à présent dans les conditions réelles est basée sur deux classes, nous espérons que le cadre présenté ici va conduire à une diversification des paradigmes et engendrer une augmentation du nombre de paradigmes multi-classes. Trois extensions naturelles basées sur la DCA ont été présentées dans ce chapitre. La première et la plus simple, que nous avons appelée DCA, ne considère que la diversité entre différentes tâches comme principe de base de l'extraction de sources. Dans la deuxième, appelée DCA-MS, la diversité issue des différentes tâches est associée à une diversité à l'intérieur de chaque tâche. Si DCA n'a pas significativement conduit à de meilleures performances que celles obtenues à l'aide de l'extension multi-classes du CSP, nous avons cependant pu montrer que DCA-MS permettait d'augmenter significativement (comparativement au CSP) les scores de classification autant dans le cas de la validation croisée que dans le cas des performances de transfert entre session. Afin de montrer la facilité d'extension de ce cadre à d'autres modèles encore plus plausibles physiologiquement, nous avons enfin présenté l'extension DCA-FR qui vise à utiliser à la fois la diversité entre les tâches et également la spécificité fréquentielle de chacune d'entre elle. Bien que nous n'ayons pas fourni d'évaluation quantitative de cette dernière méthode, une augmentation de près de 1% des taux de classification dans le cas du transfert entre sessions semble la rendre intéressante. Néanmoins l'augmentation de la complexité qu'elle engendre ne justifie peut-être pas toujours sa mise en place. Il est également possible de considérer l'utilisation des trois diversités présentées ici de manière conjointe.

9 Conclusion

En résumé, nous avons présenté ici un cadre théorique simple pour augmenter les scores de classification. Ce cadre est particulièrement utile dans le cas de paradigmes multi-classes. Les différents principes de séparation qui ont été présentés dans ce chapitre ont conduit à des améliorations substantielles pour certains sujets et plus modestes pour d'autres. La généralité de ce cadre fournit néanmoins un moyen simple et efficace de développer des nouveaux modèles d'extraction de sources non-stationnaires. Or cette flexibilité est importante pour les différents paradigmes d'ICMs. Les performances des méthodes ne semblent en effet pas homogènes d'un sujet à l'autre et la possibilité d'essayer simplement un grand nombre de méthodes pour chacun des sujets pourrait être un moyen efficace dans les conditions réelles pour prendre en compte la diversité des signaux cérébraux acquis. Le [chapitre IV](#) pourra venir en complément de ce présent chapitre en fournissant une méthode de filtrage spatial linéaire pour le débruitage des signaux EEG. Bien que le cadre présenté ici n'ait été appliqué qu'au cas de l'imagerie motrice, il est possible de considérer n'importe quelle tâche cognitive. Enfin, nous utiliserons ce cadre dans le cas de paradigmes moins contraints dans les [chapitres V](#) et [VI](#).

10 Références du chapitre

Afsari, B. (2008). Sensitivity analysis for the problem of matrix joint diagonalization. *SIAM J. Matrix Anal. Appl.*, 30(3):1148–1171.

- Baron, J. et Li, Y. (2003). Notes on the use of R for psychology experiments and questionnaires. Rapport technique, University of Pennsylvania.
- Belouchrani, A., Abed-Meraim, K., Cardoso, J.-F. et Moulines, E. (1997). A blind source separation technique using second-order statistics. *IEEE Trans. Signal Process.*, 45(2):434–444.
- Birbaumer, N. et Cohen, L. G. (2007). Brain-computer-interfaces (BCI): Communication and restoration of movement in paralysis. *J. Physiol.*, 579(3):621–636.
- Blankertz, B., Dornhege, G., Krauledat, M., Müller, K.-R. et Curio, G. (2007). The non-invasive berlin brain-computer interface: fast acquisition of effective performance in untrained subjects. *Neuroimage*, 37(2):539–550.
- Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M. et Müller, K.-R. (2008). Optimizing spatial filters for robust EEG Single-Trial analysis. *IEEE Signal Process. Mag.*, 25(1):41–56.
- Brunner, C., Naeem, M., Leeb, R., Graimann, B. et Pfurtscheller, G. (2007). Spatial filtering and selection of optimized components in four class motor imagery EEG data using independent components analysis. *Pattern Recognit. Lett.*, 28(8):957–964.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46.
- Congedo, M., Gouy-Pailler, C. et Jutten, C. (2008a). On the blind source separation of human electroencephalogram by approximate joint diagonalization of second order statistics. *Clin. Neurophysiol.*, 119(12):2677–2686.
- Darmois, G. (1947). Analyse des liaisons de probabilité. Dans *Proc. Internat. Statist. Conferences*, pages 231–240.
- Decety, J. (1996a). The neurophysiological basis of motor imagery. *Behav. Brain Res.*, 77(1-2):45–52.
- Decety, J. (1996b). Do imagined and executed actions share the same neural substrate?. *Brain Res. Cogn. Brain Res.*, 3(2):87–93.
- Dornhege, G., Blankertz, B., Curio, G. et Müller, K.-R. (2004). Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms. *IEEE Trans. Biomed. Eng.*, 51(6):993–1002.
- Dornhege, G., Blankertz, B., Krauledat, M., Losch, F. et Curio, G. et al. (2006). Combined optimization of spatial and temporal filters for improving brain-computer interfacing. *IEEE Trans. Biomed. Eng.*, 53(11):2274–2281.
- Ehrsson, H. H., Geyer, S. et Naito, E. (2003). Imagery of voluntary movement of fingers, toes, and tongue activates corresponding body-part-specific motor representations. *J. Neurophysiol.*, 90(5):3304–3316.
- Everitt, B. S. (2005). *An R and S-Plus Companion to Multivariate Analysis*. Springer, Première édition.
- Fukunaga, K. (1990). *Introduction to statistical pattern recognition (2nd ed.)*. Academic Press Professional, Inc., San Diego, CA, USA.
- Gray, R. M. (2006). Toeplitz and circulant matrices: A review. *Foundations and Trends in Communications and Information Theory*, 2(3):155–239.
- Grosse-Wentrup, M. et Buss, M. (2008). Multiclass common spatial patterns and information theoretic feature extraction. *IEEE Trans. Biomed. Eng.*, 55(8):1991–2000.
- Jankelowitz, S. K. et Colebatch, J. G. (2002). Movement-related potentials associated with self-paced, cued and imagined arm movements. *Exp. Brain Res.*, 147(1):98–107.
- Jutten, C. et Héroult, J. (1991). Blind separation of sources, part 1: an adaptive algorithm based on neuromimetic architecture. *Signal Process.*, 24(1):1–10.
- Krauledat, M., Tangermann, M., Blankertz, B. et Müller, K.-R. (2008). Towards zero training for brain-computer interfacing. *PLoS ONE*, 3(8):e2967.
- Lemm, S., Blankertz, B., Curio, G. et Müller, K.-R. (2005). Spatio-spectral filters for improving the classification of single trial EEG. *IEEE Trans. Biomed. Eng.*, 52(9):1541–1548.

- Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F. et Arnaldi, B. (2007a). A review of classification algorithms for EEG-based brain-computer interfaces. *J. Neural Eng.*, 4(2):R1–R13.
- Lotze, M. et Halsband, U. (2006). Motor imagery. *J. Physiol. Paris*, 99(4-6):386–395.
- Matsuoka, K., Ohya, M. et Kawamoto, M. (1995). A neural net for blind separation of nonstationary signals. *Neural Netw.*, 8(3):411–419.
- Michelon, P., Vettel, J. M. et Zacks, J. M. (2006). Lateral somatotopic organization during imagined and prepared movements. *J. Neurophysiol.*, 95(2):811–822.
- Müller-Gerking, J., Pfurtscheller, G. et Flyvbjerg, H. (1999). Designing optimal spatial filters for single-trial EEG classification in a movement task. *Clin. Neurophysiol.*, 110(5):787–798.
- Naeem, M., Brunner, C., Leeb, R., Graimann, B. et Pfurtscheller, G. (2006). Separability of four-class motor imagery data using independent components analysis. *J. Neural Eng.*, 3(3):208–216.
- Nijholt, A., Tan, D., Pfurtscheller, G., Brunner, C. et Millán, J. d. R. et al. (2008). Brain-computer interfacing for intelligent systems. *IEEE Intell. Syst.*, 23(3):72–79.
- Parra, L. et Spence, C. (2001). *Independent Components Analysis, Principles and Practice*, chapitre Separation of non-stationary natural signals, pages 135–157. Cambridge University Press.
- Pfurtscheller, G., Brunner, C., Schlögl, A. et Lopes da Silva, F. H. (2006). Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks. *Neuroimage*, 31(1):153–159.
- Pfurtscheller, G. et Neuper, C. (2001). Motor imagery and direct brain-computer communication. *Proc. IEEE*, 89(7):1123–1134.
- Pham, D.-T. (2001). Blind separation of instantaneous mixture of sources via the gaussian mutual information criterion. *Signal Process.*, 81(4):855–870.
- Pham, D.-T. et Cardoso, J.-F. (2001). Blind separation of instantaneous mixtures of nonstationary sources. *IEEE Trans. Signal Process.*, 49(9):1837–1848.
- Souloumiac, A. (2009). Non-orthogonal joint diagonalization by combining Givens and hyperbolic rotations. *IEEE Trans. Signal Process.*, 57(6):2222–2231.
- Tichavsky, P. et Yeredor, A. (2009). Fast approximate joint diagonalization incorporating weight matrices. *IEEE Trans. Signal Process.*, 57(3):878–891.
- Vigário, R., Särelä, J., Jousmäki, V., Hämäläinen, M. et Oja, E. (2000). Independent component approach to the analysis of EEG and MEG recordings. *IEEE Trans. Biomed. Eng.*, 47(5):589–593.
- Vollgraf, R. et Obermayer, K. (2006). Quadratic optimization for simultaneous matrix diagonalization. *IEEE Trans. Signal Process.*, 54(9):3270–3278.
- Wang, H. et Zheng, W. (2008). Local temporal common spatial patterns for robust single-trial EEG classification. *IEEE Trans. Neural Syst. Rehabil. Eng.*, 16(2):131–139.
- Wang, Y., Zhang, Z., Li, Y., Gao, X. et Gao, S. et al. (2004b). BCI competition 2003–data set IV: an algorithm based on CSSD and FDA for classifying single-trial EEG. *IEEE Trans. Biomed. Eng.*, 51(6):1081–1086.
- Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G. et Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clin. Neurophysiol.*, 113(6):767–791.
- Wu, W., Gao, X. et Gao, S. (2005). One-Versus-the-Rest (OVR) algorithm: An extension of common spatial patterns (CSP) algorithm to multi-class case. Dans *Conf. Proc. IEEE. Eng. Med. Biol. Soc.*, pages 2387–2390.
- Ziehe, A., Laskov, P., Nolte, G. et Müller, K.-R. (2004). A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind source separation. *J. Mach. Learn. Res.*, 5(Jul):777–800.
- Jahanshahi, M. et Hallett, M., éditeurs (2003). *The Bereitschaftspotential: Movement-Related Cortical Potential* Kluwer Academic/Plenum Publishers, New-York, USA.

Kandel, E. R. et Schwartz, J. H., éditeurs (1985). *Principles of Neural Science* Elsevier, New-York, Second édition.

Chapitre IV

ELIMINATION DES ARTEFACTS OCULAIRES



1	Introduction	84
2	Méthode	86
2.1	Détection des périodes de contamination	86
2.2	Transformation linéaire	88
2.3	Séparation entre le signal et le bruit	89
2.4	Extension itérative	90
3	Évaluation à partir de données simulées	91
3.1	Génération des données	91
3.2	Méthode d'évaluation	92
3.3	Résultats	93
4	Données réelles	93
4.1	Détection des périodes d'activité oculaire	94
4.2	Projection linéaire	95
4.3	Signaux débruités	96
5	Application : classification de signaux non-bruités	97
5.1	Données et méthode	97
5.2	Résultats	99
6	Discussion et conclusion	100
6.1	De l'importance de débruiter les données	100
6.2	Implémentation de la méthode	100
6.3	Conclusion	101
7	Références du chapitre	102

UNE SECONDE APPROCHE de filtrage spatial est abordée dans ce chapitre. Le cadre général de cette méthode a été initialement proposé par Dr. Reza Sameni (ancien doctorant de GIPSA-lab) pour le débruitage et l'extraction de signaux cardiaques du fœtus. Les travaux présentés dans ce chapitre sont issus de la collaboration avec Dr. Sameni et ont pour but le débruitage des signaux électroencéphalographiques (EEG) contaminés par des artefacts oculaires. Le critère qui est utilisé pour ce débruitage est innovant et la méthode qui en résulte s'avère compétitive par rapport à certaines méthodes couramment utilisées. Le débruitage des signaux est un problème très important pour toutes les applications utilisant l'EEG, nous l'abordons ici selon le point de vue des Interfaces Cerveau-Machines (ICMs). Basée sur des principes mathématiques relativement simples, cette partie propose une approche didactique du problème du débruitage par une méthode de filtrage spatial adaptée au problème.

L'application du cadre théorique proposé par le Dr. Reza Sameni au problème du débruitage des signaux EEG est originale et vient compléter de manière très spécifique ce cadre par des extensions adaptées à notre problème. Nous proposons deux déclinaisons de la méthode pour le débruitage adapté et mettons en place un critère statistique pour détecter les moments de non-activité oculaire. Nous montrons que cette méthode s'avère extrêmement compétitive en comparaison des méthodes existantes bien que l'augmentation de performances qu'elle permet d'obtenir dans le cas des ICMs est ici relativement faible dans le cas des données du [chapitre III](#).

1 Introduction

L'électroencéphalographie (EEG) est l'une des techniques la plus utilisée pour analyser et interpréter l'activité électrique cérébrale de l'homme. Comme nous l'avons vu dans les chapitres précédents, on utilise souvent des mesures spectrales ou topographiques qui reflètent l'activité du cerveau. Or ces mesures sont dans le cas de l'EEG, particulièrement contaminées par des signaux électriques extérieurs au cerveau. Ces interférences peuvent rendre l'interprétation des signaux difficiles. Ce problème est récurrent en particulier dans le contexte des ICMs puisque la classification des signaux afin d'inférer la tâche mentale réalisée par le sujet peut être très perturbée par des artefacts extérieurs ([Thulasidas et al., 2004](#) et [Boudet, 2008](#)).

Parmi les différents types d'interférences qui peuvent troubler la mesure de l'activité cérébrale, les artefacts oculaires sont ceux qui ont la plus grande amplitude (voir la [figure IV.1](#)). On peut les observer lors du clignement des yeux, de saccades oculaires ou encore d'un mouvement volontaire du regard pour observer un élément dans le champ visuel du sujet. Leurs caractéristiques respectives ainsi que la manière dont ils contaminent l'EEG sont différentes selon le type d'artefact oculaire. On attribue par exemple au clignement des yeux un changement de conductance en raison du mouvement de la paupière sur la cornée. Ceci génère un signal transitoire visible essentiellement sur les électrodes frontales et pariétales dont l'amplitude peut atteindre dix fois celle des signaux EEG. Les mouvements oculaires, quant à eux, peuvent être volontaires (lents), saccadés ou encore constitués de micro-saccades. On les attribue aux champs dipolaires formés par le couple cornée/rétine. Les saccades oculaires sont les mouvements les plus rapides produits par le corps humain et peuvent atteindre des vitesses angulaires de l'ordre de $1000^\circ/\text{s}$. Une saccade oculaire en réponse à un événement imprévu met en moyenne environ 200 ms pour s'initier et dure ensuite entre 20 et 200 ms. Chacun de ces artefacts est caractérisé par une grande amplitude qui contamine un ensemble non négligeable d'électrodes. Ces phénomènes transitoires sont considérés comme les artefacts les plus importants dans les études EEG ([Romero et al., 2008](#) et [Romero et al., 2009](#)). Afin d'enregistrer l'activité oculaire, on utilise souvent l'électrooculographie (EOG) qui consiste en un ensemble d'électrodes placées autour des yeux destinées à enregistrer les différents dipôles électriques produits par les activités oculaires. Dans le meilleur des cas, on emploie jusqu'à trois électrodes EOG. La première dans le coin gauche de l'œil gauche, la deuxième à la base du front entre les deux sourcils et la dernière à droite de l'œil droit.

Une manière courante d'enlever les artefacts oculaires des enregistrements EEG est d'utiliser l'Analyse en Composantes Indépendantes (ACI) ([Delorme et al., 2007](#) et [Boudet, 2008](#)) à partir des données multivariées EEG. L'idée est d'appliquer une méthode d'ACI puis d'enlever des signaux les composantes les plus corrélées avec un canal EOG de référence ([Ille et al., 2002](#), [Ille, 2001](#), [Jung et al., 2000a](#) et [Vigário et al., 2000](#)). Cependant, il n'est pas toujours aisé d'associer les composantes extraites par ACI avec l'EOG de manière automatique et non-supervisée. De plus, les mesures EEG peuvent être relativement bruitées, or l'ACI est basée sur une mesure d'indépendance (et non sur une mesure de la propriété du signal), par conséquent le bruit présent dans l'EEG peut être amplifié

Données contaminées par des artefacts oculaires

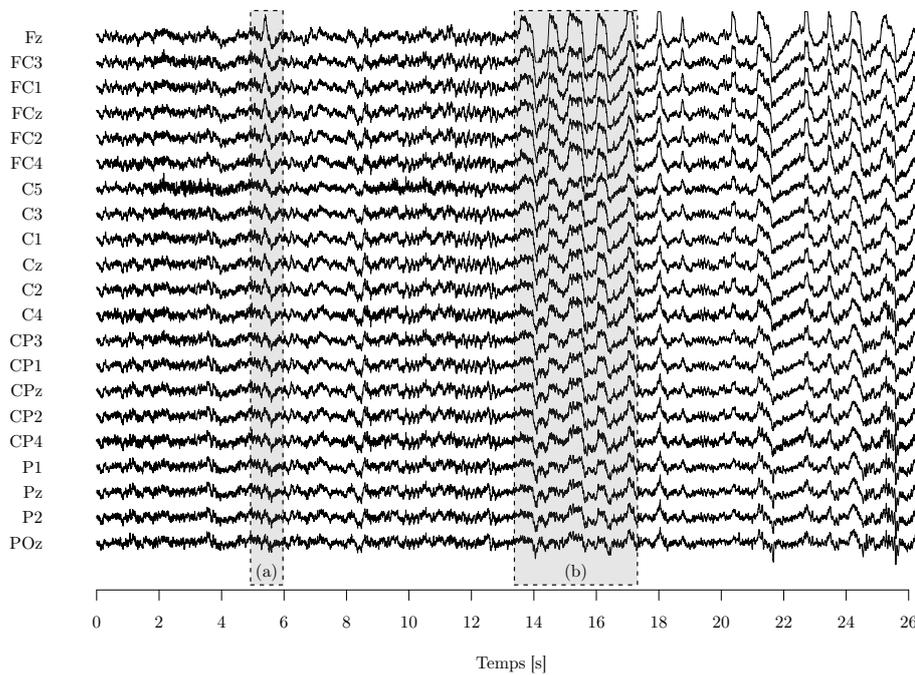


Figure IV.1 Exemple d'enregistrement EEG comportant 21 capteurs. Les signaux sont contaminés par des artefacts oculaires (le sujet devait bouger les yeux). (a) on voit une contamination par un clignement d'yeux. (b) ces signaux de grande amplitude sont caractéristiques d'un mouvement brusque des yeux.

par cette méthode. Enfin, la plupart des méthodes d'ACI est aveugle au bruit Gaussien. Celui-ci peut donc être largement dispersé dans l'ensemble des canaux EEG avec une telle méthode.

Une autre méthode courante pour enlever les artefacts oculaires des enregistrements EEG est d'utiliser une approche basée sur la soustraction (Romero et al., 2008). L'idée est alors d'utiliser les enregistrements EOG pour enlever les artefacts oculaires de l'EEG en utilisant une simple soustraction. Cette méthode vise donc seulement à trouver le facteur de projection de l'EOG sur les canaux EEG pour enlever le signal EOG mis à la bonne échelle. Cependant il n'y a aucune raison pour que le signal enregistré au niveau des yeux soit complètement exempt de tout signal EEG. Ainsi, en enlevant le signal EOG il se peut que l'on enlève également une partie intéressante de l'EEG.

Dans des travaux récents (Sameni et al., 2008a, Sameni, 2008 et Sameni et al., 2009), les auteurs ont utilisé la décomposition en valeur propres généralisée (Golub et van Loan, 1996) afin de séparer les signaux cardiaques d'une mère et de son fœtus à partir des enregistrements effectués au niveau de la paroi abdominale de la mère. Dans ces travaux, l'avantage de la décomposition en valeurs propres généralisée sur des méthodes telles que l'ACI était de fournir un ensemble de filtres spatiaux ordonnés selon un critère imposé. Ainsi, cette méthode pouvait être utilisée de manière non-supervisée et automatique. Une des forces de cette méthode résidait également dans la possibilité d'appliquer une transformation non-linéaire à certaines des composantes (pour les extraire ou les débiter). Le rang des données pouvait ainsi être conservé et on pouvait envisager d'appliquer récursivement la méthode sans que le rang du résultat n'en soit altéré. Dans le chapitre qui suit, nous allons développer une idée similaire, fondée également sur la décomposition en valeurs propres généralisée, pour la détection automatique et le nettoyage des artefacts oculaires dans les enregistrements EEG multicanaux.

2 Méthode

De même que dans les chapitres précédents, les signaux EEG enregistrés au niveau de Ns capteurs sont notés $\mathbf{x}(t)$. Dans cette partie nous disposons également d'un signal EOG que nous noterons $\varphi(t)$.

2.1 Détection des périodes de contamination

Dans une première étape, nous cherchons à détecter les moments d'activité oculaire. En raison de la nature relativement abrupte des activités EOG, il est possible de réaliser cette détection de manière assez simple en utilisant le canal EOG $\varphi(t)$. Pour ce faire, définissons $\Phi(t)$ comme la puissance moyenne (ou la variance) du signal réel EOG $\varphi(t)$ pendant un intervalle de temps défini par une fenêtre de longueur w autour de t

$$\Phi(t) = \frac{1}{w} \sum_{\tau=-w/2}^{w/2} \varphi(t - \tau)^2. \quad (\text{IV.1})$$

Le problème de détection d'activité oculaire revient donc à déterminer un seuil ℓ qui permette de distinguer entre les deux hypothèses

- \mathcal{H}_0 : aucune activité oculaire détectée par l'EOG ;
- \mathcal{H}_1 : période d'activité oculaire ;

en considérant le test simple suivant

$$\Phi(t) \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{>}} \ell \quad (\text{IV.2})$$

pour chaque temps t . Selon les jeux de données considérés, plusieurs cas de figure se présentent alors à nous.

Dans la suite de ce chapitre, nous appellerons $\Phi_{\mathcal{H}_0}(t)$ (respectivement $\Phi_{\mathcal{H}_1}(t)$) la puissance moyenne du signal EOG $\varphi_{\mathcal{H}_0}(t)$ (respectivement $\varphi_{\mathcal{H}_1}(t)$) lorsque que le sujet n'a pas d'activité oculaire (respectivement lorsque le sujet est en période d'activité oculaire). Afin de simplifier les notations, nous confondrons par la suite les variables aléatoires avec les signaux, réalisations de ces variables aléatoires.

2.1.1 Méthode semi-automatique

L'observation du signal $\varphi(t)$ pendant une courte période permet généralement de fixer ℓ de manière satisfaisante. Notons de plus qu'une légère imprécision sur le choix de ℓ ne porte pas préjudice à la méthode dans la mesure où l'estimation pourra ensuite être affinée à l'aide de la procédure récursive.

2.1.2 Information sur $\Phi_{\mathcal{H}_0}(t)$

Dans certains cas de figure que nous avons rencontrés au cours de cette thèse, les paradigmes ICMs incluaient une période initiale durant laquelle le sujet devait fermer les yeux. Si l'on suppose que $\varphi_{\mathcal{H}_0}(t) \sim \mathcal{N}(0, \sigma_{\varphi_0}^2)$ alors on peut sans difficulté estimer la variance σ_{φ}^2 de la loi de φ (méthode des moments, maximum de vraisemblance ou méthode Bayésienne) lorsque le sujet n'a pas d'activité oculaire. De plus on sait que, sous l'hypothèse d'indépendance des $\varphi(t)$

$$\sum_{-w/2}^{w/2} \frac{\varphi_{\mathcal{H}_0}(t)^2}{\hat{\sigma}_{\varphi_0}^2} \sim \chi_w^2 \quad \text{ou encore} \quad \sum_{-w/2}^{w/2} \frac{\varphi_{\mathcal{H}_0}(t)^2}{\hat{\sigma}_{\varphi_0}^2} \sim \mathcal{G}\left(\frac{w}{2}, 2\right) \quad (\text{IV.3})$$

où χ_w^2 est la distribution « Chi square » à w degrés de liberté et $\mathcal{G}(w/2, 2)$ est la distribution \mathcal{G} (Gamma) de paramètres de forme $w/2$ et d'échelle 2. Enfin, par les propriétés de la loi \mathcal{G} , on obtient la loi de $\Phi(t)$

$$\Phi_{\mathcal{H}_0}(t) \sim \mathcal{G}\left(\frac{w}{2}, \frac{2\hat{\sigma}_{\varphi_0}^2}{w}\right). \quad (\text{IV.4})$$

Cette loi nous permet donc de fixer le seuil ℓ de manière à contrôler l'erreur réalisée lorsqu'on décide \mathcal{H}_1 alors que le sujet n'avait aucune activité oculaire. On a en effet la probabilité de fausses détections, c'est-à-dire de décider \mathcal{H}_1 alors que \mathcal{H}_0 est vraie (erreur de première espèce), qui vaut

$$\mathbb{P}(\mathcal{H}_1 | \mathcal{H}_0) = \int_{\ell}^{\infty} x^{\frac{w}{2}-1} \frac{e^{-\frac{xw}{2\hat{\sigma}_{\varphi_0}^2}}}{\frac{2\hat{\sigma}_{\varphi_0}^2}{w} \Gamma\left(\frac{w}{2}\right)} dx = 1 - \int_{-\infty}^{\ell} x^{\frac{w}{2}-1} \frac{e^{-\frac{xw}{2\hat{\sigma}_{\varphi_0}^2}}}{\frac{2\hat{\sigma}_{\varphi_0}^2}{w} \Gamma\left(\frac{w}{2}\right)} dx, \quad (\text{IV.5})$$

où Γ est la fonction Gamma. L'expression IV.5 donne donc une manière numérique (la plupart des logiciels de calcul sont capables de donner des approximations de l'intégrale) de définir ℓ en fonction de $\mathbb{P}(\mathcal{H}_1 | \mathcal{H}_0)$ et de la variance estimée du processus.

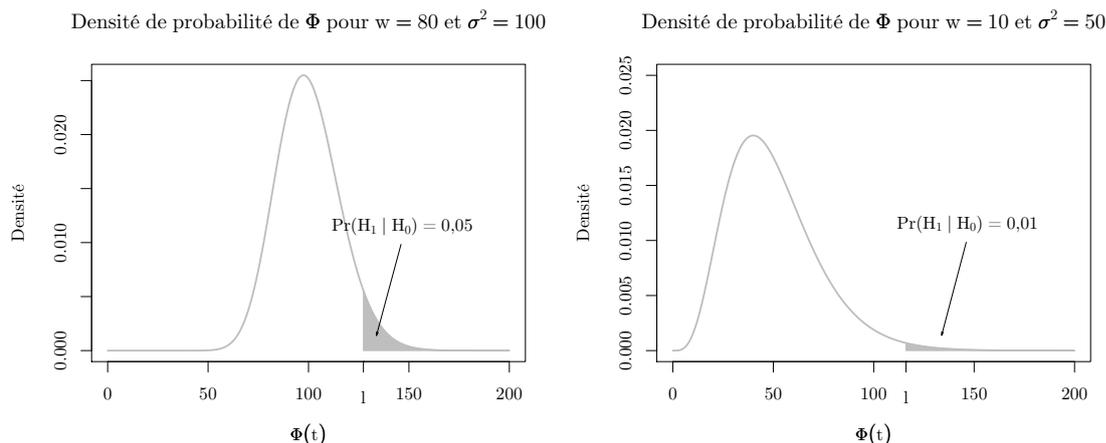


Figure IV.2 Distributions de probabilités de $\Phi(t)$ dans deux conditions distinctes. Sur les deux graphes, on a également indiqué le seuil ℓ qui permet d'obtenir des taux d'erreurs respectifs de 0.05 à gauche et de 0.01 à droite. Ce taux est égal à l'aire sous la courbe entre ℓ et ∞ . Cette aire est représentée de manière grisée sur chacune des figures.

2.1.3 Cas général

Dans le cas où aucune information pertinente ne peut être utilisée pour déterminer la loi de $\Phi(t)$ sous \mathcal{H}_0 , alors on doit faire des hypothèses supplémentaires afin de modéliser les lois de $\Phi(t)$ sous \mathcal{H}_0 et $\Phi(t)$ sous \mathcal{H}_1 . Il est relativement rare en pratique que nous ne puissions pas nous ramener au cas précédent. Cependant, dans le cas où l'on doit modéliser $\Phi(t)$ sous \mathcal{H}_1 de manière explicite, on peut alors utiliser à nouveau les calculs de la section précédente en supposant que $\varphi_{\mathcal{H}_1}(t) \sim \mathcal{N}(0, \sigma_{\varphi_1}^2)$. On peut également envisager d'utiliser une autre loi de probabilité. Considérant que ce cas est peu probable, nous ne rentrerons pas dans les détails d'une telle approche.

2.1.4 Résumé

Dans l'ensemble des trois cas précédents, la procédure utilisée, ainsi qu'un test du rapport des vraisemblances entre les deux hypothèses, doit permettre de déterminer les périodes d'activité oculaire, c'est-à-dire l'ensemble

$$t_a = \{t \in \mathbb{R} \mid \Phi(t) > \ell\}. \quad (\text{IV.6})$$

2.2 Transformation linéaire

Après avoir défini les périodes d'activité oculaire, nous cherchons à présent une transformation linéaire des signaux EEG, $\mathbf{x}(t)$, de telle sorte que le résultat de cette transformation ressemble le plus possible à l'EOG (dans le sens où la puissance des signaux extraits est concentrée pendant les périodes d'activité oculaire). Si nous appelons \mathbf{w} cette transformation linéaire et $y(t) = \mathbf{w}^T \mathbf{x}(t)$, le résultat de la transformation linéaire, alors on peut écrire mathématiquement la fonction de coût correspondant à notre but par

$$\zeta(\mathbf{w}) = \frac{\mathbb{E}_{t_a} \{y^2(t_a)\}}{\mathbb{E}_t \{y^2(t)\}} \quad (\text{IV.7})$$

où $\mathbb{E}_t \{.\}$ représente la moyenne sur la période d'observation totale tandis que $\mathbb{E}_{t_a} \{.\}$ représente la moyenne sur la période définie par t_a telle que définie dans l'expression IV.6. L'idée sous-jacente à cette fonction de coût est de trouver les combinaisons linéaires des signaux d'entrée $\mathbf{x}(t)$ dont l'énergie est maximale durant les périodes d'activité oculaire tout en minimisant globalement l'énergie de chaque signal de sortie. En utilisant la définition de $y(t)$, l'équation IV.7 peut alors être transformée en

$$\zeta(\mathbf{w}) = \frac{\mathbf{w}^T \mathbb{E}_{t_a} \{\mathbf{x}(t_a)\mathbf{x}(t_a)^T\} \mathbf{w}}{\mathbf{w}^T \mathbb{E}_t \{\mathbf{x}(t)\mathbf{x}(t)^T\} \mathbf{w}}, \quad (\text{IV.8})$$

qui forme un quotient de Rayleigh (Strang, 1988 et Golub et van Loan, 1996) et peut être optimisée en réalisant la diagonalisation jointe de deux matrices : d'une part la matrice de covariance des canaux EEG sur la totalité de la période considérée, d'autre part la matrice de covariance des mêmes données mais uniquement pendant les périodes d'activité oculaire. Chacune d'elles est définie respectivement par

$$C_x = \mathbb{E}_t \{\mathbf{x}(t)\mathbf{x}(t)^T\}, \quad (\text{IV.9})$$

$$A_x = \mathbb{E}_{t_a} \{\mathbf{x}(t_a)\mathbf{x}(t_a)^T\}. \quad (\text{IV.10})$$

Une interprétation de cette étape est que l'on cherche à décorréler les composantes de manière *globale* aussi bien que *locale*. La décorrélation globale s'apparente à un blanchiment. De plus, en diagonalisant également A_x , on s'assure que les composantes seront aussi décorrélées localement pendant les périodes d'activité oculaire. On impose ainsi que les composantes extraites n'aient pas de redondance (au moins jusqu'aux statistiques du deuxième ordre).

La matrice W maximisant la fonction de coût de l'expression IV.8 diagonalise la paire de matrices (A_x, C_x) , elle est obtenue par la méthode de la décomposition en valeurs propres généralisée

$$\begin{cases} W A_x W^T = \Lambda \\ W C_x W^T = I \end{cases} \quad (\text{IV.11})$$

où Λ est une matrice diagonale contenant les valeurs propres sur sa diagonale. Celles-ci peuvent donc être ordonnées de manière décroissante, traduisant la qualité de chaque vecteur propre généralisé associé selon le critère IV.8.

Les projections linéaires sont alors obtenues par une simple multiplication matricielle

$$\mathbf{y}(t) = W^T \mathbf{x}(t), \quad (\text{IV.12})$$

dont les éléments de \mathbf{y} correspondent aux transformations linéaires des données originelles $\mathbf{x}(t)$, ordonnées selon leur degré de ressemblance avec l'EOG lors des périodes d'activité oculaire. Ceci signifie que $y_1(t)$ est la composante qui contient le plus d'information du point de vue de l'activité oculaire tandis que $y_{N_s}(t)$ est la composante qui en contient le moins. Notons enfin que $y_1(t)$ correspond à la valeur propre de Λ la plus grande tandis que $y_{N_s}(t)$ correspond à la valeur propre de Λ la plus petite.

2.3 Séparation entre le signal et le bruit

L'étape suivante consiste à enlever les artefacts oculaires des signaux \mathbf{y} . Cette opération consiste à considérer l'ensemble de signaux \mathbf{y} comme composé d'une partie contenant le signal EEG et d'une partie contenant les artefacts oculaires

$$\mathbf{y}(t) = \underbrace{\mathbf{y}_s(t)}_{\text{signal EEG}} + \underbrace{\mathbf{y}_n(t)}_{\text{artefacts oculaires}} \quad (\text{IV.13})$$

Nous présentons ici deux méthodes qui permettent de réaliser ce débruitage.

2.3.1 Mise à zéro

Or on sait, d'après les traitements effectués dans les parties précédentes, que les premières composantes de \mathbf{y} sont celles qui contiennent le plus d'artefacts oculaires. On peut donc considérer que la [décomposition IV.13](#) s'écrit sous la forme

$$\begin{pmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_{N_s}(t) \end{pmatrix} = \begin{pmatrix} \approx 0 \\ \approx 0 \\ \vdots \\ y_{N_s,s}(t) \end{pmatrix} + \begin{pmatrix} y_{1,n}(t) \\ y_{2,n}(t) \\ \vdots \\ \approx 0 \end{pmatrix} \quad (\text{IV.14})$$

Une méthode simple et efficace de débruiter \mathbf{y} consiste donc à mettre à zéro les M premières composantes. L'idée de cette méthode est d'utiliser l'ordre fourni par la décomposition propre généralisée pour éliminer uniquement les composantes les plus contaminées (les premières). Le choix du nombre de composantes à éliminer dépend alors des données, nous avons observé dans plusieurs cas qu'une seule était suffisante pour débruiter de manière très satisfaisante. Ce nombre est lié à la dimension intrinsèque du sous-espace engendré par les signaux EOG. On peut résumer la transformation que met en place cette étape de manière mathématique par

$$z_i(t) = \begin{cases} 0 & \text{si } i \leq M \\ y_i(t) & \text{si } i > M \end{cases} \quad (\text{IV.15})$$

où $z_i(t)$ représente l'ensemble des signaux résultant de la transformation. De cette façon, $\mathbf{z}(t)$ contient la partie $\mathbf{y}_s(t)$ telle que définie dans l'[équation IV.13](#). Notons que, de cette transformation, résulte une perte de M du rang des données. Cette réduction peut être évitée en définissant une transformation plus spécifique au type d'artefact recherché. On peut par exemple envisager d'utiliser une transformation en ondelettes pour identifier et éliminer les atomes temps-fréquences les plus contaminés.

2.3.2 Débruitage par ondelettes

On peut en effet utiliser une méthode plus fine pour détecter les artefacts oculaires présents dans les premières composantes extraites par décomposition en valeurs propres

généralisée. Nous allons à nouveau utiliser uniquement les premières composantes mais allons enlever de celles-ci seulement les signaux oculaires. La procédure repose sur le fait que les signaux oculaires ont une représentation dans le domaine des ondelettes qui est relativement parcimonieuse. Par conséquent, la plupart des coefficients en ondelettes d'un signal oculaire pur peut être fixée à zéro sans que l'information contenue dans le signal n'en soit altérée outre mesure. La procédure employée s'appuie donc sur la transformée en ondelettes discrète de la composante considérée (Mallat, 1999). En se basant sur la propriété (Lindgren et al., 1983)

$$X \sim \mathcal{N}(0, 1) \quad \implies \quad \lim_{n \rightarrow \infty} \mathbb{P} \left(\left\{ \max_{i \in [1..n]} |x_i| < \sqrt{2 \log(n)} \right\} \right) = 1, \quad (\text{IV.16})$$

où $\{x_i\}$ sont les réalisations de la variable aléatoire X , on fixe un seuil (Donoho et Johnstone, 1995)

$$\tau = \sqrt{2 \log(n)} \sigma, \quad (\text{IV.17})$$

où σ est la racine de la variance estimée du processus. Cette procédure élimine, jusqu'à une certaine échelle d'ondelette (typiquement l'échelle maximale moins 3) tous les coefficients négligeable de la représentation. La transformée inverse des coefficients seuillés (Mallat, 1999) donne alors le signal oculaire pur qui est soustrait de la composante original (dans cet procédure, notons que le signal EEG est considéré comme le bruit alors que le signal oculaire est le signal d'intérêt). De même que dans le cas de la mise à zéro des M premières composantes de la section précédente, la procédure de débruitage par seuillage des coefficients en ondelette est appliquée au M premières composantes extraites par décomposition en valeurs propres généralisée.

2.3.3 Retour à l'espace des capteurs

Les signaux multivariés \mathbf{z} représentent donc la partie débruitée du signal original. Cependant l'espace de représentation de ces signaux (la base, ou l'interprétation de chaque dimension) n'est pas celle des capteurs originaux. Il peut donc être intéressant de transformer à nouveau les signaux \mathbf{z} afin de revenir dans l'espace originel avant de continuer les traitements. Il suffit, pour ce faire, d'utiliser la relation IV.12 en observant que la matrice W^T peut être inversée pour donner

$$\mathbf{x}(t) = (W^T)^{-1} \mathbf{y}(t). \quad (\text{IV.18})$$

Par conséquent, à partir de $\mathbf{z}(t)$, on peut revenir dans l'espace des capteurs en appliquant la transformation inverse

$$\mathbf{x}^{(1)}(t) = (W^T)^{-1} \mathbf{z}(t). \quad (\text{IV.19})$$

2.4 Extension itérative

Après avoir enlevé les composantes qui contenaient le plus d'artefacts et reprojété les signaux débruités résultant sur l'espace des capteurs, on peut améliorer encore les résultats en répétant la même procédure d'une manière récursive. On peut, lors de la deuxième itération, utiliser le signal $y_1(t)$ (qui est celui qui ressemble le plus à la composante oculaire du signal) afin de réestimer les périodes d'activité oculaire. En répétant cette procédure un certain nombre de fois, on peut alors obtenir une meilleure estimation des artefacts oculaires.

Notons que cette procédure est particulièrement intéressante puisqu'elle permet d'étendre la méthode au cas où aucun signal EOG de référence n'est disponible. L'idée est alors

de partir d'une estimation relativement restrictive des périodes d'activité oculaire, et ensuite de raffiner cette estimation au fur et à mesure des itérations. Dans ce cas, la première phase de détection des périodes d'activité oculaire peut être réalisée en utilisant un seuil de détection volontairement surestimé. On s'assure ainsi qu'on détecte uniquement les contaminations les plus fortes, qui correspondent aux périodes où l'activité oculaire est la plus puissante. La diagonalisation conjointe permet alors d'extraire la composante qui contient le plus d'activité oculaire. Cette première composante peut alors être utilisée à la place du signal de référence EOG afin d'itérer la méthode. Lors de la deuxième itération, l'algorithme fonctionne alors de la même manière que si un signal EOG était à notre disposition.

La [figure IV.3](#) résume de manière schématique l'ensemble des étapes de la méthode.

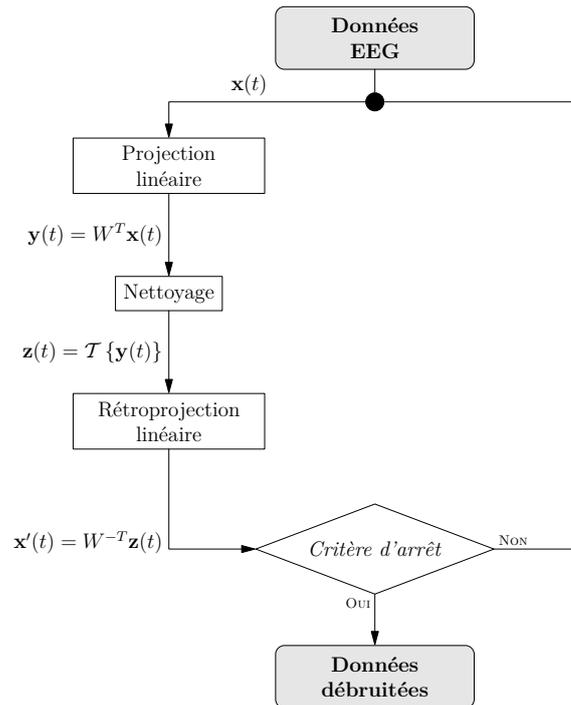


Figure IV.3 Représentation schématique résumant la procédure décrite. Les données EEG sont d'abord projetées, puis débruitées et enfin l'inverse de la première projection est appliquée afin que les signaux soient à nouveau dans l'espace des capteurs. La possibilité d'itérer ce processus un certain nombre de fois est également représenté sur la figure.

3 Évaluation à partir de données simulées

Afin d'évaluer notre méthode, nous comparons ses performances par rapport à une méthode très employée pour réaliser le débruitage des signaux EEG : FastICA ([Vigário et al., 2000](#), [Hyvärinen, 1999](#) et [Hyvärinen et al., 2001](#)). Dans cette première section, nous évaluons notre méthode à partir de signaux générés artificiellement. Ceci nous permet de quantifier de manière exacte les performances de notre méthode.

3.1 Génération des données

Afin de simuler un signal EEG relativement réaliste, nous considérons le modèle

$$\mathbf{x}(t) = \underbrace{\mathbf{x}_s(t)}_{\text{signal EEG « pur »}} + \beta \underbrace{\begin{pmatrix} k_1 \\ k_2 \\ \vdots \\ k_{N_s} \end{pmatrix}}_{\text{composante oculaire}} x_n(t) \quad (\text{IV.20})$$

dans lequel les différentes composantes sont générées de la manière suivante :

1. $\mathbf{x}_s(t)$: on utilise une portion de données réelles ne contenant aucun artefact pour entraîner un modèle autorégressif multivarié d'ordre $p = 8$ et de type Yule-Walker (Anderson et al., 1998). Le signal $\mathbf{x}_s(t)$ de longueur N_t est alors généré à partir du filtre autorégressif résultant et d'un bruit blanc Gaussien multivarié. Dans toute la suite de cette simulation, nous fixons $N_s = 4$ (nombres de capteurs) ;
2. $x_n(t)$: afin de générer le signal oculaire univarié, une portion d'un signal EOG décrivant le déroulement temporel d'un clignement d'œil est répétée de manière aléatoire afin de générer un signal oculaire de taille N_t . Mathématiquement, on peut considérer les clignements des yeux comme une succession d'événements sans mémoire¹, ainsi le temps entre deux clignements suit une loi exponentielle dont le paramètre est fixé à $\lambda = 0,1$ Hz. Le choix de ce paramètre revient à fixer le temps moyen entre deux occurrences puisque la moyenne d'une distribution exponentielle vaut $1/\lambda = 10$ s. A chaque occurrence de l'événement aléatoire dont le temps d'attente est décrit par la loi de poisson, on génère un Dirac, l'ensemble formant donc un peigne de Dirac dont l'espacement varie. Le signal $x_n(t)$ est alors généré en convoluant le signal d'un clignement « modèle » avec le peigne de Dirac.
3. β : ce paramètre réel est un facteur d'échelle du signal oculaire permettant d'ajuster le rapport signal sur bruit.
4. $\{k_i\}_{i \in [1..N_s]}$: vecteur caractérisant les différents gains du signal $x_n(t)$ au niveau de chacun des capteurs. Pour tout i , $k_i \sim \mathcal{N}(1; 0,3)$ (loi normale de moyenne 1 et de variance 0,3).

Notons qu'un modèle autorégressif multivarié a été préféré à un ensemble de modèles autorégressifs univariés, ce dernier ne permettant pas de capturer les dépendances spatiales entre les différents capteurs.

3.2 Méthode d'évaluation

La procédure présentée dans ce chapitre est comparée à une méthode utilisée couramment pour débruiter les signaux EEG : FastICA. Afin de réaliser cette comparaison, nous allons évaluer l'erreur résiduelle entre les signaux débruités par chacune des méthodes et les signaux originaux générés par le filtre autorégressif multivarié. Définissons pour cela les signaux représentant cette erreur résiduelle (les signaux indicés ICA sont relatifs à la méthode d'analyse en composantes indépendantes tandis que les signaux indicés GEVD sont relatifs à la méthode de ce chapitre - d'après Generalized Eigenvalue Decomposition, traduction anglaise de « décomposition en valeurs propres généralisée »)

¹ La probabilité d'attendre $s+t$ secondes entre deux clignements d'yeux sachant qu'on a déjà attendu t secondes est la même que la probabilité d'attendre s secondes à partir du début de la période d'attente. Ce modèle n'est bien sûr pas physiologiquement très plausible mais rend suffisamment bien compte des observations

$$\begin{aligned}\varepsilon_{\text{ICA}}(t) &= \mathbf{x}_{\text{ICA}}(t) - \mathbf{x}_s(t) \quad , \\ \varepsilon_{\text{GEVD}}(t) &= \mathbf{x}_{\text{GEVD}}(t) - \mathbf{x}_s(t) \quad ,\end{aligned}\tag{IV.21}$$

où $\mathbf{x}_{\text{ICA}}(t)$ représente le signal multivarié débruité par la méthode FastICA tandis que $\mathbf{x}_{\text{GEVD}}(t)$ représente le signal multivarié débruité par la procédure décrite au début de ce chapitre. L'erreur résiduelle est donc un signal multivarié de dimension N_s . On évalue ensuite la variance résiduelle du i -ème signal unidimensionnel de $\varepsilon_{i,\text{ICA}}(t)$ (respectivement $\varepsilon_{i,\text{GEVD}}(t)$) comme

$$\begin{aligned}\sigma_{i,\text{ICA}} &= \mathbb{E} \{ \varepsilon_{i,\text{ICA}}^2(t) \} - \mathbb{E} \{ \varepsilon_{i,\text{ICA}}(t) \}^2 \quad , \\ \sigma_{i,\text{GEVD}} &= \mathbb{E} \{ \varepsilon_{i,\text{GEVD}}^2(t) \} - \mathbb{E} \{ \varepsilon_{i,\text{GEVD}}(t) \}^2 \quad .\end{aligned}\tag{IV.22}$$

Le critère de performance permettant alors de comparer les deux méthodes correspond à la moyenne pour l'ensemble des canaux du rapport des variances $\sigma_{i,\text{GEVD}}/\sigma_{i,\text{ICA}}$. Exprimé en décibels, ce critère se note mathématiquement

$$\mathcal{Q} = 10 \log \left(\frac{1}{N_s} \sum_{i=1}^{N_s} \frac{\sigma_{i,\text{GEVD}}}{\sigma_{i,\text{ICA}}} \right).\tag{IV.23}$$

Ainsi, lorsque \mathcal{Q} est positif, alors la variance résiduelle moyenne des signaux débruités par GEVD est plus importante que celle des signaux débruités par ICA : la méthode ICA est donc meilleure. A l'opposé, lorsque \mathcal{Q} est négatif, la procédure GEVD présentée dans ce chapitre s'avère meilleure en moyenne.

3.3 Résultats

Rappelons que, pour ces simulations, le nombre de capteurs est $N_s = 4$. Le rapport signal sur bruit varie par pas de 10 entre -10 et 30 dB. Ces valeurs de rapport signal sur bruit sont réalistes et tiennent compte de l'amplitude maximum des signaux EOG par rapport aux signaux EEG ainsi que de la relative rareté des occurrences des activités oculaires. Nous avons considéré que l'amplitude maximale d'un signal oculaire pouvait atteindre 100 fois l'amplitude d'un signal EEG. En raison de la simplicité de la simulation présentée ici, seule la mise à zéro de la première composante extraite par projection linéaire était nécessaire. De plus, l'extension itérative de la méthode n'a pas été employée car la qualité du signal utilisé pour détecter les périodes d'activité oculaire était parfaite ($x_n(t)$).

Pour la méthode FastICA, on a utilisé une recherche en parallèle des composantes indépendantes (par opposition à une méthode par déflation). Les signaux $\mathbf{x}(t)$ ainsi que $x_n(t)$ étaient concaténés en entrée. Le nombre de composantes indépendantes recherchées étaient de $N_s + 1$ et celle qui ressemblait le plus à $x_n(t)$ était enlevée.

La [figure IV.4](#) présente un résumé des résultats obtenus à partir de 200 simulations pour chacune des valeurs distinctes de bruit. Les données sont représentées sous la forme de « boîtes à moustache » qui illustrent schématiquement les positions respectives des différents quantiles (médiane, quartiles et 5-95 %). Ces deux figures montrent une claire supériorité de la méthode présentée dans ce chapitre par rapport à FastICA pour la plupart des valeurs de bruit. Néanmoins l'avantage se fait plus réduit pour les niveaux de bruit plus faibles. Les deux méthodes tendent donc à être équivalentes lorsqu'il y a peu d'artefacts oculaires ou bien que leur amplitude est faible.

4 Données réelles

Nous illustrons à présent la méthode à partir de données réelles. Les données utilisées sont les mêmes que dans le [chapitre III](#). Nous nous concentrons ici sur la première session

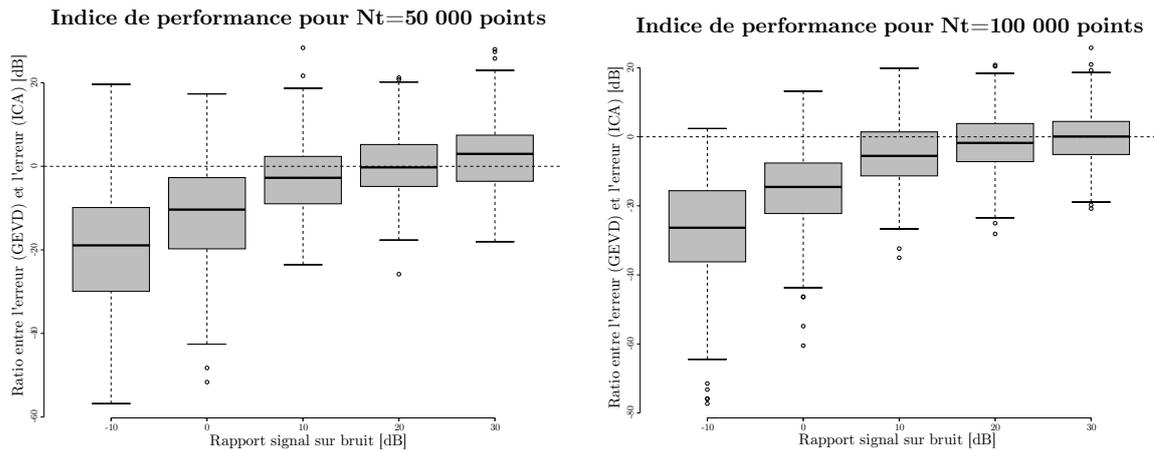


Figure IV.4 Performances de la méthode présentée dans ce chapitre par rapport aux performances d'une méthode basée sur l'analyse en composantes indépendantes (FastICA). Chaque des boîtes à moustache a pour but de représenter de manière schématique la distribution des scores obtenus. Le rectangle de la boîte représente la médiane (trait plein au milieu de la boîte) ainsi que les quartiles à 25 et 75 % (bords de la boîte). Les traits en pointillés s'étendent sur un intervalle dont le but est de représenter un intervalle de confiance à 95 % dans le cas de la comparaison entre deux médianes. Cet intervalle est fondé sur une hypothèse gaussienne de la distribution de la médiane. Les \circ symbolisent les valeurs en dehors de cet intervalle de confiance.

du sujet A01 uniquement. Selon la nomenclature utilisée lors de la « BCI Competition 2008 », cette session correspond aux données intitulées A01T.

4.1 Détection des périodes d'activité oculaire

Les données utilisées dans cette section contiennent des périodes très importantes pour l'analyse des signaux oculaires. Trois types de périodes ont ainsi été successivement imposées au sujet (voir [figure IV.5](#)) :

- yeux ouverts : le sujet devait rester concentré avec les yeux ouverts, cette période contient donc des clignements d'yeux mais peu de mouvements oculaires.
- yeux fermés : le sujet devait fermer les yeux pendant une durée déterminée et essayer de bouger les yeux le moins possible. Cette période est exempte de clignements d'yeux. Cependant de légers mouvements oculaires sont présents pour la plupart des sujets, imposant souvent à un intervenant extérieur de sélectionner la période la moins contaminée. Celle-ci va nous permettre d'utiliser la théorie mise en place au début de ce chapitre concernant l'évaluation de $\Phi_{\mathcal{H}_0}$.
- mouvements oculaires : pendant cette période, le sujet devait réaliser des mouvements oculaires. Cette période est donc très dense en artefacts oculaires de toute sorte.

Commençons donc par utiliser la période pendant laquelle le sujet a les yeux fermés afin de déterminer $\Phi_{\mathcal{H}_0}$. Nous avons fait l'hypothèse d'une distribution gaussienne du signal EOG pendant cette période pour en déduire la distribution théorique de l'énergie $\Phi_{\mathcal{H}_0}$. La [figure IV.6 \(gauche\)](#) montre que l'hypothèse de Gaussianité est respectée. En effet, la distribution des quantiles théoriques suit bien la distribution théorique des quantiles d'une loi Gaussienne. Ceci est mis en évidence par la ligne en rouge. A partir de l'estimation de la variance σ_{ϕ_0} , on a tracé sur la [figure IV.6 \(droite\)](#) la distribution théorique de $\Phi_{\mathcal{H}_0}$ (voir l'[expression IV.4](#)) pour une fenêtre de longueur $w = 20$ (en bleu). Sur cette courbe théorique, on a superposé l'histogramme des valeurs réelles estimées de $\Phi_{\mathcal{H}_0}(t)$ lors de

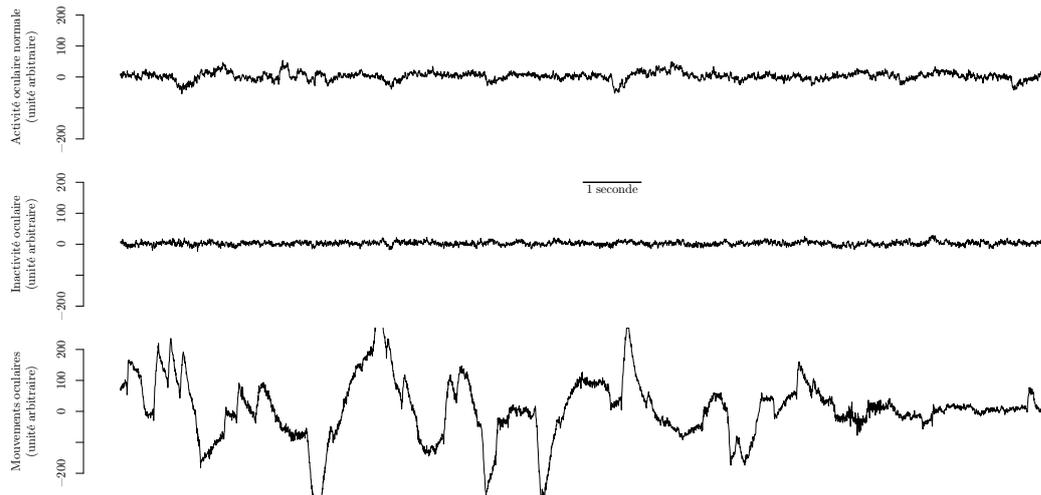


Figure IV.5 Illustration des trois phases différentes au début de la session. En haut, le sujet avait les yeux ouverts, au milieu, il gardait les yeux fermés et en bas, il devait réaliser des mouvements oculaires quelconques.

la période pendant laquelle le sujet avait les yeux fermés. Cette superposition montre que la distribution théorique semble surestimer la distribution réelle. Deux causes principales peuvent expliquer ce phénomène : il se peut que l'hypothèse iid sous-jacente au calcul de la distribution théorique de $\Phi_{\mathcal{H}_0}$ ne soit en fait pas vérifiée ; ou alors le sujet n'a pas complètement respecté la consigne et des micro-mouvements ont pu modifier la distribution. Cependant, la première hypothèse semble plus crédible au vue du graphe quantile-quantile présenté sur la [figure IV.6 \(gauche\)](#). Le calcul du coefficient de Hurst nous confirme la violation de l'hypothèse iid puisque ce coefficient vaut environ 0,68 (un signal iid aurait eu un coefficient de Hurst de 0.5). Ce coefficient montre un effet longue mémoire du processus EOG. De manière intuitive, ceci signifie que le processus a tendance à moins « varier » d'un point à un autre que dans le cas où on serait en présence d'un bruit blanc Gaussien iid. Ainsi il n'est pas surprenant que la densité de probabilité de l'énergie du processus iid soit surévaluée par rapport à la densité réelle. D'un point de vue théorique, il faudrait reprendre les développements pour trouver la loi de $\Phi_{\mathcal{H}_0}(t)$ en appliquant les mêmes équations après décomposition en ondelettes du processus $\varphi(t)$ ([Beran, 1994](#)). Ce biais n'est pas gênant dans notre cas puisqu'on a vu qu'une surestimation du seuil pouvait être compensée par d'autres techniques.

En choisissant un seuil ℓ tel que $\mathbb{P}(\mathcal{H}_1 | \mathcal{H}_0) = 0.001$, l'opération de seuillage permet alors d'identifier les périodes d'activités oculaires lorsque le sujet n'a pas les yeux fermés. Cette opération est illustrée sur la [figure IV.7](#).

4.2 Projection linéaire

La procédure de seuillage précédente nous permet alors les deux matrices de covariances définies par les [équations IV.9](#) et [IV.10](#) et de réaliser la décomposition en valeurs propres généralisée afin d'obtenir les vecteurs de projection. La [figure IV.8](#) représente les trois premières ainsi que les trois dernières composantes extraites. On peut faire plusieurs remarques à propos de cette figure. Tout d'abord, il semble que, dans ce cas particulier, la dimension des signaux EOG soit de moins de 3. Le nombre de composantes qui semblent contaminées est en effet de deux. On peut également observer que les deux premières composantes, bien que reflétant au mieux les artefacts oculaires, semblent également contenir des signaux EEG d'amplitude relativement faible dans le cas de la première composante mais assez importante dans le cas de la deuxième composante. Cette observation va plutôt

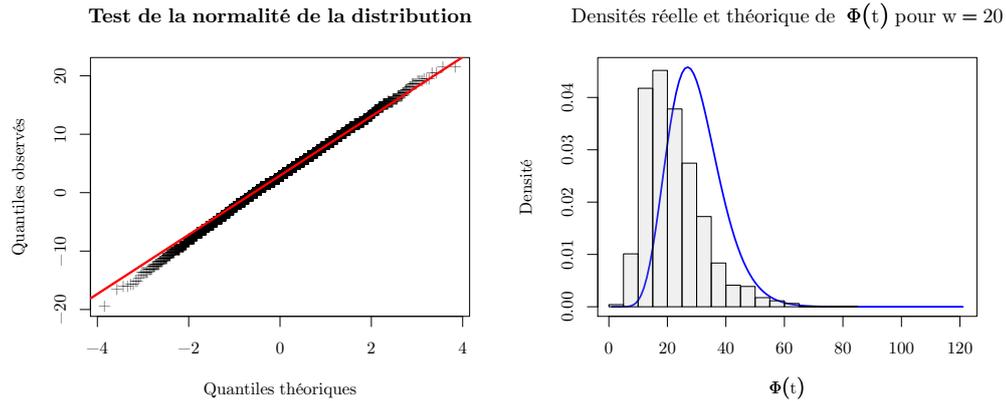


Figure IV.6 La figure de gauche est un graphe quantile-quantile qui permet de vérifier la normalité de la distribution de $\varphi(t)$ lorsque le sujet n'a pas d'activité oculaire importante. Sur la figure de droite, on évalue la validité de la distribution théorique de $\Phi_{\mathcal{H}_0}$. En bleu, on a représenté la distribution théorique et l'histogramme en gris représente la densité observée à partir des données réelles.

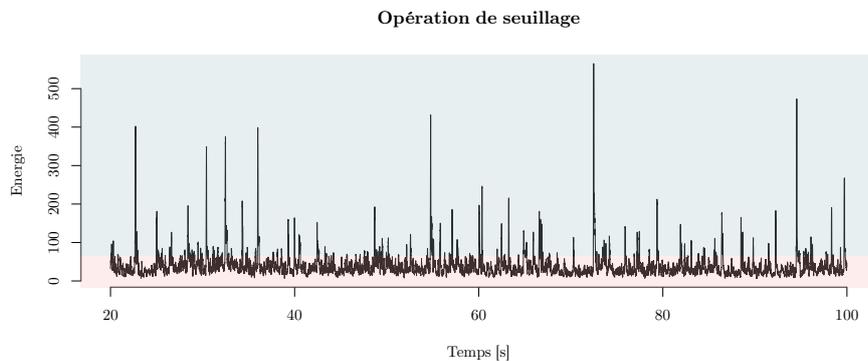


Figure IV.7 Illustration montrant l'opération de seuillage de l'énergie. Sur fond bleu sont représentées les parties du signal qui sont considérées comme de l'activité oculaire et en rouge les parties considérées comme des périodes d'inactivité oculaire.

dans le sens de la nécessité d'utiliser la méthode plus fine de débruitage basée sur les ondelettes plutôt que la mise à zéro des premières composantes. Une mise à zéro signifierait en effet une perte irréversible d'une partie non-négligeable des données EEG. D'une manière plus générale, l'intérêt de la décomposition en valeurs propres généralisée est ici bien mis en évidence. On voit bien que les signaux ressemblant le plus à l'EOG ont été extraits par une simple projection linéaire et classés selon leur qualité. Les derniers signaux présentés sur la [figure IV.8](#) sont donc les composantes qui ressemblent le moins à l'EOG.

4.3 Signaux débruités

Nous avons identifié les composantes qui ressemblent le plus à l'EOG (deux premières composantes extraites). Nous allons maintenant appliquer les deux méthodes différentes de débruitage proposées précédemment et rétro-projeter les signaux débruités dans l'espace des capteurs afin de comparer le résultat avec les signaux initiaux. Ces deux phases sont illustrées sur la [figure IV.9](#). Le débruitage de deux des 22 canaux utilisés lors de l'enregistrement est présenté. Ces deux canaux sont respectivement Fz et POz, ils ont été choisis pour leurs différences de position. Alors que Fz est positionné de manière frontale et est donc très sensible aux artefacts oculaires, POz est plus éloigné, le rendant moins

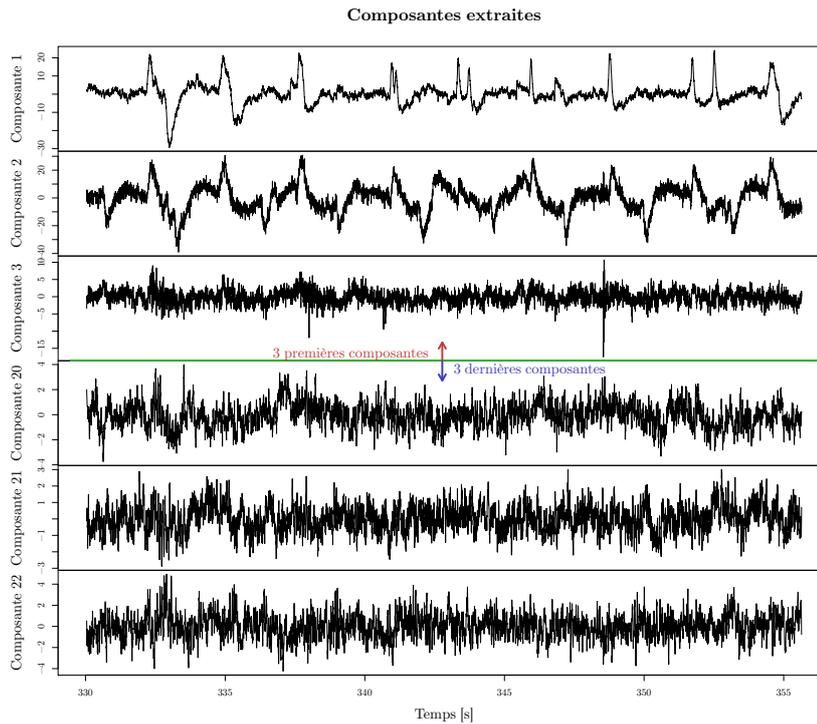


Figure IV.8 Trois premières et trois dernières composantes extraites par décomposition en valeurs propres généralisée. Attention : on a utilisé des échelles différentes pour chaque composante afin de faire mieux apparaître les détails de chacune d'elles.

sujet aux artefacts considérés ici. Les deux figures de gauche ([figure IV.9 \(a,c\)](#)) illustrent le débruitage des signaux recueillis au niveau de Fz et POz en utilisant la méthode de débruitage consistant à remettre à zéro les deux premières composantes. Tandis que les deux figures de droite ([figure IV.9 \(b,d\)](#)) illustrent le cas de l'utilisation d'une méthode de débruitage par ondelettes. Chacune des quatre figures représente, de haut en bas, les signaux originaux, les signaux débruités ainsi que les résidus (soustraction des deux précédents). Notons que la supériorité de la méthode basée sur les ondelettes semblent ici évidente. Aucune trace de signal EEG ne peut être détectée dans les résidus alors que les artefacts oculaires ont bien été identifiés. Cette propriété n'est pas respectée par la méthode consistant à mettre à zéro les deux premières composantes puisqu'on peut observer de l'EEG dans les signaux résiduels. Précisons enfin que la période représentée sur la [figure IV.9](#) est en dehors de la période qui a servi à entraîner les méthodes et trouver les filtres spatiaux.

En conclusion, nous avons illustré la méthode présentée dans ce chapitre pas à pas en détaillant chaque étape. La qualité de la méthode, combinée à un débruitage basé sur les ondelettes semble excellente sur cet exemple. Nous allons à présent évaluer l'intérêt de notre approche pour la classification des signaux dans le cas d'une expérience d'ICM.

5 Application : classification de signaux non-bruités

5.1 Données et méthode

Afin de tester l'intérêt du débruitage des artefacts oculaires dans des conditions réelles, on peut tout simplement observer les variations de performances de classification des signaux issus d'une ICM. Nous avons donc considéré à nouveau les signaux de la « BCI Competition 2008 » et avons utilisé deux méthodes différentes

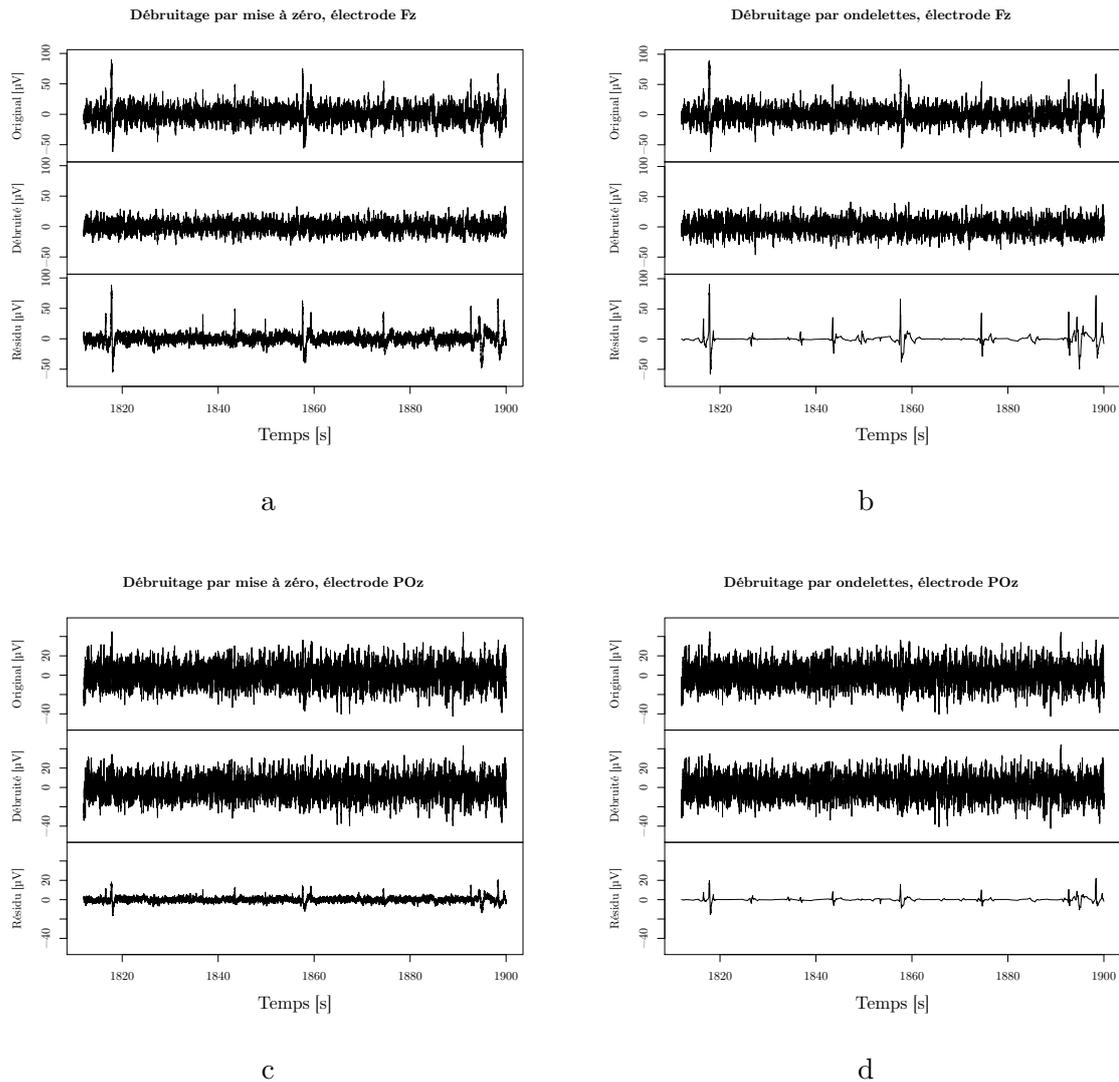


Figure IV.9 Sur chacune des quatre figures sont représentés les signaux originaux (haut), signaux débruités (milieu) ainsi que les résidus (bas). a et c) Débruitage respectivement des électrodes Fz et POz en utilisant la méthode de mise à zéro des deux premières composantes extraites. b et d) Débruitage des électrodes Fz et POz en utilisant une méthode de nettoyage des composantes basée sur la décomposition en ondelettes.

1. classique : on applique la méthode classique présentée dans le [chapitre III](#). La méthode de filtrage spatial utilisée est celle qui a été appelée DCA dans le [chapitre III](#). Cependant, l'étape de préfiltrage des données entre 5 et 35 Hz n'est pas utilisée et on étend la bande de fréquences de la phase d'extraction des caractéristiques afin qu'elle commence à 3 Hz au lieu de 5 Hz. N.B : d'autres paramètres concernant l'utilisation du classifieur ainsi que certaines positions et tailles des fenêtres d'analyse ont été modifiés entre les deux chapitres. Les résultats ne sont donc pas directement comparables. Néanmoins, si les changements de paramètres de la méthode ont entraîné certaines grosses différences pour certains sujets, les performances globales restent à peu près identiques.
2. classique+débruitage : préalablement à l'étape précédente, on applique une étape de débruitage visant à enlever les artefacts oculaires. Cette étape utilise la méthode décrite dans le présent chapitre. La méthode classique est ensuite appliquée sur les signaux débruités de la même manière qu'en 1. Le débruitage par ondelettes a été utilisé pour

l'ensemble des sujets, aucune itération de la méthode n'a été utilisée et le nombre de composantes débruitées était de deux pour l'ensemble des sessions. Les filtres linéaires pour chacune des sessions ont été obtenus à l'aide de la partie initiale des données (aucune tâche n'était réalisée pendant ce temps) et ensuite utilisés sans mise à jour pour l'ensemble des données de la session. Chaque début de session consistait en trois phases oculaires distinctes (normale, yeux fermés et mouvements des yeux).

Rappelons seulement que les données sont issues d'une expérience d'ICM à laquelle 9 sujets ont pris part, réalisant chacun deux sessions (deux jours différents). La tâche proposée au sujet consistait en l'imagination de quatre types de mouvements. Dans cette évaluation, nous considérerons les performances de classification uniquement dans le cas du transfert entre sessions.

Trois canaux EOG étaient disponibles pour ces données, afin de revenir au cas d'un unique canal, nous considérons à chaque fois la moyenne des trois canaux afin de créer une référence EOG artificielle unique.

5.2 Résultats

Une analyse de variance à l'aide d'un modèle similaire à ceux utilisés dans le chapitre précédent (permettant de tenir compte de l'hétérogénéité entre les sujets) montre un effet marginalement significatif de la méthode utilisée ($p = 0.075$). Une analyse détaillée des sujets telle que présentée sur la [figure IV.10](#) montre également une différence relativement faible entre les deux approches proposées ici. Néanmoins, les performances obtenues à l'aide du débruitage sont pour l'ensemble des sessions (excepté 2 sessions sur 16) au moins équivalentes aux performances obtenues sans débruitage. Ceci montre que la méthode de débruitage n'altère pas les données EEG et ne peut être que bénéfique pour augmenter les scores de classification.

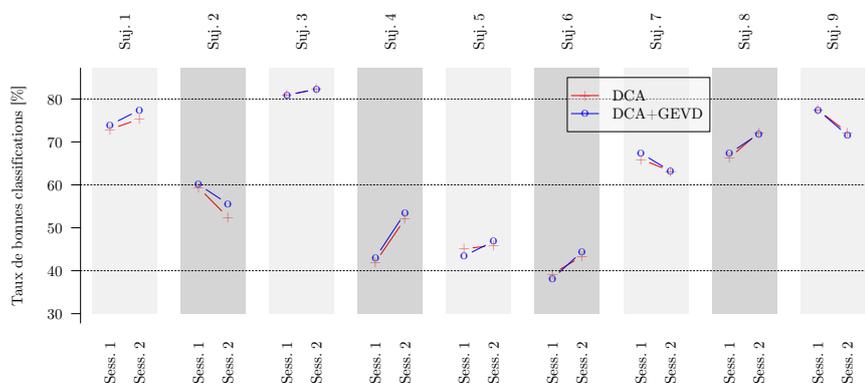


Figure IV.10 Résumé des performances obtenues par les deux méthodes considérées pour chacun des sujets dans les cas 2 → 1 (Sess. 1) et 1 → 2 (Sess. 2). DCA représente la méthode classique et DCA+GEVD la méthode classique couplée à l'étape de débruitage.

Nous espérons que la méthode de débruitage pourrait permettre de faire une meilleure utilisation des signaux à basse fréquence (en dessous de 5 Hz) puisque ce sont les plus contaminés par les artefacts oculaires. La [figure IV.11](#) montre l'utilisation fréquentielle des données par le classifieur pour chacune des tâches réalisées. Celle-ci montre que le débruitage ne modifie pas significativement l'utilisation des fréquences, et, en particulier, que les fréquences faibles ne sont pas plus utilisées dans le cas de l'emploi de la méthode de débruitage couplée à la méthode de classification et d'extraction de caractéristiques classique.

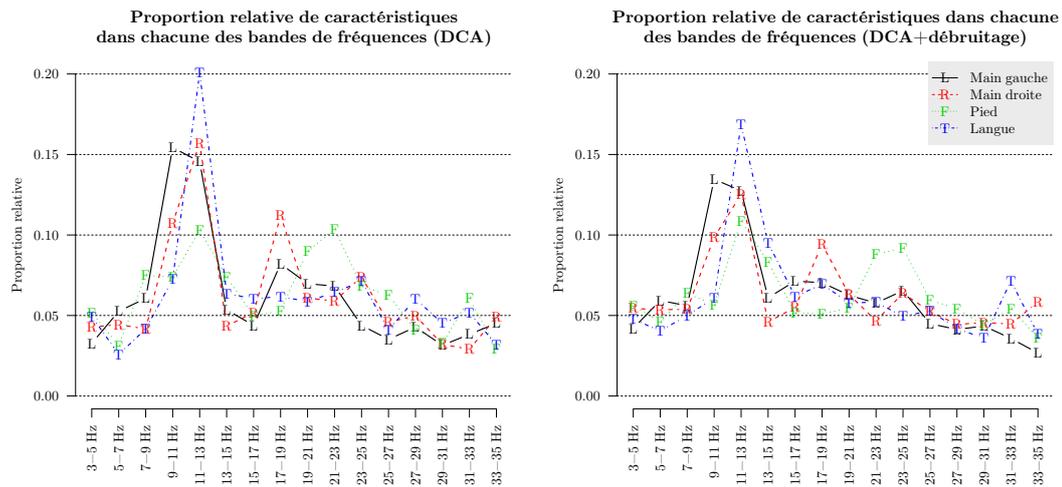


Figure IV.11 Analyse de l'utilisation des bandes de fréquences par le classifieur. Pour chacune des courbes représentant l'utilisation des bandes pour classifier chaque tâche, la somme de tous les points d'une courbe est de un, de telle sorte que nous pouvons observer la proportion relative de l'utilisation de chacune des bandes.

6 Discussion et conclusion

6.1 De l'importance de débruiter les données

Le gain de performance relativement faible de notre méthode présentée dans la section précédente ne doit pas inciter à conclure à un effet négligeable de cette étape. En effet, nous avons présenté la méthode dans un cas simple réalisé en laboratoire et peu d'essais étaient contaminés par des artefacts oculaires (moins de 8% (Naeem et al., 2006 et Schlögl et al., 2007)). L'utilisation d'une telle méthode devient primordiale dans le cas d'un enregistrement en conditions réelles. Les systèmes ICMs ont en effet du mal à se développer dans des contextes plus libres que les laboratoires, la contamination des données par différents artefacts oculaires et mouvements peut en être en partie la cause.

Notons également que l'intérêt de débruiter les données revêt toute son importance dans le cas du suivi et de l'analyse en temps réel des signaux EEG. Un paradigme à contraintes temporelles (tel que celui utilisé dans les données de ce chapitre) permet au sujet de réaliser des clignements d'yeux entre deux tâches (lorsque le système n'analyse pas les données EEG). Mais dans le cas d'une analyse en continu, le sujet doit pouvoir cligner des yeux de manière libre. Ainsi nous pensons que l'utilisation d'une méthode telle que celle présentée dans ce chapitre s'avère primordiale pour l'utilisation des ICMs dans des contextes libres de la vie de tous les jours.

Un important projet dans lequel GIPSA-lab est un des partenaires principaux a été récemment lancés afin d'utiliser de manière conjointe l'oculométrie et l'EEG. L'oculométrie est une technique qui permet de suivre de manière très précise les différents mouvements des yeux. Elle est essentiellement destinée à mesurer les micro-saccades qui sont des mouvements presque invisibles de l'œil. Nous pensons que la technique présentée dans ce chapitre pourra être envisagée pour nettoyer les signaux EOG et trouver les micro-saccades dans l'EEG à partir des informations données par l'oculométrie.

6.2 Implémentation de la méthode

La méthode proposée dans ce chapitre peut être facilement implémentée pour un traitement automatique en temps réel des signaux EEG. En effet, le seul paramètre qui

doit être fixé est le nombre de composantes à débruiter. Ce nombre peut sans difficulté être choisi parmi un, deux ou trois dans la plupart des cas. Nous avons vu que la taille des fenêtres qui permettent de calculer l'énergie des signaux EEG est automatiquement prise en compte dans la distribution théorique IV.4. La méthode peut donc, dans le cas où l'on peut indiquer une période de relative inactivité oculaire du sujet, être utilisée de manière complètement automatique.

Une des limitations la plus importante de la méthode réside dans la quantité de données nécessaires pour entraîner les filtres spatiaux. En effet, la méthode nécessite le calcul de la matrice de covariance des données lors des périodes d'activité oculaires. Cette estimation requiert donc une quantité importante de données afin que l'estimation soit suffisamment robuste. On peut d'ailleurs envisager l'utilisation de méthodes robustes pour évaluer la matrice de covariance des données lors des périodes d'activité oculaires afin de compenser le manque de données.

La complexité des calculs nécessaires à l'application de la méthode est très faible. La décomposition en valeurs propres généralisée a une complexité de l'ordre d'une décomposition en valeurs propres et n'est réalisée qu'une seule et unique fois au début de l'enregistrement. Lors du débruitage en temps réel, la décomposition en ondelettes se fait de manière très efficace à l'aide d'étages de filtres successifs (Mallat, 1999). La complexité algorithmique globale de la méthode est donc très faible. Celle-ci peut sans difficulté être utilisée pour un nombre de capteurs beaucoup plus important.

6.3 Conclusion

Nous avons présenté une méthode automatique et simple permettant d'enlever les artefacts oculaires des enregistrements EEG. Cette méthode présente un cas particulier d'un cadre plus général visant à décomposer les signaux multivariés en sous-espaces indépendants (Tsalaile et al., 2009, Sameni et al., 2008b, Sameni, 2008 et Sameni et al., 2009). Ce cadre flexible permet de choisir le critère permettant de trouver les combinaisons linéaires, la fonction permettant de débruiter et l'utilisation de la partie itérative pour affiner les résultats. Dans le cas particulier présenté ici, nous avons utilisé la non-stationnarité pour trouver la combinaison linéaire. Ce critère nous a conduit à un quotient de Rayleigh (Strang, 1988, Golub et van Loan, 1996 et Gray, 2006) généralisé dont la solution est donnée par la méthode de la décomposition en valeurs propres généralisée. Dans d'autres applications, d'autres contrastes pourraient également être utilisés, par exemple la périodicité (extraction de signaux cardiaques) ou encore un contraste basé sur les diversités spectrales des signaux.

Nous avons présenté deux différentes méthodes de débruitage. Alors que la première engendre une diminution du rang des données du nombre de composantes mises à zéro, la deuxième basée sur la transformée en ondelettes discrète permet de conserver le rang des données.

Les différents exemples d'utilisation qui ont été présentés dans ce chapitre ont montré l'intérêt de la méthode pour le débruitage. Tout d'abord, dans le cas de données simulées, nous avons montré que notre méthode s'avère plus précise qu'une méthode classique basée sur l'analyse en composantes indépendantes (ACI). L'exemple suivant basé sur des données réelles a permis d'illustrer de manière détaillée chaque étape de la méthode et a montré la supériorité du débruitage par ondelettes par rapport à une simple mise à zéro des premières composantes extraites. Enfin, dans le but de classifier des signaux EEG d'une expérience d'ICM, la dernière application a montré que notre méthode pouvait dans le cas général, engendrer de faibles gains de performances mais ne présentait aucun inconvénient à être utilisée de manière systématique.

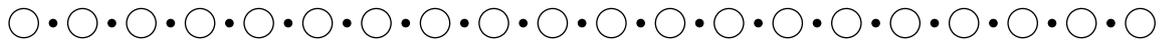
7 Références du chapitre

- Anderson, C. W., Stolz, E. A. et Shamsunder, S. (1998). Multivariate autoregressive models for classification of spontaneous electroencephalographic signals during mental tasks. *IEEE Trans. Biomed. Eng.*, 45(3):277–286.
- Beran, J. (1994). *Statistics for long-memory processes*. Chapman & Hall, New-York, USA.
- Boudet, S. (2008). *Filtrage d'artefacts par analyse multicomposante de l'électroencéphalogramme de patients épileptiques*. Thèse de doctorat, Université de Lille 1.
- Delorme, A., Sejnowski, T. et Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *Neuroimage*, 34(4):1443–1449.
- Donoho, D. L. et Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Am. Statist. Ass.*, 90(432):1200–1224.
- Golub, G. et van Loan, C. (1996). *Matrix Computations*. The Johns Hopkins University Press, Troisième édition.
- Gray, R. M. (2006). Toeplitz and circulant matrices: A review. *Foundations and Trends in Communications and Information Theory*, 2(3):155–239.
- Hyvärinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neural Netw.*, 10(3):626–634.
- Hyvärinen, A., Karhunen, J. et Oja, E. (2001). *Independent Component Analysis*. Wiley-Interscience.
- Ille, N. (2001). *Artifact correction in continuous recordings of the electro- and magnetoencephalogram by spatial filtering*. Thèse de doctorat, Universität Mannheim.
- Ille, N., Berg, P. et Scherg, M. (2002). Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies. *J. Clin. Neurophysiol.*, 19(2):113–124.
- Jung, T., Humphries, C., Lee, T., McKeown, M. et Iragui, V. et al. (2000a). Removing electroencephalographic artifacts by blind source separation. *J. Psychophysiol.*, 37:163–178.
- Lindgren, G., Leadbetter, M. et Rootzén, H. (1983). *Extremes and related properties of stationary sequences and processes*. Springer, New-York, USA.
- Mallat, S. (1999). *A Wavelet Tour of Signal Processing, Second Edition (Wavelet Analysis & Its Applications)*. Academic Press.
- Naeem, M., Brunner, C., Leeb, R., Graimann, B. et Pfurtscheller, G. (2006). Separability of four-class motor imagery data using independent components analysis. *J. Neural Eng.*, 3(3):208–216.
- Romero, S., Mañanas, M. A. et Barbanoj, M. J. (2008). A comparative study of automatic techniques for ocular artifact reduction in spontaneous EEG signals based on clinical target variables: A simulation case. *Comput. Biol. Med.*, 38(3):348–360.
- Romero, S., Mañanas, M. et Barbanoj, M. (2009). Ocular reduction in EEG signals based on adaptive filtering, regression and blind source separation. *Ann. Biomed. Eng.*, 37(1):176–191.
- Sameni, R. (2008). *Extraction of Fetal Cardiac Signals from an Array of Abdominal Maternal Recordings*. Thèse de doctorat, Institut Polytechnique de Grenoble.
- Sameni, R., Jutten, C. et Shamsollahi, M. B. (2009). A deflation procedure for subspace decomposition. *IEEE Trans. Signal Process.*. Submitted (in review process).
- Sameni, R., Jutten, C. et Shamsollahi, M. B. (2008a). Multichannel electrocardiogram decomposition using periodic component analysis. *IEEE Trans. Biomed. Eng.*, 55(8):1935–1940.
- Sameni, R., Shamsollahi, M. B. et Jutten, C. (2008b). Model-based bayesian filtering of cardiac contaminants from biomedical recordings. *Physiol. Meas.*, 29(5):595–613.

- Schlögl, A., Keinrath, C., Zimmermann, D., Scherer, R. et Leeb, R. et al. (2007). A fully automated correction method of EOG artifacts in EEG recordings. *Clin. Neurophysiol.*, 118(1):98–104.
- Strang, G. (1988). *Linear Algebra and Its Applications*. Brooks/Cole, 3rd édition.
- Thulasidas, M., Guan, C., Ranganatha, S., Wu, J. et Zhu, X. et al. (2004). Effect of ocular artifact removal in brain computer interface accuracy. Dans *Conf. Proc. IEEE. Eng. Med. Biol. Soc.*, pages 4385–4388.
- Tsalaile, T., Sameni, R., Sanei, S., Jutten, C. et Chambers, J. (2009). Sequential blind source extraction for quasi-periodic signals with time-varying period.. *IEEE Trans. Biomed. Eng.*, 56(3):646–655.
- Vigário, R., Särelä, J., Jousmäki, V., Hämmäläinen, M. et Oja, E. (2000). Independent component approach to the analysis of EEG and MEG recordings. *IEEE Trans. Biomed. Eng.*, 47(5):589–593.

Chapitre V

FILTRAGE DES ÉTATS MENTAUX PAR MODÈLE DE MARKOV



1	Introduction	106
2	Formulation théorique	107
2.1	Modèle d'évolution des états mentaux, cadre variationnel bayésien	107
2.2	Résumé de l'algorithme de filtrage des probabilités	114
2.3	Illustration du filtrage variationnel bayésien	115
3	Méthode générale pour les ICMs	117
3.1	Phase d'entraînement	117
3.2	Phase d'évaluation	117
3.3	Résumé	117
4	Apprentissage de l'état « repos »	118
4.1	Utilisation des périodes de repos	118
4.2	Définition implicite	119
5	Données expérimentales	120
5.1	« BCI Competition III, dataset 4b »	120
5.2	« BCI Competition IV, dataset 1 »	121
6	Résultats	121
6.1	Méthode d'évaluation	121
6.2	« BCI Competition III »	123
6.3	« BCI Competition IV »	127
7	Discussion et conclusion	129
7.1	Utilisation des bandes de fréquences	130
7.2	Améliorations dues au filtrage variationnel	130
8	Références du chapitre	132

NOUS AVONS CONSACRÉ les chapitres précédents à plusieurs étapes fondamentales de traitement pour la conception d'Interfaces Cerveau-Machines (ICMs). Nous nous sommes notamment intéressés au débruitage des signaux ([chapitre IV](#)) ainsi qu'aux méthodes de filtrage spatial permettant d'extraire les sources cérébrales associées à la réalisation des tâches mentales ([chapitre III](#)). De plus, l'extraction de caractéristiques permettant de résumer l'activité cérébrale à quelques valeurs principales ainsi que la classification de ces valeurs a été décrite et abordée dans les différents chapitres et approfondie dans [l'annexe A](#).

Cependant la plupart des méthodes mises en place jusqu'à présent se sont concentrées sur des paradigmes ICMs relativement contraints dans la mesure où les sujets devaient envoyer leur commande mentale à des instants imposés par le système (ICMs dites synchrones). Dans le cas de paradigmes sans contrainte temporelle, l'analyse de l'activité

cérébrale doit être réalisée en continu (ICMs dites asynchrones). Les problèmes engendrés par ce type de système sont multiples. Dans ce chapitre, nous montrons que les méthodes qui ont été décrites depuis le début de ce manuscrit peuvent être appliquées au cas de paradigmes non contraints.

Nous adaptons ces méthodes à la problématique des interfaces non-contraintes en ajoutant une étape de traitement inédite visant à contrôler la dynamique en sortie du système selon le principe neurophysiologiquement plausible de relative stabilité des états mentaux du sujet. Cette étape a pour but de régulariser la sortie du système afin de compenser les variabilités engendrées par l'analyse en continu de l'activité cérébrale. Cette problématique est abordée en utilisant un modèle de Markov caché dans un cadre variationnel bayésien.

Ces travaux ont été réalisés en collaboration avec le Dr. Jérémie Mattout (INSERM, Lyon).

1 Introduction

Depuis plus de 20 ans, experts en traitement du signal ainsi que neurophysiologistes travaillent côte à côte pour fournir aux personnes souffrant de graves maladies motrices des outils pour restaurer la communication ou des capacités motrices minimales ([Birbaumer et Cohen, 2007](#)). Si les paradigmes relativement contraints, tels que ceux utilisés pour les données des chapitres précédents, ont permis d'obtenir des taux de classification élevés, le caractère peu flexible de ce type de systèmes a jusqu'à présent limité leur utilisation par une large population de patients. Or la principale limitation des ICMs classiques réside dans la contrainte imposée au sujet d'envoyer sa commande au moment où le système l'impose et non au moment où le sujet le désire.

Nous avons vu au cours du [chapitre II](#) qu'un effort significatif visait aujourd'hui à rendre les ICMs plus flexibles. Ceci passe tout d'abord par la suppression de la contrainte temporelle imposant au sujet l'instant de sa commande ([Townsend et al., 2004](#)). Bien que cette problématique soit apparue il y a une dizaine d'années ([Mason et Birch, 2000](#)), seules quelques équipes dans le monde travaillent sur ce problème et tentent de résoudre les difficultés qu'il soulève ([Mason et al., 2006](#), [Bashashati et al., 2007b](#), [Birch et al., 2003](#), [Scherer et al., 2008](#) et [Galán et al., 2008](#)). Parmi ces difficultés, l'une des principales est engendrée par la nécessité, pour le système, de décider si le sujet est en train d'envoyer une commande ou s'il se trouve dans un état naïf (du point de vue du système). Or, s'il est relativement aisé de discriminer entre deux tâches déterminées dont on connaît bien les caractéristiques, il est beaucoup plus difficile de discriminer entre une tâche et un état neutre, dont la variabilité est très importante. Afin de pallier ce problème, [Townsend et al. \(2004\)](#) ont proposé d'utiliser ce qu'on appelle un « dwell time ». L'idée de ce système est d'analyser l'activité cérébrale en continu mais de ne prendre une décision que lorsqu'on est sûr qu'une certaine tâche est maintenue pendant une période relativement longue. Le taux de fausses détections est alors contrôlé par le temps de « dwell time ».

Cependant, l'inconvénient de cette approche est qu'elle ralentit le flux d'information du système. Elle tente néanmoins d'utiliser l'information *a priori* qui consiste à considérer que la variation d'état mental du sujet doit être relativement lente et peut permettre de pallier la grande variabilité qui serait engendrée par une classification en continu. Dans ce chapitre, nous proposons donc d'étendre cette idée en proposant une méthode basée sur un modèle de Markov caché. L'idée sous-jacente est d'utiliser à la fois la relative stabilité *a priori* des états mentaux du sujet et également les transitions *a priori* entre les différents états mentaux pour filtrer la sortie du classifieur afin de rendre le système stable.

Différents travaux ont déjà tenté d'utiliser des chaînes de Markov cachées dans le cadre des ICMs mais dans un but de classification ([Obermaier et al., 2001](#)) en utilisant la vraisemblance par rapport à plusieurs modèles correspondant à chacune des classes. Cette

approche générative (un modèle par classe) souffre néanmoins d'une perte de puissance du modèle qui est utilisé pour générer les données plutôt que pour discriminer les différentes classes. Cet inconvénient a été pointé par [Chiappa \(2006\)](#) qui a proposé d'utiliser un modèle de Markov discriminatif plutôt que génératif ([Chiappa et Bengio, 2004](#)). Dans le même but, certains auteurs ont également proposé d'utiliser une description continue des états plutôt que discrète, proposant donc des filtres de Kalman pour la classification des signaux EEG ([Tsui et al., 2009](#), [Sykacek et al., 2004](#) et [Sykacek et al., 2003](#)). Bien que l'outil présenté dans ce chapitre soit également basé sur des chaînes de Markov cachées, le but est ici de proposer une méthode de filtrage de la sortie du classifieur et non une méthode de classification.

2 Formulation théorique

Afin de simplifier les notations dans le présent chapitre, nous adoptons une notation différente de celle utilisée dans les chapitres précédents. Ainsi, nous noterons $x_t \in \mathbb{R}^{N_s}$ les données EEG, enregistrées à l'aide de N_s capteurs. A chaque instant t , l'état mental du sujet sera appelé l_t . Nous nous placerons dans le cas d'une interface à commandes discrètes, c'est-à-dire l_t prend des valeurs parmi un alphabet fini dont on peut définir une bijection avec l'ensemble $\{1, 2, \dots, c\}$. Enfin, nous supposerons disposer d'un système capable, à partir d'une fenêtre de données x_t autour de chaque instant t , de fournir une estimation de la probabilité que le sujet soit dans chacun des états mentaux. La probabilité pour que le sujet soit dans l'état $i \in \{1, 2, \dots, c\}$ à l'instant t sera notée $p_{i,t}$. L'ensemble des probabilités estimées à l'instant t sera rassemblé sous la forme d'un vecteur

$$p_t = \begin{pmatrix} p_{1,t} \\ p_{2,t} \\ \vdots \\ p_{c,t} \end{pmatrix} \quad \text{avec} \quad \forall t \in \mathbb{R} \quad \sum_{i=1}^c p_{i,t} = 1. \quad (\text{V.1})$$

Bien que l_t soit une valeur discrète dans l'intervalle $\{1, 2, \dots, c\}$, nous la noterons dans la suite du chapitre de manière préférentielle sous la forme d'une variable indicatrice $l_t \in \{0, 1\}^c$ telle que

$$\forall i \in \{1, 2, \dots, c\} \quad l_{i,t} = \begin{cases} 1 & \text{si } l_t = i \\ 0 & \text{sinon} \end{cases}. \quad (\text{V.2})$$

2.1 Modèle d'évolution des états mentaux, cadre variationnel bayésien

Le but de ce chapitre est de proposer une méthode afin d'inférer l_t à partir des observations p_t . Pour ce faire, nous allons considérer un modèle incluant une dépendance d'ordre un entre la suite des variables latentes l_t . La loi décrivant cette dépendance sera notée $f(l_t | l_{t-1})$. De plus, les variables d'observation sont considérées comme indépendantes conditionnellement aux variables latentes l_t . Ce modèle est donc un cas classique de modèle de Markov d'ordre un ([Rabiner, 1989](#)) et peut être représenté sous forme graphique par le schéma de la [figure V.1](#).

Un tel modèle nécessite donc de définir les éléments suivants :

- loi d'évolution du système, $f(l_t | l_{t-1})$: cette loi décrit l'évolution interne du système. Dans notre cas, comme l_t est une variable discrète, la définition de la loi d'évolution revient donc à définir une matrice de transition $T_t \in \mathbb{R}^{c \times c}$ dont les éléments $[t]_{i,j,t} = \mathbb{P}(l_t = i | l_{t-1} = j)$.
- loi d'observation, $f(p_t | l_t)$: cette loi décrit la loi de probabilité des observations sachant que l'état latent du système est connu.

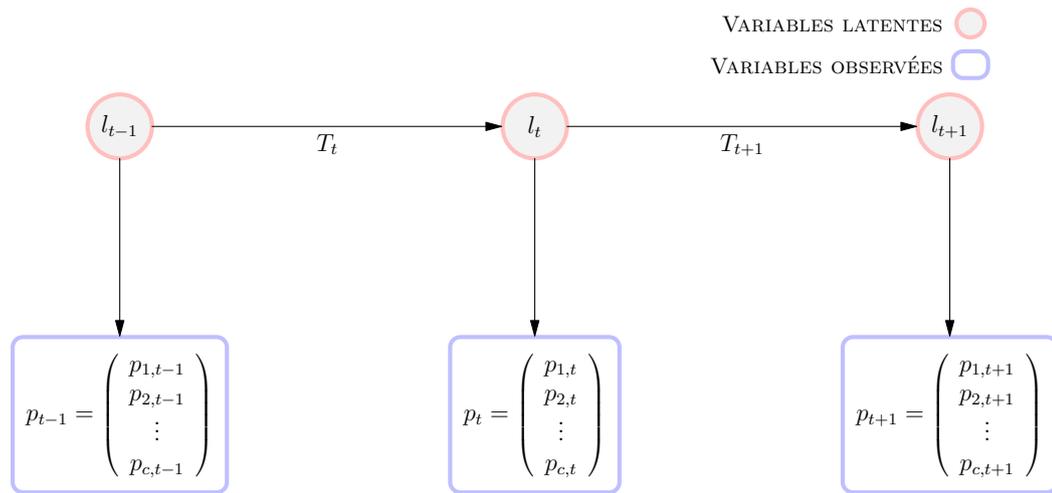


Figure V.1 Modèle graphique décrivant les interactions entre les différentes variables du modèle. l_t représente une variable latente du système et peut être interprétée comme l'état mental du sujet. Tandis que p_t décrit l'estimation des probabilités pour que le sujet soit dans chacun des états mentaux proposés.

- initialisation du système : on doit enfin déterminer l'état initial du système, c'est-à-dire l_0 .

Les chaînes de Markov soulèvent trois types de problèmes (Rabiner, 1989). Tout d'abord, étant donnée une séquence d'observations et un modèle, comment calculer la probabilité pour que ce modèle puisse être responsable de cette séquence ? Ce premier problème permet de faire de la sélection de modèle. C'est également dans ce sens que sont utilisées les chaînes de Markov pour la classification. Deuxièmement, étant donnée une séquence d'observations, comment inférer la séquence d'états du système (variables latentes) qui l'a générée ? Nous nous intéressons ici à ce deuxième type de problème, c'est-à-dire que nous désirons inférer l'état mental du sujet sachant la séquence de probabilités observées. Un troisième problème revient à se demander comment ajuster les paramètres du modèle de façon que la probabilité d'une séquence d'observation soit maximale. Ce dernier point ne sera pas abordé ici.

Notons enfin que le nom de chaîne de Markov cachée fait référence au caractère inobservable de la variable latente dont l'évolution est régie par une chaîne de Markov d'ordre un (caractérisée par la matrice de transition T_t). Dans la suite de ce chapitre, nous nous plaçons dans un cadre variationnel bayésien. Si l'inférence variationnelle est utilisée depuis plusieurs années dans des contextes « hors-ligne » (Ghahramani et Beal, 2001 et MacKay, 1997), la formulation du problème lié au modèle précédent dans un contexte de traitement des données en temps réel est relativement récente (Šmídl et Quinn, 2005 et Šmídl et Quinn, 2008). En nous basant sur le cadre théorique général de traitement des données en temps réel développé par Šmídl et Quinn (2008) et en exploitant deux des exemples de ces auteurs, nous détaillons le traitement variationnel associé au modèle précédent V.1 afin d'inférer l'état mental du sujet à partir de probabilités observées. Les équations qui suivent sont presque identiques à celles trouvées dans chacune des références Šmídl et Quinn (2005) et Šmídl et Quinn (2008). Cependant dans un cas (Šmídl et Quinn, 2005), l'absence d'un terme de modulation de l'apprentissage de la matrice de transition rendait le système trop instable, alors que dans Šmídl et Quinn (2008), le modèle d'observation utilisé est plus général que celui nécessaire dans notre cas.

2.1.1 Le cadre variationnel bayésien

Dans le contexte bayésien, on appelle modèle un ensemble de croyances (connaissances *a priori*) dont chacun des éléments contient une information sur la manière avec laquelle sont générées les données. Le modèle \mathcal{M} est décrit par un ensemble de paramètres Θ . Ce modèle gouverne la génération des données, qui peuvent ne pas être directement observables, on les appelle alors « cachées ». Lorsque l'on désire utiliser un certain modèle afin d'analyser des données observées, les différents paramètres du modèle doivent être ajustés et les valeurs des variables cachées inférées afin d'expliquer au mieux les données observées. Ces deux problèmes sont équivalents aux deux derniers problèmes formulés par [Rabiner \(1989\)](#) et rappelés dans le paragraphe précédent. Or, si l'on modélise les croyances par des densités de probabilité, alors le théorème de Bayes ([Bayes, 1763](#)) décrit la manière dont les croyances et les variables cachées doivent être mises à jour en fonction des données observées. Celui-ci s'exprime par

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)}. \quad (\text{V.3})$$

Littéralement, cette formule met en évidence le fait que notre connaissance à propos de y après avoir vu x (loi *a posteriori* $p(y|x)$) s'exprime en fonction de l'ancienne connaissance à propos de y (loi *a priori* $p(y)$) modifiée par la probabilité qu'on observe les données x selon le modèle précédent (vraisemblance $p(x|y)$). Le dénominateur permet de s'assurer que la loi *a posteriori* se comporte comme une probabilité, il est aussi appelé « vraisemblance marginale ».

Ainsi, dans un contexte bayésien, l'inférence consiste à calculer la loi *a posteriori* selon les variables cachées, alors que l'apprentissage consiste à calculer la loi *a posteriori* selon l'ensemble des paramètres du modèle. Cependant, le cadre bayésien s'avère vite très lourd en calculs lorsque le nombre de variables devient important et les dépendances complexes. C'est pourquoi on préfère en général utiliser des approximations plus ou moins brutales. Parmi les plus brutales, on peut citer le principe du maximum de vraisemblance ou du maximum *a posteriori*, qui sont toutes deux des approximations ponctuelles des paramètres (comparativement aux méthodes basées sur les lois de probabilité des paramètres). L'algorithme connu sous le nom d'« expectation-maximisation » (EM) fournit une manière efficace d'optimiser les vraisemblances dans ces deux cas. L'inconvénient de ces méthodes réside dans l'impossibilité de définir des *a priori* sur les paramètres du modèle, ce qui peut conduire à des valeurs aberrantes. Parmi les autres approximations, on peut citer les méthodes de Monte-Carlo qui s'avèrent très lourdes à mettre en œuvre.

Plusieurs de ces méthodes pourraient permettre de réaliser l'inférence des variables latentes dans le cas du modèle présenté dans le paragraphe précédent. Nous avons choisi d'adopter ici le cadre variationnel bayésien pour sa simplicité et les possibilités d'extensions qu'il offre dans le cas où certains paramètres du modèle doivent être inférés en fonction des données. Très populaire en physique statistique, l'approximation variationnelle repose sur l'idée centrale d'approcher les lois *a posteriori* selon les différents poids résumant les paramètres du modèle et des variables cachées par des formes qui rendent les calculs réalisables. L'optimisation de ces lois approchées repose alors sur la minimisation entre les lois approchées et les lois exactes ([Choudrey, 2002](#) et [Šmídl et Quinn, 2005](#)). Le problème se ramène donc à choisir un espace de distributions dans lequel l'ensemble des éléments rend les calculs réalisables, ainsi qu'une mesure de similarité permettant de quantifier la différence entre la vraie distribution et la distribution approchée. L'approximation variationnelle conduit à utiliser la divergence de Kullback-Leibler comme mesure de similarité. Quant à l'espace des distributions, celui-ci est défini comme l'espace dans lequel les densités *a posteriori* sont conditionnellement indépendantes. Mathématiquement, on écrit ([Šmídl et Quinn, 2005](#))

$$\mathbb{F}_c = \{f(\theta_1, \theta_2 | D) \mid f(\theta_1, \theta_2 | D) = f(\theta_1 | D)f(\theta_2 | D)\}, \quad (\text{V.4})$$

où D est l'ensemble des données et θ_1, θ_2 sont deux ensembles disjoints de paramètres. Notons que cette condition implique d'ores et déjà que le nombre de paramètres soit strictement plus grand que un. L'approximation variationnelle repose alors sur le théorème suivant dont la preuve est fournie dans Šmídl et Quinn (2005) :

Théorème 2 (VARIATIONNEL BAYÉSIEN). *Soit $f(\theta | D)$ la loi a posteriori du paramètre multivarié θ . On suppose que θ peut être partitionné en q sous-vecteurs tel que :*

$$\theta = [\theta_1^T, \theta_2^T, \dots, \theta_q^T]^T. \quad (\text{V.5})$$

Soit $\check{f}(\theta | D)$ une distribution de probabilités approchée choisie parmi l'ensemble des distributions telles que

$$\check{f}(\theta | D) = \check{f}(\theta_1, \theta_2, \dots, \theta_q | D) = \prod_{i=1}^q \check{f}(\theta_i | D). \quad (\text{V.6})$$

Alors le minimum de la divergence de Kullback-Leibler définie par

$$\tilde{f}(\theta | D) = \arg \min_{\check{f}(\cdot)} \text{KL} \left(\check{f}(\theta | D) \parallel f(\theta | D) \right), \quad (\text{V.7})$$

est atteint en

$$\forall i \in [1 \dots q] \quad \tilde{f}(\theta_i | D) \propto \exp \left(\mathbb{E}_{\check{f}(\theta_{[1 \dots q] \setminus i} | D)} \{ \ln (f(\theta, D)) \} \right). \quad (\text{V.8})$$

$\tilde{f}(\theta | D)$ est appelée l'approximation variationnelle tandis que $\tilde{f}(\theta_i | D)$ représente l'ensemble des lois marginales variationnelles.

Ce théorème est à la base de l'approximation variationnelle. Nous allons à présent utiliser ce théorème dans le cadre du modèle de la figure V.1. La méthodologie utilisée suit les différentes étapes explicitées dans Šmídl et Quinn (2005) et Šmídl et Quinn (2008).

2.1.2 Reformulation variationnelle du problème

2.1.2.1 Modèle d'évolution du système

Étant donné le modèle présenté sur la figure V.1 il est naturel de considérer la loi d'évolution

$$f(l_t | l_{t-1}, T_t) = \mathcal{M}u_{l_t}(T_t l_{t-1}) \quad (\text{V.9})$$

où $\mathcal{M}u(\cdot)$ est la loi multinomiale. La loi multinomiale est la généralisation multidimensionnelle de la loi binomiale. Rappelons que la distribution multinomiale d'un vecteur c -dimensionnel l_t , où $l_{i,t} \in \mathbb{N}$ et $\sum_{i=1}^c l_{i,t} = \gamma$ s'écrit d'une manière générale

$$f(l_t | \alpha) = \mathcal{M}u_{l_t}(\alpha) = \frac{1}{\zeta_{l_t}(\alpha)} \prod_{i=1}^c \alpha_i^{l_{i,t}} \chi_{\mathbb{N}^c}(l_t), \quad (\text{V.10})$$

où

$$\zeta_{l_t}(\alpha) = \frac{\prod_{i=1}^c l_{i,t}!}{\gamma!},$$

$\chi_{\mathbb{N}^c}$ est la fonction indicatrice de \mathbb{N}^c et $\alpha = [\alpha_1, \dots, \alpha_c]^T$ est un ensemble de nombres positifs de somme unitaire.

Dans notre cas, comme l_t a été définie comme une variable indicatrice et prend uniquement les valeurs 0 ou 1, on a $\zeta_{l_t}(\alpha) = 1$ quel que soit l'état mental du sujet.

2.1.2.2 Modèle d'observation

On modélise également les probabilités sachant les variables latentes par une loi de Dirichlet. On note

$$f(p_t | l_t) = \mathcal{D}i_{p_t}(1_{c,1} + \rho l_t), \quad (\text{V.11})$$

où $1_{c,1}$ est un vecteur colonne contenant uniquement des 1. Le paramètre ρ permet d'influencer la manière avec laquelle on peut inférer la valeur de la variable latente à partir des observations qui sont les probabilités. Ainsi, plus ρ est grand, plus les valeurs observées des probabilités influenceront les variables latentes.

Rappelons que la distribution de Dirichlet d'une variable c -dimensionnelle α de \mathbb{R}^c , notée $[\alpha_1, \dots, \alpha_c]^T$, s'écrit de manière générale

$$f(\alpha | \beta) = \mathcal{D}i_{\alpha}(\beta) = \frac{1}{\zeta_{\alpha}(\beta)} \prod_{i=1}^c \alpha_i^{\beta_i-1} \chi_{\Delta_c}(\alpha), \quad (\text{V.12})$$

où Δ_c est appelé le simplexe des probabilités et est défini comme

$$\Delta_c = \left\{ \alpha \mid \alpha_i \geq 0, \sum_{i=1}^c \alpha_i = 1 \right\},$$

$\chi_{\Delta_c}(\alpha)$ vaut 1 lorsque α appartient au simplexe des probabilités et 0 sinon. Cette écriture permet de sélectionner uniquement les valeurs de α qui forme des probabilités. Le vecteur de paramètres $\beta = [\beta_1, \dots, \beta_c]^T$, $\forall i \beta_i > 0$, $\sum_{i=1}^c \beta_i = \gamma$ permet de définir la constante de normalisation

$$\zeta_{\alpha}(\beta) = \frac{\prod_{i=1}^c \Gamma(\beta_i)}{\Gamma(\gamma)},$$

où $\Gamma(\cdot)$ est la fonction Gamma.

2.1.2.3 Modèle d'évolution de la matrice de transition

Enfin, le modèle d'évolution de la matrice de transition T_t est défini comme

$$\forall i \in \{1, 2, \dots, c\} \quad f(T_{i,t} | T_{i,t-1}) = \mathcal{D}i_{T_{i,t}}(\kappa T_{i,t-1}), \quad (\text{V.13})$$

où $T_{i,t}$ représente l'ensemble des vecteurs colonnes formés par la matrice T_t . Ainsi, on a

$$f(T_t | T_{t-1}) = \prod_{i=1}^c f(T_{i,t} | T_{i,t-1}).$$

Le premier moment d'une distribution de Dirichlet est donné par

$$\hat{\alpha}_i = \mathbb{E}_{f(\alpha|\beta)} \{\alpha_i\} = \frac{\beta_i}{\gamma},$$

où rappelons que $\sum_{i=1}^c \beta_i = \gamma$. Par conséquent la [définition V.13](#) implique que $\mathbb{E}\{T_{i,t}\} = \hat{T}_{i,t} = T_{i,t-1}$. Le paramètre κ est primordial dans le modèle puisqu'il conditionne la rapidité des variations de la matrice des transitions. Ainsi, pour une valeur de κ élevée, la matrice de transition T_t aura des variations lentes alors que des valeurs de κ faibles autoriseront des variations rapides. Ce comportement est illustré sur la [figure V.2](#).

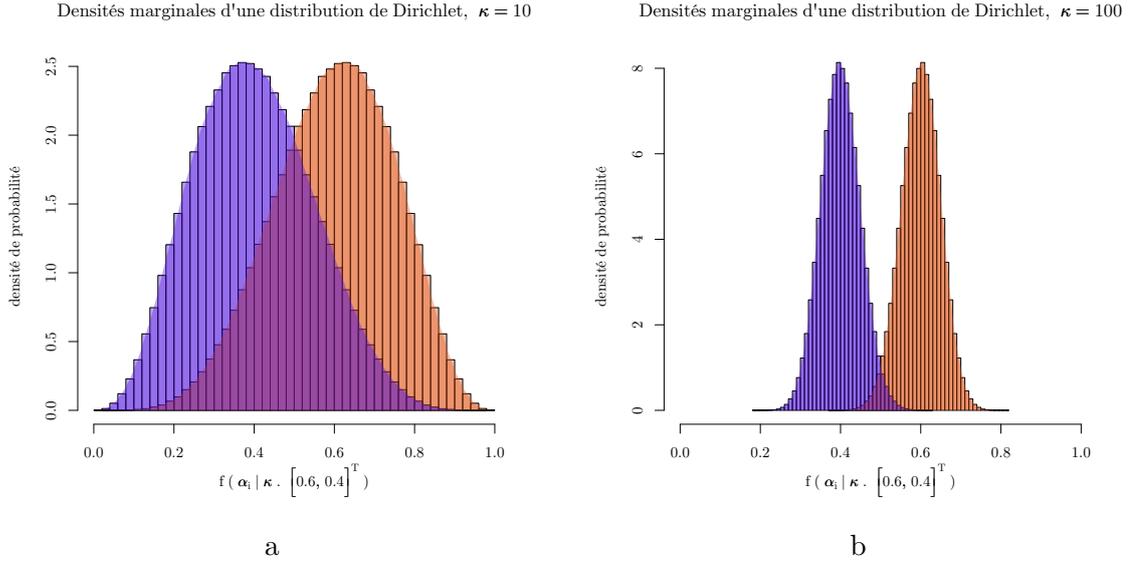


Figure V.2 Densités marginales d'une distribution de Dirichlet en dimension 2 pour différentes valeurs de κ .

2.1.3 Filtrage variationnel

Afin de pouvoir se placer dans le cadre du [théorème 2](#), l'indépendance conditionnelle entre l_t et T_t est imposée. On a ainsi $f(l_t, T_t | p_t) \approx \tilde{f}(l_t | p_t) \tilde{f}(T_t | p_t)$. Or le [modèle d'évolution V.9](#), pour $l_t \in \chi_{\mathbb{N}^c}$ s'écrit

$$\begin{aligned} f(l_t | l_{t-1}, T_t) &= \mathcal{M}u_{l_t}(T_t l_{t-1}) \\ &= \prod_{i=1}^c [T_t l_{t-1}]_i^{l_{i,t}} \end{aligned} \quad (\text{V.14})$$

où $[T_t l_{t-1}]_i$ représente le i -ème élément du vecteur colonne $T_t l_{t-1}$. On peut donc écrire

$$\begin{aligned} f(l_t | l_{t-1}, T_t) &= \prod_{i=1}^c \exp(l_{i,t} \ln [T_t l_{t-1}]_i) \\ &\propto \exp(l_t^T \ln(T_t) l_{t-1}). \end{aligned} \quad (\text{V.15})$$

Cette forme a été obtenue en utilisant le fait que l_t vaut soit zéro soit un. Une telle factorisation est nécessaire afin d'appliquer le théorème variationnel de ([Šmídl et Quinn, 2008](#)).

De même, le modèle d'évolution de T_t de l'[équation V.13](#) s'écrit, pour toute matrice stochastique¹ T_t ,

$$f(T_t | T_{t-1}) = \prod_{j=1}^c \left[\frac{\Gamma(\kappa)}{\prod_{k=1}^c \Gamma(\kappa t_{k,j,t-1})} \prod_{i=1}^c t_{i,j,t}^{\kappa t_{i,j,t-1} - 1} \right]. \quad (\text{V.16})$$

En prenant le logarithme naturel et en utilisant ses propriétés, on obtient alors

$$\ln(f(T_t | T_{t-1})) = \sum_{j=1}^c \left\{ \ln \left(\frac{\Gamma(\kappa)}{\prod_{k=1}^c \Gamma(\kappa t_{k,j,t-1})} \right) + \sum_{i=1}^c (\kappa t_{i,j,t-1} - 1) \ln(t_{i,j,t}) \right\}. \quad (\text{V.17})$$

¹ On appelle matrice stochastique une matrice dont la somme des éléments de chacune des colonnes vaut 1.

Alors en enlevant la constante $\Gamma(\kappa)$ qui sera retrouvée par normalisation des probabilités, on aboutit à la simplification

$$\ln(f(T_t | T_{t-1})) = - \sum_{i=1}^c \sum_{j=1}^c \ln(\Gamma(\kappa t_{i,j,t-1})) + \text{tr} [\ln T_t^T (\kappa T_{t-1} - 1_{c,c})] + \text{constante.} \quad (\text{V.18})$$

On peut donc exprimer le modèle d'évolution, en utilisant l'équation V.15 et V.18, sous la forme donnée dans Šmídl et Quinn (2008) (sans les détails de calcul)

$$\ln(f(l_t, T_t | l_{t-1}, T_{t-1})) = l_t^T \ln(T_t) l_{t-1} - \sum_{i=1}^c \sum_{j=1}^c \ln(\Gamma(\kappa t_{i,j,t-1})) + \text{tr} [\ln T_t^T (\kappa T_{t-1} - 1_{c,c})] + \text{Cste} \quad (\text{V.19})$$

Alors, en appliquant le théorème 2 comme expliqué dans Šmídl et Quinn (2008), et en écrivant le modèle d'observation sous la forme

$$f(p_t | l_t) \propto \exp(\rho \ln(p_t)^T l_t), \quad (\text{V.20})$$

on obtient les formes des distributions *a posteriori* (N.B. :d'après nos calculs les formes données dans Šmídl et Quinn (2008) contiennent des erreurs)

$$\alpha_t = p_t^\rho \circ \exp(\widehat{\ln(T_t)}^T \widehat{l_{t-1}}) \quad (\text{V.25})$$

$$\tilde{f}(l_t | p_t) = \mathcal{M}u(\alpha_t) \quad (\text{V.21})$$

$$Q_t = \kappa \widehat{T_{t-1}} + \widehat{l_t} \widehat{l_{t-1}}^T \quad (\text{V.26})$$

$$\tilde{f}(T_t | p_t) = \mathcal{D}i(Q_t) \quad (\text{V.22})$$

$$\beta_t = \alpha_{t-1} \circ \exp(\widehat{\ln(T_t)}^T \widehat{l_t}) \quad (\text{V.27})$$

$$\tilde{f}(l_{t-1} | p_t) = \mathcal{M}u(\beta_t) \quad (\text{V.23})$$

$$\begin{aligned} \phi(T_{t-1}) &= \text{tr} \left[\kappa \widehat{\ln T_t}^T (T_{t-1} - 1_{c,c}) \right] \\ &- \sum_{i=1}^c \sum_{j=1}^c \ln(\Gamma(\kappa t_{i,j,t-1})) \quad (\text{V.28}) \end{aligned}$$

$$\tilde{f}(T_{t-1} | p_t) \propto \exp(\phi(T_{t-1})) \quad (\text{V.24})$$

Les expressions V.21, V.22, V.23 sont entièrement déterminées par les paramètres dont les valeurs exactes sont données de manière immédiate par les formules V.25, V.26 et V.27. Cependant, comme des dépendances entre ces différents paramètres existent, ils sont calculés de manière itérative en donnant une valeur initiale arbitraire (l'utilisation de la valeur du paramètre au pas de temps précédent permet de converger plus rapidement vers la solution). Les valeurs des différents moments nécessaires à ces expressions sont données par les moments théoriques des lois obtenues. Ainsi on a

$$\begin{aligned} \widehat{l_t} &\propto \alpha_t \\ \widehat{l_{t-1}} &\propto \beta_t \\ \widehat{\ln t_{i,j,t}} &= \psi(q_{i,j,t}) - \psi(1_{1,c} Q_t 1_{c,1}) \end{aligned} \quad (\text{V.29})$$

où $\psi(\cdot)$ est la fonction digamma et $q_{i,j,t}$ est l'élément de la ligne i et de la colonne j de la matrice Q_t .

Contrairement aux trois formes précédentes, l'expression V.24 n'est pas une forme de distribution connue et nécessite l'évaluation de $\tilde{f}(T_{t-1} | p_t)$ de manière approchée en utilisant une grille décrivant l'espace des valeurs possibles de T_{t-1} . Si cette évaluation est possible en dimension deux, auquel cas T_{t-1} a deux degrés de liberté et on peut décrire relativement bien l'espace de la matrice stochastique T_{t-1} à l'aide d'une grille de 50×50 , les calculs deviennent vite très lourds dans les dimensions supérieures (grille de 50^6 points en dimension trois). Ce n'est pas acceptable dans notre objectif de temps réel. Afin de

remédier à ce problème, Šmídl et Quinn (2008) proposent d'utiliser une forme approchée de la distribution $\tilde{f}(T_{t-1} | p_t)$

$$\tilde{f}(T_{t-1} | p_t) \approx \tilde{f}(T_{t-1} | p_{t-1}) = \mathcal{D}i_{T_{t-1}}(Q_{t-1}). \quad (\text{V.30})$$

Celle-ci permet d'obtenir une formule exacte pour le calcul de chacun des paramètres décrivant entièrement les lois *a posteriori*.

2.2 Résumé de l'algorithme de filtrage des probabilités

Afin de clarifier la procédure permettant de filtrer une suite de probabilités observées selon la méthode exposée dans ce chapitre, nous avons résumé dans l'[algorithme V.1](#) les différentes étapes. Dans cet algorithme, on peut noter que les équations sont itérées cinq fois afin de converger vers la solution. Ce faible nombre d'itérations est néanmoins largement suffisant pour atteindre la convergence.

ENTRÉES : $P \in \mathbb{R}^{c \times N_t}$, $\alpha_0 \in \mathbb{R}^c$, $\beta_0 \in \mathbb{R}^c$, $T_0 \in \mathbb{R}^{c \times c}$
 PARAMÈTRES : $\rho \in \mathbb{N}$, $\kappa \in \mathbb{N}$
Hypothèses : $\sum_{i=1}^c [P]_{i,t} = \sum_{i=1}^c p_{i,t} = 1$, les observations sont des probabilités
 $\sum_{i=1}^c \alpha_{i,0} = 1$ et $\sum_{i=1}^c \beta_{i,0} = 1$
 T_0 est une matrice stochastique (colonnes unitaires)

Procédure : vbhmm

```

 $l_t \leftarrow \beta_0$ 
 $Q_t \leftarrow T_0$ 
 $\ln T_t \leftarrow \ln T_0$ 
tant que  $t \leq N_t$  faire
   $l_{t-1} \leftarrow l_t$ 
   $Q_{t-1} \leftarrow Q_t$ 
   $T_{t-1} \leftarrow \mathbb{E}(\mathcal{D}i_{T_{t-1}}(Q_{t-1}))$ 
  tant que  $i \leq 5$  faire
     $\alpha_t \leftarrow p_t^\rho \circ \exp(\ln T_t^T l_{t-1})$ 
     $l_t \leftarrow \alpha_t / \sum_{i=1}^c \alpha_{i,t}$ 
     $Q_t \leftarrow \kappa T_{t-1} + l_t l_{t-1}^T$ 
     $\ln T_t \leftarrow \psi(Q_t) - \psi(\sum_{i,j} Q_t)$ 
     $\beta_t \leftarrow l_{t-1} \circ \exp(\ln T_t^T l_t)$ 
     $l_{t-1} \leftarrow \beta_t / \sum_{i=1}^c \beta_{i,t}$ 
  fin tant que
  stocker  $l_t$ ,  $l_{t-1}$ ,  $T_t$  et  $T_{t-1}$ 
fin tant que

```

Fin de procédure

SORTIES : $\{l_t\}_{t \in [1..N_t]}$, $\{l_{t-1}\}_{t \in [1..N_t]}$, $\{T_t\}_{t \in [1..N_t]}$, $\{T_{t-1}\}_{t \in [1..N_t]}$

Algorithme V.1 Filtrage variationnel bayésien

Afin de pouvoir utiliser la méthode dans des systèmes réels, il est primordial de ne pas utiliser le calcul de la distribution de T_{t-1} . En effet, sur un ordinateur relativement puissant (processeur Xeon cadencé à 3 GHz), en considérant des observations en dimension 2, il faut en moyenne 8,6s afin d'inférer l_t tandis que le résultat est obtenu en moyenne en 0,006s dans le cas où l'on utilise l'approximation approchée. De plus, seule une différence de résultat infime a été observée entre les deux méthodes (erreur relative de l'ordre de

10^{-4}). Les versions qui ont permis de réaliser ces comparaisons étaient toutes deux codées en langage R.

Afin de rendre le temps de calcul nécessaire à l'inférence d'un état latent du sujet à partir d'une observation négligeable, nous avons implémenté une version de l'algorithme présenté en C. Nous avons ainsi encore pu gagner un facteur 10, en réduisant le temps de calcul moyen nécessaire au traitement d'une nouvelle observation p_t (en dimension 2 également) à 0,0008 s. Sachant qu'une nouvelle classification dans le domaine des interfaces cerveau-machines est généralement disponible toute les 100 ms au mieux, ces chiffres montrent que l'algorithme mis en œuvre dans ce chapitre peut sans problème être utilisé dans un système fonctionnant en temps réel.

2.3 Illustration du filtrage variationnel bayésien

Afin de bien comprendre le fonctionnement et l'intérêt de l'algorithme, nous présentons une illustration à partir de probabilités calculées lors d'une expérience réelle comportant deux tâches mentales distinctes (figure V.3). La figure montre les probabilités observées (en bleu) ainsi que les estimations de \hat{l}_t (en rouge) et \widehat{l}_{t-1} (en vert) pendant environ 200 secondes pour trois valeurs distinctes de κ . Dans cette illustration, la valeur de ρ a été fixée à deux, mais des valeurs légèrement différentes ne changent pas substantiellement les résultats. Les valeurs initiales des paramètres α_0 , β_0 et T_0 ont été choisies de façon qu'elles soient non-informatives ($\alpha_0 = \beta_0 = T_{i,0} = [0, 5; 0, 5]^T$).

La figure V.3 montre clairement l'influence de κ sur le filtrage. Lorsque $\kappa = 20$, on voit que l'apprentissage est extrêmement rapide (5 premières secondes), la matrice de transition estimée tend alors rapidement vers des valeurs en régime stationnaire telles que $\mathbb{P}(l_t = i \mid l_{t-1} = i) \gg \mathbb{P}(l_t = i \mid l_{t-1} = j, j \neq i)$. Cependant ces valeurs subissent de fortes variations au cours du temps. Dans le cas où $\kappa = 200$, on voit que l'apprentissage de la matrice de transition prend plus de temps (10 à 20 secondes). Cependant, lorsque celles-ci sont relativement stabilisées, elles ne changent pas considérablement, permettant au système d'entrer dans un régime relativement stationnaire. On peut observer sur la période de temps entre 170 et 190 s que la configuration $\kappa = 200$ permet plus de transition que dans le cas où $\kappa = 20$. Enfin, nous avons choisi le cas extrême $\kappa = 100000$ afin d'illustrer le fonctionnement de l'algorithme lorsque la matrice de transition ne varie presque pas à partir des valeurs initiales (ici non-informatives). Ce cas illustre deux choses : tout d'abord, dans le cas où la matrice de transition est connue, on peut choisir κ de sorte que l'évolution autour des valeurs initiales soit faible, ceci peut être très intéressant dans le cas de paradigmes relativement contraints imposant implicitement une matrice de transition fixe ; de plus, on peut fixer la matrice de transition de façon que les décisions soient uniquement influencées par les valeurs des observations.

Dans les différents cas de figure, il est intéressant d'observer l'inertie du système, c'est-à-dire la tendance du système à rester dans un état tandis que le seuillage des probabilités observées aurait engendré un état différent. Afin d'observer ce phénomène, nous avons matérialisé en pointillé noir le seuil de décision correspondant au principe de maximum de vraisemblance appliqué à chacune des fenêtres séparément. Par exemple, si l'on observe la période de temps entre 60 et 80 s, on voit que la première configuration $\kappa = 20$ n'entraîne aucun changement d'« état » de \widehat{l}_{t-1} , on peut donc parler d'une inertie importante. La deuxième configuration $\kappa = 200$ montre quant à elle des changements d'états de cette même variable. Enfin, le dernier cas extrême entraîne autant de variations que les observations. Les deux dernières illustrations sont donc représentatives de systèmes ayant respectivement une inertie moyenne et nulle.

En raison de l'influence analogue du paramètre ρ et d'une plage très restreinte de valeurs cohérentes, nous ne détaillons pas son rôle en détail. Entre les sujets, la valeur de ρ peut néanmoins varier entre 2 et 5. Les valeurs choisies ont été fixées selon le critère

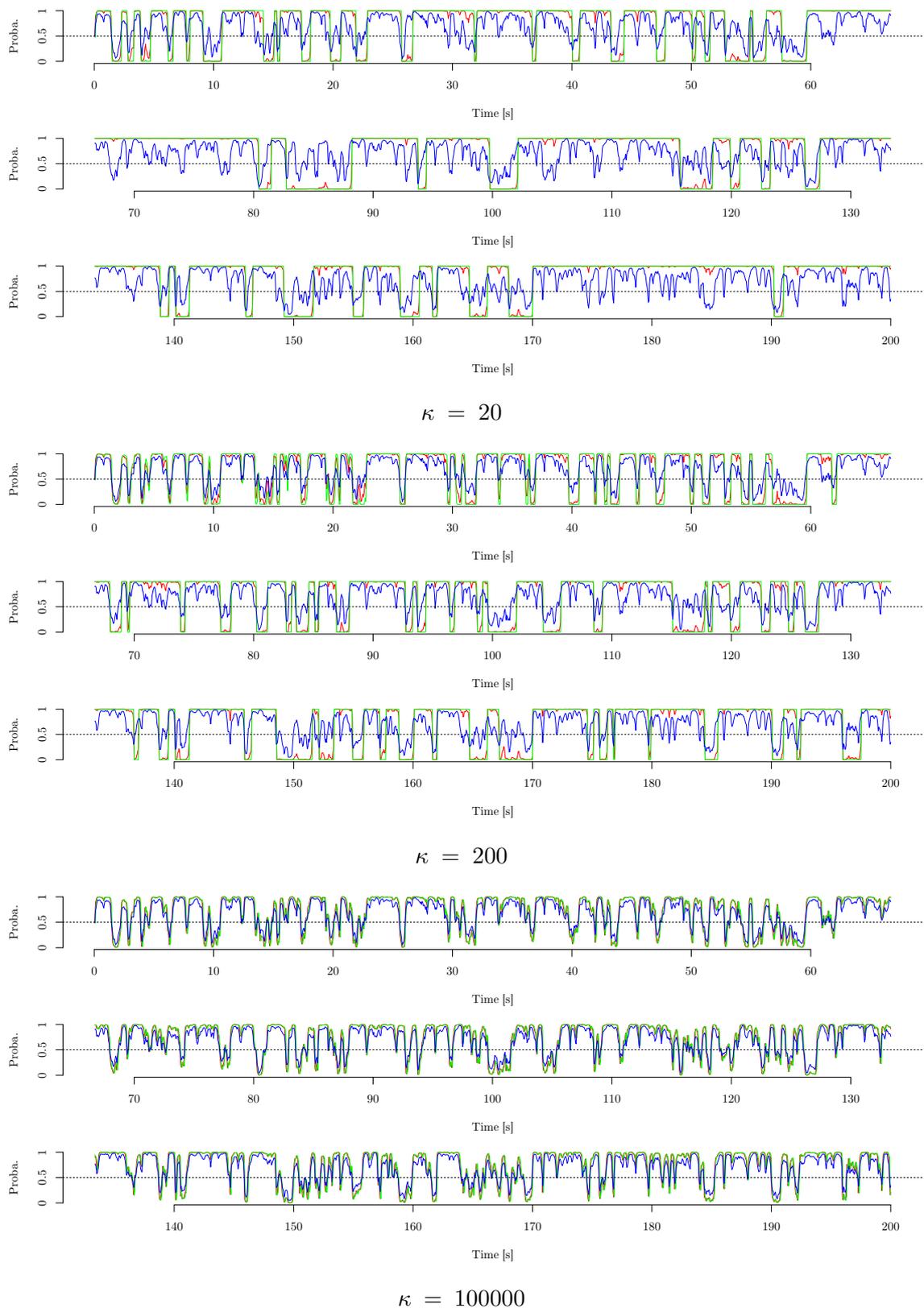


Figure V.3 Illustration du fonctionnement de l'algorithme à partir de probabilités issues de données réelles. De l'expérience qui comportait deux tâches, nous nous intéressons uniquement à la probabilité que le sujet soit dans l'une d'entre elle (l'autre peut être déduite à partir de celle-ci). Chacun des trois graphiques illustre le comportement de l'algorithme pour trois valeurs de κ distinctes. En bleu sont représentées les observations p_t , en rouge \widehat{l}_t et en vert \widehat{l}_{t-1} . Les conditions initiales ont été fixées de sorte qu'elles soient non-informatives.

suivant : de bonnes performances lors de la session d'entraînement conduit à des valeurs de ρ plus grandes. On met ainsi plus de poids sur les probabilités calculées par le classifieur et moins sur l'influence du modèle de Markov. Ce paramètre permet donc une légère adaptation du système à la maîtrise du sujet.

3 Méthode générale pour les ICMs

Nous allons à présent décrire comment intégrer l'algorithme précédent dans le flot de traitement d'une ICM. La méthode est très similaire à celle adoptée dans le [chapitre III](#), cependant, la méthode d'apprentissage du classifieur diffère quelque peu.

3.1 Phase d'entraînement

Les ICMs sans contrainte temporelle, de même que leurs homologues synchrones, doivent comporter une phase d'apprentissage qui permet de fixer les paramètres de la méthode de traitement (apprentissage du classifieur, choix des filtres spatiaux). Afin de choisir ces paramètres, on utilise pendant la phase d'entraînement un protocole synchrone. Cela permet de contrôler les périodes de réalisation des tâches mentales et ainsi de maîtriser plus précisément l'apprentissage des paramètres.

La phase d'entraînement est conforme à la méthode développée dans le [chapitre III](#) exceptée sur deux points importants. Premièrement, aucun filtre de prétraitement n'est utilisé. Les raisons de ce choix apparaîtront dans la suite. De plus, on ne désire pas classer un essai entier mais plutôt trouver l'intervalle de temps ainsi que la valeur du paramètre de régularisation qui conduisent aux taux de classification les plus élevés. Il est donc ici plus opportun d'utiliser une validation croisée afin de trouver la valeur du paramètre de régularisation ainsi que la fenêtre temporelle commune à chacun des essais qui permettent d'obtenir les meilleurs taux de classification.

Une fois obtenues la position de la meilleure fenêtre temporelle globale pour les essais ainsi que la valeur de régularisation adaptée ([algorithme V.2](#)), on utilise alors l'ensemble des données d'entraînement pour obtenir les filtres spatiaux ainsi les paramètres de la régression logistique sur la fenêtre idéale. Ces différents paramètres sont stockés afin d'être par la suite utilisés lors de la phase d'évaluation pendant laquelle aucune information temporelle ne sera connue. Notons que l'utilisation d'un ensemble d'entraînement réduit (lors de la validation croisée de l'[algorithme V.2](#)) pour trouver la meilleure valeur de régularisation qui sera utilisée pour le jeu d'entraînement total pose un problème. En effet, la taille des données influe sur la valeur optimale du paramètre de régularisation. Ce problème est abordé et on donne des indications pour le résoudre dans l'[annexe A](#).

3.2 Phase d'évaluation

La phase d'évaluation a ensuite pour but de classer de manière continue les données EEG. Pour ce faire, il suffit de classer les données de chaque fenêtre successive de la phase d'évaluation. Cette classification est réalisée à partir des sources obtenues par projection à l'aide des filtres calculés lors de la phase d'apprentissage et des paramètres du classifieur. Nous obtenons ainsi, pour chaque fenêtre d'analyse, une probabilité observée pour que le sujet soit dans chacune des tâches mentales de l'expérience. Ces probabilités observées sont ensuite filtrées selon la méthode expliquée au début de ce chapitre et en appliquant l'[algorithme V.1](#).

3.3 Résumé

Le [schéma V.4](#) résume les étapes d'apprentissage et d'évaluation en faisant apparaître le filtrage variationnel bayésien.

ENTRÉES : Données d'apprentissage $X \in \mathbb{R}^{N_i \times N_s}$, $(\mathcal{T}_i)_{i \in [1..N_j]}^j$ la suite des intervalles pendant lesquels le sujet réalisait la tâche j ($j \in [1..c]$),

PARAMÈTRES : N_{cv} nombre de répétitions de la procédure

N_e taille des fenêtres temporelles pour la classification

r_e taux de recouvrement des fenêtres

Procédure : apprentissage du paramètre de régularisation du classifieur et meilleur intervalle temporel.

tant que $k \leq N_{cv}$ **faire**

Choisir $\mathcal{T}r = \bigcup_{j=1}^c \mathcal{T}_{\mathcal{P}[1..N_j]}^j$ les essais d'entraînement (\mathcal{P} partie aléatoire)

entraîner les filtres spatiaux à l'aide de $\mathcal{T}r$ (équation III.23)

choisir les meilleurs filtres spatiaux (équation III.17)

projeter les données (équation III.22)

calculer la puissance de chaque source pour chaque fenêtre glissante de chaque essai

entraîner un classifieur par position de fenêtres à partir de $\mathcal{T}r$ (annexe A)

classifier chaque fenêtre de chaque essai de l'ensemble d'évaluation

stocker le taux de classification pour chaque fenêtre et chaque régularisation

fin tant que

Fin de procédure

SORTIES : position de la fenêtre temporelle et valeur de régularisation telles que le taux de classification soit maximum

Algorithme V.2 Apprentissage des paramètres lors de la phase d'entraînement. Lors de cette phase, le paradigme est synchrone, nous disposons donc des indications temporelles à propos des états mentaux du sujet.

4 Apprentissage de l'état « repos »

Lors de la phase d'évaluation, le système doit être capable de distinguer entre la réalisation d'une tâche mentale et un état de repos du sujet. Par conséquent, le système doit apprendre à reconnaître l'état de repos. Or, si la définition de l'état mental correspondant à la réalisation d'une tâche est relativement facile, il n'en est pas de même pour la définition de l'état de repos qui se fonde sur une opposition à un état d'activité. Nous proposons deux méthodes afin de réaliser cet entraînement, celles-ci seront ensuite comparées en utilisant les données réelles dans les paragraphes suivants.

4.1 Utilisation des périodes de repos

Une première approche consiste à utiliser les périodes entre chaque tâche comme des périodes définissant le repos. Lors de la phase d'entraînement (voir figure V.4), on considère alors une classe de plus que le nombre de classes proposées au sujet.

L'avantage de cette approche est la définition explicite de moments de repos qui permet d'obtenir une probabilité pour que le sujet soit dans un tel état. Cependant, cette approche fait deux hypothèses relativement fortes : tout d'abord que l'état de repos entre deux périodes d'activité n'est pas trop variable au cours de la session d'entraînement ; de plus que les états de repos au cours des périodes d'entraînement et d'évaluation sont relativement similaires. Cette deuxième condition implique que la concentration soit à peu près identique et que les paradigmes soient similaires entre les deux phases. Or il est courant que cette condition ne soit pas respectée, notamment car les contraintes imposées lors de la phase d'entraînement (paradigme synchrone) sont relâchées lors de la phase

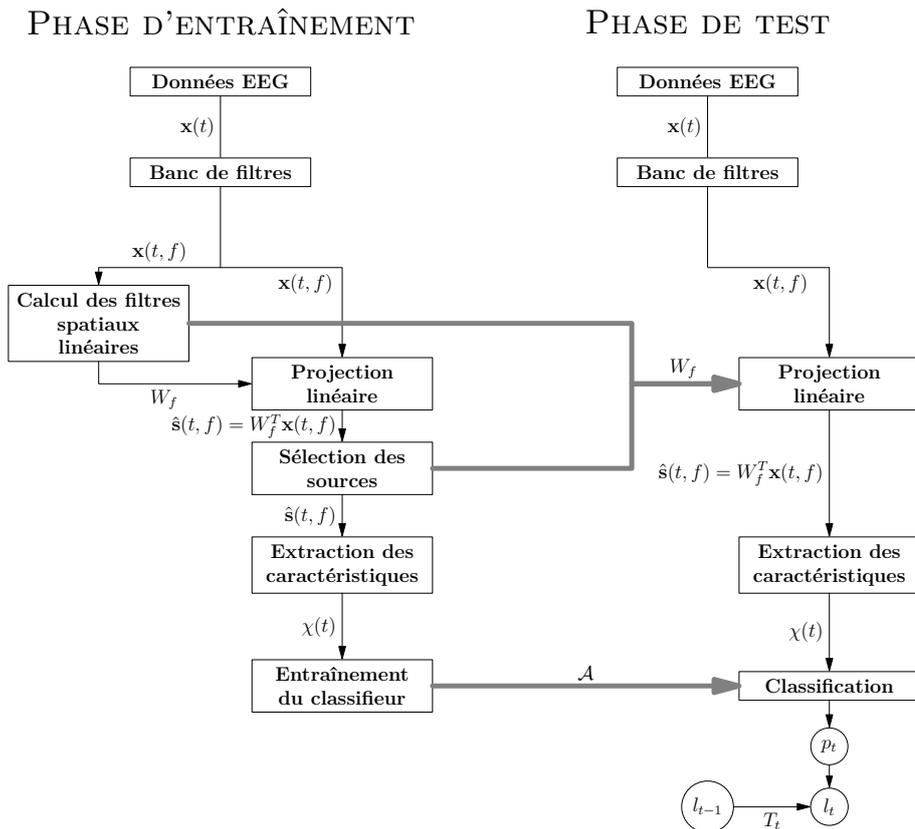


Figure V.4 Schéma bloc résumant les différentes phases de traitement dans les étapes d'entraînement et de test en incluant le filtrage des probabilités par la méthode présentée dans ce chapitre.

d'évaluation (pas de contrainte temporelle). La [figure V.5](#) illustre les périodes qui sont utilisées afin de définir l'état de repos.

Définition de l'état de repos

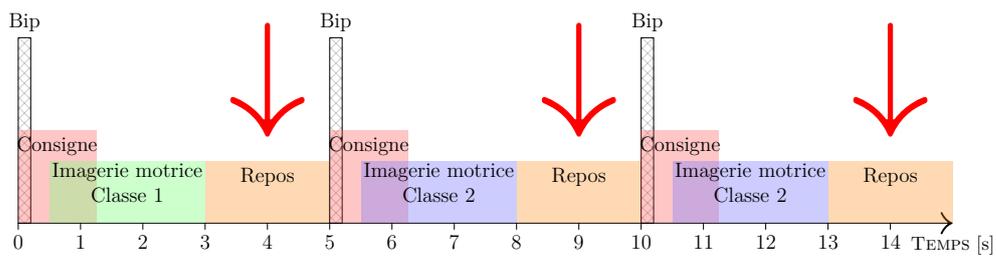


Figure V.5 Utilisation des périodes d'inactivité afin de définir l'état de repos.

4.2 Définition implicite

La deuxième méthode que nous proposons consiste à définir une probabilité *a priori* que le sujet soit au repos (ou symétriquement en activité). Si l'on définit \mathcal{I} (respectivement \mathcal{R}) l'événement « le sujet est en activité » (respectivement « le sujet est au repos »), alors, lorsqu'on entraîne les algorithmes en utilisant uniquement les tâches mentales « actives », on peut interpréter les probabilités résultantes comme des estimations des probabilités de chacune des classes **sachant que le sujet est en activité \mathcal{I}** . Or pendant la phase d'évaluation (qui ne contient pas d'information sur les moments pendant lesquels le sujet réalise sa tâche), on est intéressé en réalité par les probabilités d'être dans chacune des

classes. Mathématiquement, si l'on appelle $\mathbb{P}(c_i)$ la probabilité que le sujet soit en train de réaliser la tâche c_i , $\mathbb{P}(\mathcal{I})$ la probabilité que le sujet soit en activité et $\mathbb{P}(\mathcal{R})$ la probabilité qu'il soit au repos, alors on peut écrire

$$\mathbb{P}(\mathcal{R}) = 1 - \mathbb{P}(\mathcal{I}) = 1 - \sum_{i=1}^c \mathbb{P}(c_i). \quad (\text{V.31})$$

Comme on est intéressé par $\mathbb{P}(c_i)$, on peut utiliser de manière très simple la règle de Bayes $\mathbb{P}(c_i) = \mathbb{P}(c_i, \mathcal{I}) = \mathbb{P}(c_i | \mathcal{I}) \mathbb{P}(\mathcal{I})$ et ainsi obtenir, pour une quelconque paire de classes $(c_i, c_j)_{i \neq j}$

$$\mathbb{P}(c_i) - \mathbb{P}(c_j) = \mathbb{P}(\mathcal{I}) [\mathbb{P}(c_i | \mathcal{I}) - \mathbb{P}(c_j | \mathcal{I})]. \quad (\text{V.32})$$

Cette dernière expression montre clairement que, lorsque la différence entre la probabilité de deux classes est faible, cela peut être dû au fait que le classifieur n'arrive pas à faire la différence entre les deux classes, ou bien que le sujet a de fortes chances d'être au repos. Dans la suite de ce chapitre, nous allons utiliser cette identité en considérant $\mathbb{P}(\mathcal{I})$ comme un paramètre du modèle considéré. Cette valeur permettra, à partir des probabilités observées $\mathbb{P}(c_i | \mathcal{I})$, d'obtenir les probabilités $\mathbb{P}(c_i)$.

Cette approche a l'avantage de fournir un paramètre qui peut être adapté en fonction de l'application ICM visée. Ainsi, dans le cas d'un système destiné à envoyer beaucoup de commandes par unité de temps, le paramètre peut être volontairement fixé à une valeur importante. De manière générale, le paradigme utilisé permet d'avoir facilement une idée de la valeur adaptée de ce paramètre. Nous verrons dans la suite qu'il peut aussi permettre d'ajuster le taux de fausses détections selon l'exigence que l'on désire. L'inconvénient de cette approche est d'imposer au classifieur d'avoir un comportement équilibré entre les différentes classes. En effet, si l'une des classes a tendance à être plus facilement reconnue par le système, alors cela va induire un biais important et rendre inadéquate la valeur *a priori* du paramètre.

5 Données expérimentales

Afin de tester la méthode présentée dans ce chapitre, ainsi que ses diverses déclinaisons, nous avons utilisé deux jeux de données qui sont disponibles librement sur Internet. Ceux-ci correspondent aux jeux de données utilisés lors des deux dernières « BCI Competitions » qui ont eu lieu respectivement en 2005 et 2008.

5.1 « BCI Competition III, dataset 4b »

Lors de la troisième compétition internationale sur les ICMs, l'un des jeux de données s'intitulait « Motor imagery: uncued classifier application » (littéralement « Imagerie motrice : classification sans information temporelle »)¹.

Ce jeu de données (Dornhege et al., 2004) comporte un unique sujet en bonne santé. Sept sessions d'enregistrement ont été réalisées par celui-ci. Les trois premières sessions sont les sessions d'entraînement, une information visuelle est fournie au sujet, lui indiquant la tâche à réaliser parmi les tâches L (main gauche), F (pied droit) et Z (langue). La tâche d'imagerie motrice dure 3,5 s. Entre chaque tâche, une période de durée variable (entre 1,75 et 2,25 s) permet au sujet de se relaxer. Les trois dernières sessions correspondent aux sessions d'évaluation. Afin de simuler un paradigme non contraint dans le temps mais qui permette néanmoins d'avoir une vérité sur l'état mental du sujet, les organisateurs

¹ Les données ainsi que la description du paradigme sont disponibles sur le [Site internet officiel de la compétition Interfaces Cerveau-Machines, édition III](#).

ont décidé d'utiliser des indices de début de tâches acoustiques plutôt que visuels. Ainsi, on s'assure que ce n'est pas la réponse engendrée par le stimulus qui a été apprise mais bien l'imagination motrice. La durée des tâches mentales lors de la période d'évaluation varie entre 1,5 et 8 s.

L'enregistrement est réalisé à l'aide d'un amplificateur BrainAmp comportant 128 électrodes Ag/Cl. L'activité cérébrale n'est enregistrée qu'au niveau de 118 positions correspondant au système 10/20 étendu (Jurcak et al., 2007). Les signaux ont été enregistrés à une fréquence d'échantillonnage de 1000 Hz et filtrés entre 0,5 et 200 Hz. Dans la suite du chapitre, nous utiliserons une version sous-échantillonnée des signaux (à 100 Hz).

Notons qu'un seul candidat a participé à la compétition pour ce jeu de données particulier et qu'aucun vainqueur n'a été désigné.

5.2 « BCI Competition IV, dataset 1 »

Lors de la compétition suivante¹, un jeu de données similaire a été mis en jeu. Le nombre de sujets a néanmoins été largement augmenté (quatre sujets réels) et le paradigme légèrement modifié (Blankertz et al., 2007).

Le sujet devait choisir deux tâches parmi l'imagerie de la main gauche, de la main droite, du pied droit, du pied gauche ou bien encore des deux pieds. Une période de fixation, avant la présentation du stimulus a été insérée dans le déroulement d'un essai. Quand au déroulement de la session de test, il est similaire en tout point à celui des données précédentes. L'enregistrement est réalisée à l'aide d'un amplificateur BrainAmp comportant 128 électrodes Ag/Cl. Mais cette fois seules 59 positions ont été utilisées. Nous utilisons à nouveau une version des données sous-échantillonnées à 100 Hz afin de diminuer la lourdeur des traitement engendrés. Enfin, notons que ce jeu de données comporte trois « sujets artificiels », dont les signaux ont été générés par des modèles mathématiques. Les organisateurs n'ont pour l'instant pas fourni la méthode qui a permis de générer ces données.

6 Résultats

Afin d'évaluer l'apport des méthodes de ce chapitre sur des données réelles, nous appliquons nos algorithmes aux deux jeux de données présentés précédemment. La question de l'évaluation des méthodes est abordée et nous proposons un ensemble de mesures afin de rendre compte du fonctionnement du système.

6.1 Méthode d'évaluation

6.1.1 Erreur quadratique moyenne

Lors de ces deux compétitions, les organisateurs ont décidé d'utiliser l'erreur quadratique moyenne afin d'évaluer les différentes soumissions de manière quantitative. On affecte à une des deux classes la valeur +1 et à l'autre la valeur -1. Pour chaque instant t de la phase d'évaluation, on connaît l'état mental théorique du sujet $\text{True}(t)$ qui est une valeur appartenant à l'ensemble $\{-1, 0, 1\}$. Si l'on appelle $c(t)$ la valeur estimée de l'état mental du sujet donné par la méthode de classification à l'instant t , alors le critère de l'erreur quadratique moyenne s'écrit

¹ Les données ainsi que la description du paradigme sont disponibles sur le [Site internet officiel de la compétition Interfaces Cerveau-Machines, édition IV](#).

$$\rho_{\text{quad}} = \frac{1}{T} \sum_{t=1}^T (\text{True}(t) - c(t))^2. \quad (\text{V.33})$$

Si ce critère a l'avantage d'être très simple à calculer, il a néanmoins l'inconvénient de ne rendre compte que très partiellement des différentes erreurs produites par la méthode de classification. De plus, cette méthode introduit un biais important dans la mesure où elle favorise les systèmes qui ont tendance à ne jamais faire de décision entre +1 et -1 et à rester toujours dans l'état repos. Ainsi, d'excellents résultats pouvaient être obtenus par des candidats ayant envoyé un vecteur ne contenant que des zéros. C'est pourquoi, en complément de cette première méthode d'évaluation « officielle », nous proposons une deuxième méthode basée sur l'analyse de la matrice de confusion.

6.1.2 Matrice de confusion

La notion de matrice de confusion a déjà été définie dans la [définition III.20](#). Dans le cas précis, si l'on appelle $l_t \in \{-1, 0, 1\}$ l'état mental estimé du sujet à l'instant t et $\text{True}(t)$ l'état réel du sujet au même instant, alors on peut définir la matrice de confusion

$$C_{\text{Préd., Réel}} = \begin{pmatrix} p_{0,0} & p_{0,-1} & p_{0,1} \\ p_{-1,0} & p_{-1,-1} & p_{-1,1} \\ p_{1,0} & p_{1,-1} & p_{1,1} \end{pmatrix} \quad (\text{V.34})$$

où $p_{a,b}$ représente la probabilité de prédire a alors que le sujet était dans l'état b . A partir de cette matrice de confusion, on peut alors définir trois mesures qui caractérisent ce que nous avons appelé l'inertie ainsi que la précision du système. Ces mesures sont définies sur la [figure V.6](#).

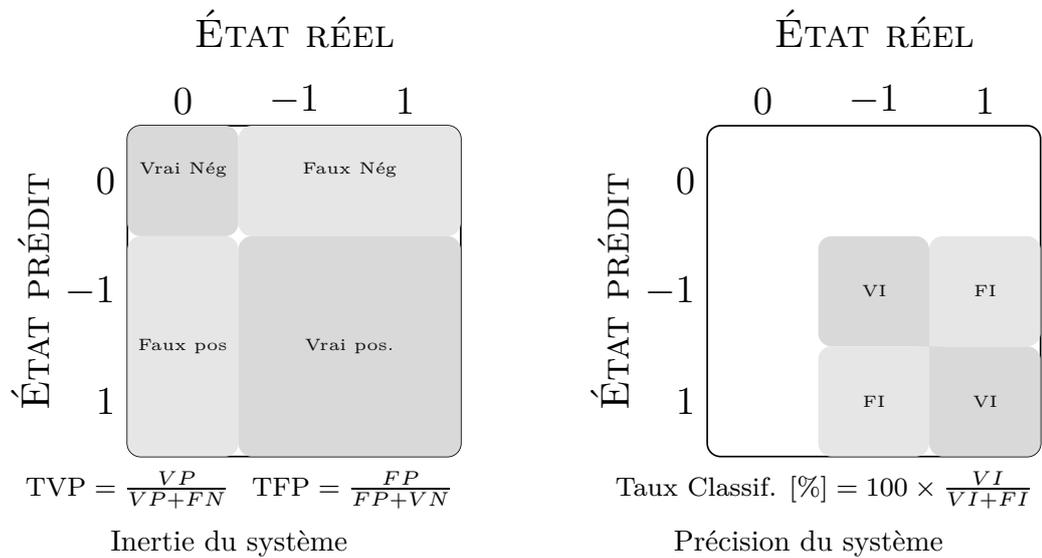


Figure V.6 Utilisation de la matrice de confusion pour évaluer les performances d'un système de classification. À gauche, les proportions de vrais négatifs (Vrai Nég.), faux négatifs (Faux Nég.), faux positifs (Faux pos.) et vrais positifs (Vrai pos.) afin de définir le taux de vrais positifs (TVP) ainsi que le taux de faux positifs (TFP). À droite, on utilise les proportions de vrais imageries (VI) et fausses imageries (FI) afin de définir le taux de bonnes classifications. Alors que les deux premières mesures tentent de rendre compte de l'inertie du système (sa propension à passer dans un état d'activité), la deuxième rend compte de sa précision lorsqu'il passe en activité (la qualité de ses classifications parmi les tâches proposées).

6.1.3 Caractéristique de fonctionnement du récepteur

Dans le cas où l'état de repos est défini en fonction d'un paramètre (définition implicite), alors on peut observer les différentes valeurs des mesures définies dans la [figure V.6](#) lorsque la valeur du paramètre parcourt un certain intervalle. Ici le paramètre $\mathbb{P}(\mathcal{I})$ peut prendre l'ensemble des valeurs réelles entre 0 et 1. Pour chacune des valeurs du paramètre, on obtient les mesures du système. Ces informations définissent ce qu'on appelle la *Caractéristique de fonctionnement du récepteur* ou en anglais *Receiver Operating Characteristic*. Nous l'appellerons courbe ROC.

6.1.4 Proportion de points mal classés

Bien que les deux mesures précédentes permettent d'avoir une évaluation très complète du comportement du système, elles s'avèrent peu pratiques pour comparer de manière quantitative deux méthodes. Une solution consiste à considérer l'aire sous la courbe ROC mais la question se pose alors de savoir comment intégrer la mesure de la précision du système. Afin de pallier ces difficultés, nous définissons la mesure

$$\rho_{\text{cat}} = \frac{\text{Nb de points mal classés}}{\text{Nb de points total}}. \quad (\text{V.35})$$

Cette mesure est appelée ρ_{cat} car la variable décrivant l'état mental du sujet est considérée comme une variable catégorielle plutôt que quantitative.

6.2 « BCI Competition III »

6.2.1 Choix des paramètres par validation croisée

Afin de choisir le meilleur intervalle d'entraînement ainsi que la meilleure valeur du paramètre de régularisation, on applique l'[algorithme V.2](#) en utilisant uniquement la session d'entraînement. Les bandes de fréquences choisies pour déterminer les filtres spatiaux recouvrent la bande entre 7 et 25 Hz, sont de largeur 2 Hz et de bandes passantes disjointes. Afin de calculer les caractéristiques de chacune des sources, on utilise une fenêtre de longueur 1 s. Les différentes fenêtres considérées sont définies par (on rappelle que la fréquence d'échantillonnage est de 100 Hz) les intervalles

$$\mathcal{F} = \{[a; a + 100 - 1] \mid a = 20k, k \in \mathbb{N}, a + 100 < 350\}. \quad (\text{V.36})$$

C'est-à-dire les fenêtres successives $[0; 99]$, $[20; 119]$, $[40; 139]$... Enfin, on considère seulement un ensemble fini de valeurs de régularisation possibles. Cet ensemble est $\{0,01; 0,03; 0,04; 0,05; 0,1\}$. Les performances obtenues pour ce sujet sont excellentes puisqu'on obtient des scores de bonne classification de l'ordre de 99 % pour les meilleurs paramètres. Ces scores sont obtenus avec un paramètre de régularisation de 0,05 et une fenêtre positionnée entre 1,3 et 2,3 s après le début de la tâche. Ces paramètres sont à présent fixés pour la suite de l'analyse de ce sujet. Notons également que le nombre de filtres spatiaux est fixé à 8 dans toutes les analyses qui suivent.

6.2.2 Apprentissage de l'état de repos

On s'intéresse tout d'abord à l'apprentissage de l'état de repos. Rappelons que deux méthodes différentes ont été proposées (voir précédemment). Dans le cas de la première méthode (apprentissage de l'état de repos), l'application de la procédure présentée dans la [figure V.4](#) (sans utiliser l'inférence par modèle de Markov) nous donne, pour chaque instant t , une probabilité pour que le sujet soit dans chacun des états -1, 0 ou 1. A partir

de ces trois probabilités $p_{-1}(t)$, $p_0(t)$ et $p_1(t)$, on crée un signal unidimensionnel $\gamma(t)$ tel que

$$\gamma_{\text{repos}}(t) = 2 \left(\frac{p_1(t)}{1 - p_0(t)} - \frac{1}{2} \right). \quad (\text{V.37})$$

Cette formule est construite de sorte que $\gamma_{\text{repos}}(t) \in [-1, 1]$ et $\gamma_{\text{repos}}(t) \rightarrow 1$ quand $p_1(t) \rightarrow 1$, $\gamma_{\text{repos}}(t) \rightarrow -1$ quand $p_{-1}(t) \rightarrow 1$ et $\gamma_{\text{repos}}(t) \rightarrow 0$ quand $p_0(t) \rightarrow 1$ ou $p_{-1}(t) \approx p_1(t)$. Remarquons que $\frac{p_1(t)}{1 - p_0(t)}$ peut s'interpréter comme la probabilité d'être dans l'état 1 sachant que le sujet est en activité (d'après la formule de Bayes).

Tandis que, dans le cas de la définition implicite de la probabilité de repos, nous obtenons seulement $\mathbb{P}(1 | \mathcal{I})$ et $\mathbb{P}(-1 | \mathcal{I})$. On définit alors de manière analogue

$$\gamma_{\text{impl}}(t) = 2 \left(\mathbb{P}(1 | \mathcal{I}) - \frac{1}{2} \right). \quad (\text{V.38})$$

Les erreurs quadratiques moyennes entre $\gamma(t)$ et les états mentaux théoriques du sujet sont alors consignées dans le [tableau V.1](#). Ces résultats montrent une assez large supériorité de la méthode fondée sur l'apprentissage explicite de l'état de repos.

$\frac{1}{T} \sum_{t=1}^T (\mathbf{True}(t) - \gamma_{\text{repos}}(t))^2$	$\frac{1}{T} \sum_{t=1}^T (\mathbf{True}(t) - \gamma_{\text{impl}}(t))^2$
0.222	0.286

Tableau V.1 Erreurs quadratiques moyennes pour chacune des deux méthodes d'apprentissage de l'état de repos.

A présent, si l'on s'intéresse aux mesures définies sur la [figure V.6](#), on peut tracer la ROC en faisant varier la probabilité d'activité *a priori*. Afin d'obtenir les matrices de confusion nécessaires à la définition de ces mesures, dans le cas où l'état de repos est appris, on considère les probabilités fournies par le classifieur $p_{-1}(t)$, $p_0(t)$ et $p_1(t)$ et on attribue à chaque instant t la classe de la fenêtre considérée

$$\text{Etat mental du sujet}(t) = \arg \max_{i \in \{-1, 0, 1\}} p_i(t). \quad (\text{V.39})$$

Dans le cas où le repos est implicite (déterminé par le choix de $\mathbb{P}(\mathcal{I})$), on utilise les propriétés rappelées dans l'[équation V.32](#) afin d'obtenir $p_{-1}(t)$, $p_0(t)$ et $p_1(t)$. De même que dans le cas du repos appris, on peut alors appliquer la [règle de décision V.39](#) afin d'obtenir la matrice de confusion en comparant les décisions par rapport aux états théoriques du sujet.

Les résultats correspondants à cette évaluation sont résumés dans la [figure V.7](#). L'ensemble de la courbe ROC (méthode d'apprentissage implicite du repos) ainsi que le point correspondant aux résultats pour l'apprentissage explicite de l'état de repos se trouvent dans la partie supérieure du graphe, montrant en premier lieu l'intérêt de chacun des systèmes par rapport à une classification basée sur la chance (système inutile). On peut observer de plus que la méthode basée sur le repos appris permet d'obtenir des performances largement supérieures au repos implicite. Ainsi, la méthode repos appris correspond à un taux de faux positifs de 0,42 et un taux de vrais positifs correspondant de 0,88. Alors que, pour ce même taux de faux positifs, la méthode repos implicite engendre un taux de vrais positifs de seulement 0,62. Les précisions de chacune des méthodes sont excellentes, supérieures à 90 %. Ces précisions montrent que la discrimination entre les deux tâches proposées ne pose pas de réels problèmes au système, lorsque celui-ci détecte une période d'activité, il est rare qu'il se trompe sur le type d'activité que le sujet est en train de réaliser.

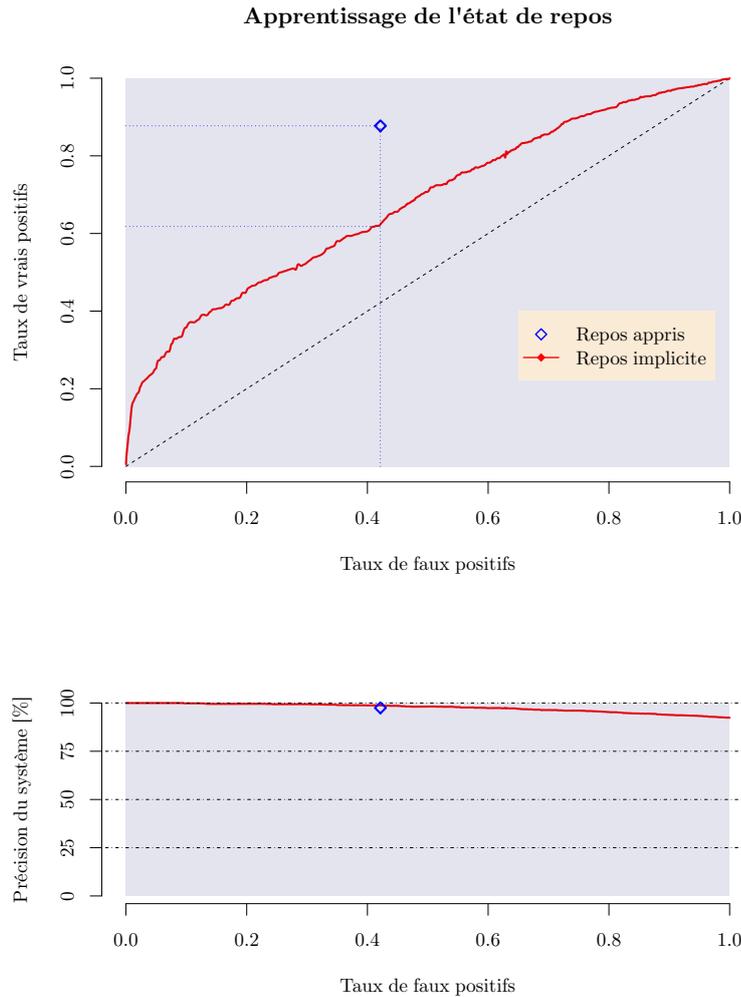


Figure V.7 En haut : en rouge la courbe ROC correspondant à la méthode de l'apprentissage implicite du repos dans le cas où le paramètre $\mathbb{P}(\mathcal{I})$ varie entre -1 et 1. En bleu, on a positionné le point correspondant à la méthode du repos appris. En bas : précision du système, c'est-à-dire le taux de bonnes classification lorsque le système est en activité. L'interprétation des couleurs est la même que pour la figure du haut.

6.2.3 Utilisation du modèle de Markov

On s'intéresse à présent à l'intérêt du filtrage variationnel des probabilités afin d'augmenter les performances du système. Pour chacune des méthodes (repos appris ou repos implicite), les probabilités $p_{-1}(t)$, $p_0(t)$ et $p_1(t)$ sont filtrées par l'algorithme V.1. Les différents paramètres utilisés sont $\rho = 3$ et $\kappa = 300$. Ceux-ci ont été fixés afin d'utiliser les informations temporelles connues sur les tâches (une tâche dure entre 1,5 et 8 s) et en se basant également sur les résultats présentés sur la figure V.3. Les densités initiales nécessaires à l'algorithme ont été choisies de manière à être non-informatives. Ainsi, on a choisi $\alpha_0 = \beta_0 = [1/3; 1/3, 1/3]^T$, et $T_0 = [[0, 7; 0, 15; 0, 15]^T; [0, 15; 0, 7; 0, 15]^T; [0, 15; 0, 15; 0, 7]^T]$.

Les courbes ROC dans le cas du repos implicite ainsi que les points de mesure correspondant à la méthode du repos appris sont présentés dans la figure V.8. On voit que l'apport du filtrage variationnel des probabilités est particulièrement clair pour des taux de faux positifs supérieurs à 0,15. Dans le cas des faibles taux de faux positifs, l'apprentissage de la matrice de transition converge vers un état dans lequel seul l'état de repos est réellement privilégié, rendant inadéquat le filtrage. Dans le cas du repos appris, l'introduction

du filtrage variationnel permet de diminuer le taux de faux positifs tout en conservant un taux de vrais positifs relativement similaire.

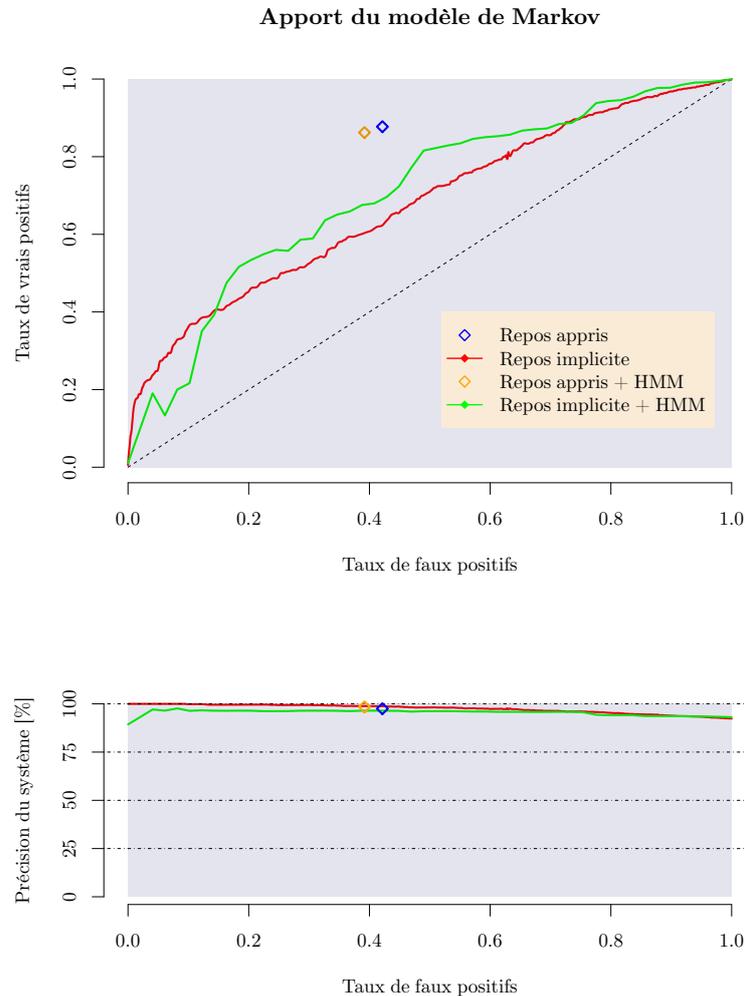


Figure V.8 En haut : en bleu, le point correspondant à la méthode du repos appris ; en orange la même séquence probabilités a été filtrée en utilisant le filtrage variationnel des probabilités ($\rho = 3$, $\kappa = 300$). En orange et rouge, les courbes ROC obtenues en faisant varier le paramètre $\mathbb{P}(\mathcal{I})$, respectivement en utilisant le seuillage des probabilités ou leur version régularisée (filtrage variationnel). En bas : précisions du système dans chacun des quatre cas considérés.

Remarque 2 (MESURES ET SIGNAUX). *Seules quelques combinaisons de signaux-mesures sont présentées dans ce chapitre parmi les différentes possibles. Ainsi, on pourrait également calculer l'erreur quadratique entre l'état mental du sujet (prenant les valeurs -1 , 0 ou 1) et les décisions obtenues par seuillage des probabilités selon le critère V.39. Cependant, une telle combinaison s'avère inadaptée. En témoignent nos résultats obtenus lors de la « BCI Competition IV », dataset 1, dans laquelle nous avons utilisé une telle méthode, méconnaissant alors le critère utilisé pour évaluer quantitativement les soumissions. Nos places furent respectivement $11^{\text{ème}}/24$ pour les données issues de sujets réels et $4^{\text{ème}}/24$ dans le cas des données artificielles. En utilisant les probabilités directement en sortie du classifieur (sans seuillage), nos places auraient été $4^{\text{ème}}/24$ pour les données issues de sujets réels et $1^{\text{er}}/24$ dans le cas des données artificielles. Ainsi, dans toute la suite, nous considérerons le critère de l'erreur quadratique en association avec un signal de probabilité et l'inertie/précision du système dans le cas des décisions. Le filtrage variationnel des*

probabilités, par construction, fournit des estimations qui sont quasiment des probabilités seuillées, la mesure adaptée est donc l'inertie/précision.

6.3 « BCI Competition IV »

Ce jeu de donnée est composé de quatre sujets réels et trois sujets dont les signaux ont été générés artificiellement. Afin de permettre la comparaison avec les résultats de la compétition, nous conservons la même notation. Les sujets sont indicés de *a* à *g*. Les sujets réels sont *a*, *b*, *f* et *g* alors que les sujets artificiels sont *c*, *d* et *e*.

6.3.1 Choix des paramètres

De même que pour les données précédentes, on commence par choisir la position optimale de la fenêtre qui servira pour entraîner le classifieur ainsi que le paramètre de régularisation du classifieur. La longueur des fenêtres glissantes considérées sont cette fois de 1,5s (taux de recouvrement de 80%). Le nombre de filtres spatiaux est de $L = 8$ et les mêmes bandes de fréquence que précédemment sont utilisées. Les paramètres obtenus par application de l'algorithme V.2 sont rapportés dans le tableau V.2

	Meil. Fenêtre	Régul.	Nbre filtres	Perf. [%]
Suj. a	1,3 → 2,8 s	0,2	8	88,5
Suj. b	2,5 → 4,5 s	0,2	8	80,5
Suj. c	1,5 → 3,0 s	0,2	8	92,5
Suj. d	1,9 → 3,4 s	0,1	8	99,0
Suj. e	2,5 → 4,0 s	0,1	8	100,0
Suj. f	0,7 → 2,2 s	0,2	8	95,3
Suj. g	0,9 → 2,4 s	0,1	8	94,8

Tableau V.2 Valeurs des différents paramètres choisis arbitrairement (nombre de filtres spatiaux) ou fixés par validation croisée (meilleure fenêtre et paramètre de régularisation du classifieur. La dernière colonne indique le taux de bonnes classification obtenu au cours de la validation croisée correspondante. Sur fond blanc sont indiqués les sujets réels, les sujets artificiels sont sur fond grisé.

6.3.2 Apprentissage de l'état de repos

De la même façon que pour les données de la « BCI Competition III », nous nous intéressons tout d'abord à l'influence de la méthode d'apprentissage de l'état de repos. Nous utilisons l'erreur quadratique pour quantifier les résultats. Ceux-ci sont présentés dans le tableau V.3. Dans le cas des sujets réels, les résultats obtenus sont relativement similaires à ceux obtenus lors de la compétition par les trois meilleurs compétiteurs. Nos résultats sont cependant clairement meilleurs dans le cas des données simulées artificiellement. Il est difficile d'interpréter pour le moment ce résultat dans l'attente de la publication des méthodes qui ont permis de générer ces données.

On peut néanmoins noter que la différence qui était apparue flagrante sur le jeu de données précédent entre les deux méthodes d'apprentissage s'est effacée à la lumière de ces nouveaux résultats. Ces deux méthodes apparaissent en effet très similaires à l'exception des sujets *b* et *c*. On ne peut donc pas conclure à la supériorité de l'une ou l'autre des méthodes. Notons cependant que l'intérêt de la méthode d'apprentissage implicite de l'état de repos fournit au concepteur du système un paramètre qui peut être modifié en temps réel afin de jouer sur l'inertie du système. Cet avantage peut s'avérer intéressant pour

certaines systèmes. On peut enfin observer une nette supériorité des résultats obtenus à partir des données simulées en comparaison de ceux obtenus à partir des sujets réels. Ces observations rappellent la difficulté de concevoir un modèle qui permet de générer des données suffisamment complexes pour ressembler aux données enregistrées sur des êtres humains.

	$\frac{1}{T} \sum (\mathbf{True}(t) - \gamma_{\text{repos}}(t))^2$	$\frac{1}{T} \sum (\mathbf{True}(t) - \gamma_{\text{impl}}(t))^2$	1 ^{er}	2 ^{ème}	3 ^{ème}
Suj. a	0,40	0,40	0,40	0,35	0,39
Suj. b	0,45	0,52	0,42	0,46	0,43
Suj. c	0,43	0,32	0,33	0,43	0,44
Suj. d	0,13	0,13	0,23	0,27	0,28
Suj. e	0,14	0,15	0,28	0,23	0,31
Suj. f	0,45	0,44	0,42	0,41	0,40
Suj. g	0,32	0,32	0,29	0,31	0,36
MOYENNE	0,407	0,425	0,382	0,383	0,397
MOYENNE	0,238	0,206	0,281	0,311	0,348

Tableau V.3 Erreurs quadratiques obtenues pour chaque méthode d'apprentissage de l'état de repos et chaque sujet. Ce tableau contient également les résultats obtenus par les trois meilleurs compétiteurs (trois dernières colonnes). Sur fond blanc sont rapportés tous les résultats relatifs aux sujets réels. Sur fond grisé, les résultats pour les données générées artificiellement.

6.3.3 Utilisation du filtrage variationnel

De même que pour le jeu de données précédent, nous nous intéressons à présent à l'apport du filtrage variationnel. Nous présentons uniquement les résultats relatifs aux quatre derniers sujets (les trois premiers étant relativement similaires) dans la [figure V.9](#). Afin d'évaluer l'apport du filtrage variationnel, nous ne considérons ici que le cas où l'état de repos est défini de manière implicite.

D'une manière générale, les résultats régularisés par filtrage variationnel sont au moins équivalents aux résultats bruts en sortie du classifieur. Les résultats pour les sujets *d* et *g* sont par exemple sensiblement égaux alors que les résultats des sujets *e* et *f* sont légèrement meilleurs en utilisant le modèle de Markov.

Afin d'évaluer quantitativement les résultats, nous considérons la mesure catégorielle définie dans l'[équation V.35](#). Nous utilisons la méthode d'apprentissage explicite de l'état de repos. Les résultats sont résumés dans le [tableau V.4](#).

Ce tableau montre que des améliorations minimales sont obtenues pour les sujets réels et légèrement plus importantes pour les données simulées. La différence de moyenne n'est pas significativement significative. Cependant, il faut noter que, afin de comparer les étiquettes prédites et les étiquettes réelles, nous avons interpolé (interpolation d'ordre zéro), dans le cas des données de la quatrième compétition, les décisions afin d'en avoir une à chaque instant *t*. Comme les étiquettes réelles sont fournies à une fréquence d'échantillonnage de 1000 Hz, le nombre d'étiquettes total est de plus de 2 millions pour chacun des sujets. Ainsi, une différence de 0,001 correspond en réalité à 2000 échantillons temporels classés correctement. En fonction de la position des échantillons bien classés, la différence du point de vue de l'ergonomie d'utilisation du système peut être largement améliorée, le filtrage empêchant des instabilités dans la réponse du classifieur.

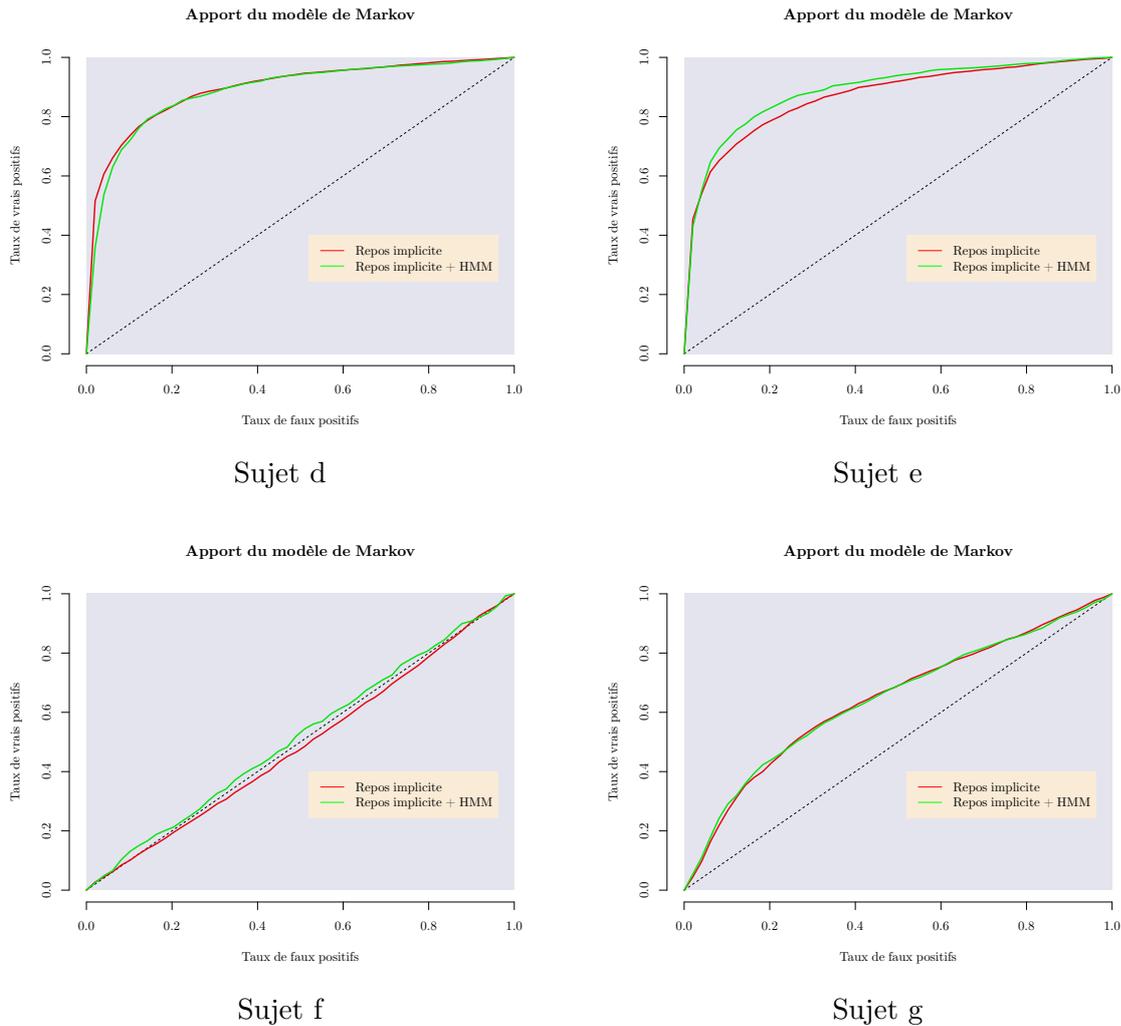


Figure V.9 Courbes ROC pour les quatre derniers sujets. L'apprentissage de l'état de repos est implicite. Les précisions ne sont pas indiquées sur ces figures car elles sont sensiblement similaires entre les deux méthodes.

	a	b	c	d	e	f	g	Moyenne	Moyenne
Brut	0,508	0,292	0,559	0,512	0,083	0,608	0,417	0,456	0,384
VBHMM	0,493	0,299	0,545	0,511	0,075	0,604	0,414	0,452	0,377

Tableau V.4 Proportions de points mal classés dans le cas où le seuillage est basé sur les probabilités brutes en sortie du classifieur ou bien sur les probabilités filtrées. L'apprentissage de l'état de repos est explicite.

7 Discussion et conclusion

Nous avons présenté une méthode très générale pour le filtrage des probabilités. S'inspirant du cadre variationnel et des exemples qui ont été présentés dans Šmídl et Quinn (2005) et Šmídl et Quinn (2008), nous avons intégré ces méthodes dans une procédure de traitement destinée à identifier les états mentaux d'un sujet à partir des signaux EEG. Nous avons montré que cette procédure pouvait permettre des améliorations significatives de performances pour certains sujets et beaucoup plus modestes pour certains autres. Pour l'ensemble des sujets, la mise en place du filtrage variationnel n'a pas détérioré les classifications.

Nous avons également évalué l'influence de la méthode utilisée pour apprendre l'état de repos. L'apprentissage explicite de cet état a permis une amélioration significative des performances pour le sujet de la troisième compétition ainsi que pour le sujet *b* de la quatrième compétition. Cependant elle a conduit à une diminution significative des performances pour le sujet *c* de la quatrième compétition. Ces résultats opposés peuvent s'expliquer par une grande variabilité de l'état de repos, entraînant ainsi une difficulté à apprendre les caractéristiques constantes de cet état. Pour la plupart des sujets, l'équivalence montrée entre ces deux méthodes laisse au concepteur le choix d'adopter une méthode comportant un paramètre ou bien une méthode sans paramètre. Dans le cas de l'apprentissage implicite de l'état de repos, le concepteur peut en effet choisir la valeur de la probabilité d'activité *a priori* afin de se positionner à un endroit précis de la courbe ROC. Ceci peut être d'une grande utilité pour adapter l'inertie du système à l'application visée. Enfin, notons que les précisions de chacune des méthodes présentées ici sont excellentes. Ainsi la difficulté de la conception des systèmes sans contrainte temporelle réside plus dans le bon apprentissage de l'état de repos plutôt que dans la difficulté à distinguer entre deux tâches d'imagerie motrice parmi le flot continu des signaux EEG.

Les charges de calcul engendrées par chacune des méthodes présentées sont très faibles. C'est une des raisons qui nous a fait choisir le cadre variationnel par rapport à d'autres méthodes dans lequel le modèle utilisé dans ce chapitre auraient également pu être traité. Dans les différents cas traités au cours de ce chapitre, toutes les étapes proposées peuvent être appliquées en temps réel. Ceci s'explique par la simplicité des équations de mise à jour des paramètres pour le filtrage variationnel. De plus, le filtrage spatial engendre, dans le fonctionnement en temps en mode évaluation du système, uniquement une opération linéaire appliquée aux données. Ainsi il ne limite pas l'application de la méthode en temps réel.

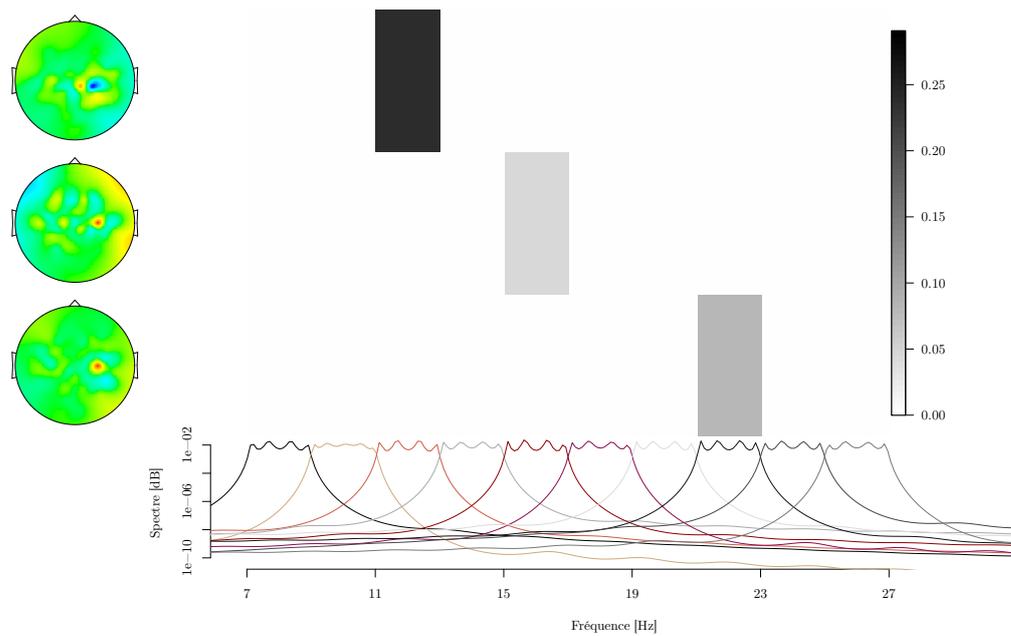
7.1 Utilisation des bandes de fréquences

Il est même possible de simplifier encore plus la méthode en limitant le nombre de bandes de fréquences utilisées. En effet, dans l'ensemble des analyses de ce chapitre, nous avons utilisé un banc de filtres recouvrant la bande entre 7 et 25 Hz, de largeurs respectives de 2 Hz et de bandes passantes disjointes. De plus, le nombre de filtres utilisés pour chaque bande de fréquence était de 8. Nous avons alors laissé au classifieur la tâche de sélectionner les meilleures caractéristiques parmi les 8×10 caractéristiques proposées. Afin de limiter le nombre de filtres dans le banc de filtres initial, nous pouvons observer *a posteriori* les caractéristiques utilisées afin de réduire ce nombre et ainsi la complexité de la méthode. Notons cependant que cette sélection réduit essentiellement la complexité de la méthode pendant l'apprentissage des paramètres mais relativement peu pendant le traitement en ligne. Le gain lors du traitement en ligne réside dans le nombre de multiplications à effectuer pour trouver les sources et également les bandes de fréquences pour lesquelles les caractéristiques doivent être évaluées.

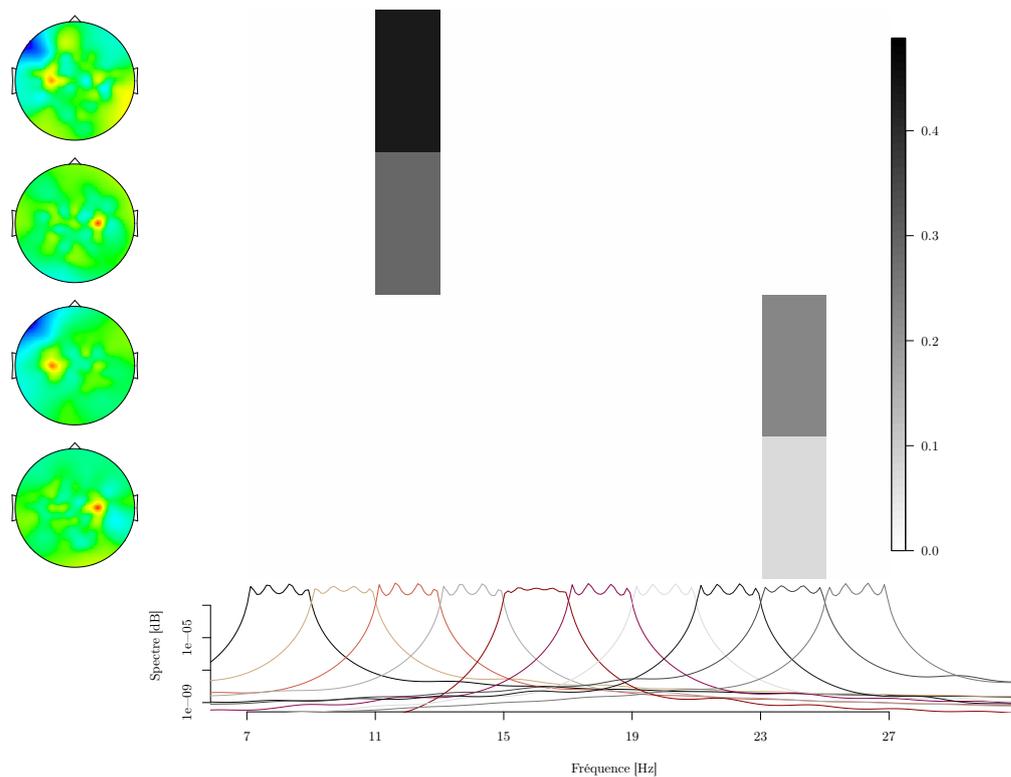
La [figure V.10](#) montre un exemple caractéristique de l'utilisation des bandes de fréquence. On s'aperçoit que seules quelques unes d'entre elles sont réellement utilisées et qu'on pourrait simplifier la procédure en considérant uniquement les bandes intéressantes.

7.2 Améliorations dues au filtrage variationnel

Les améliorations apportées par l'utilisation du filtrage variationnel des probabilités peuvent apparaître relativement modestes. Néanmoins, il faut noter que de petites améliorations peuvent, dans un système réel, s'avérer importantes pour le contrôle. En effet, le filtrage contribue à stabiliser la sortie du classifieur et empêche donc les sauts d'états



sujet a



sujet e

Figure V.10 Coefficients non nuls de la régression logistique. Les coefficients (rectangles gris dont la valeur est codée selon la valeur absolue du coefficient) sont positionnés sur une grille. En ordonnée sont représentées les sources (poids au niveau des capteurs EEG) correspondant aux coefficients non nuls. En abscisse est indiquée la fréquence à laquelle les coefficients se rapportent. Les bandes de fréquence sont matérialisées par les réponses impulsionnelles des filtres de Butterworth utilisés pour filtrer les signaux afin d'obtenir les caractéristiques. Les données sont relatives au sujet *a* (haut) et *e* (bas) de la compétition 2008 (pas de classe repos apprise). Dans les deux cas, les tâches sont les imaginations de la main gauche et de la main droite.

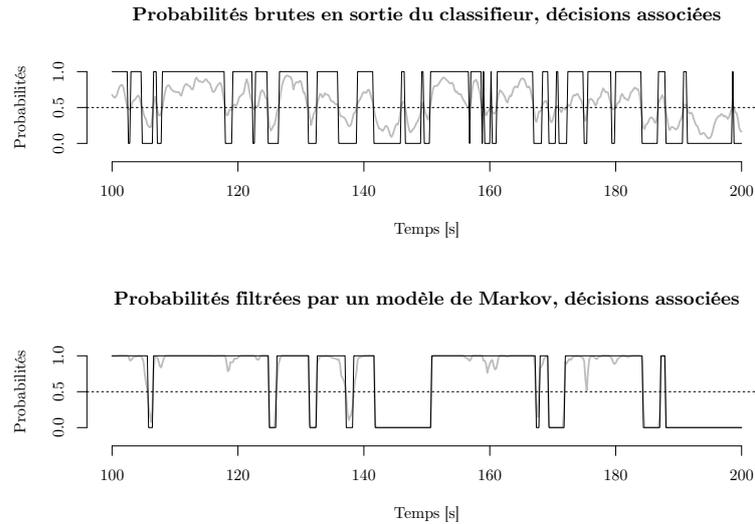


Figure V.11 Probabilités ainsi que les décisions associées en sortie du classifieur (haut) ainsi qu’après filtrage variationnel. La discrimination implique seulement deux classes. Par conséquent nous n’avons représenté que la probabilité pour chaque segment d’appartenir à la classe 1. Ainsi, lorsque la probabilité est proche de 1, il y a une grande probabilité pour que le sujet réalise la tâche 1. À l’opposé, lorsque la probabilité est proche de 0, il y a une grande probabilité pour que le sujet réalise la tâche 2. Le seuil de décision entre les deux classes est matérialisé par un trait en pointillés en 0,5.

inopinés qui peuvent nuire au bon fonctionnement. Ainsi, même si quelques points temporels seulement sont régularisés à l’aide de cette méthode, les stabilisations apportées peuvent ne pas être négligeables.

8 Références du chapitre

- Šmídl, V. et Quinn, A. (2008). Variational Bayesian filtering. *IEEE Trans. Signal Process.*, 56(10):5020–5030.
- Šmídl, V. et Quinn, A. (2005). *The Variational Bayes Method in Signal Processing (Signals and Communication Technology)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.
- Bashashati, A., Ward, R. K. et Birch, G. E. (2007b). Towards development of a 3-state self-paced brain-computer interface. *Comput. Intell. Neurosci.*, 2007(1):84386.
- Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philos. Trans. R. Soc.*, 53:370–418.
- Birbaumer, N. et Cohen, L. G. (2007). Brain-computer-interfaces (BCI): Communication and restoration of movement in paralysis. *J. Physiol.*, 579(3):621–636.
- Birch, G. E., Mason, S. G. et Borisoff, J. F. (2003). Current trends in brain-computer interface research at the neil squire foundation. *IEEE Trans. Neural Syst. Rehabil. Eng.*, 11(2):123–126.
- Blankertz, B., Dornhege, G., Krauledat, M., Müller, K.-R. et Curio, G. (2007). The non-invasive berlin brain-computer interface: fast acquisition of effective performance in untrained subjects. *Neuroimage*, 37(2):539–550.
- Chiappa, S. (2006). *Analysis and Classification of EEG Signals using Probabilistic Models for Brain Computer Interfaces*. Thèse de doctorat, École Polytechnique Fédérale de Lausanne.

- Chiappa, S. et Bengio, S. (2004). HMM and IOHMM modeling of EEG rhythms for asynchronous BCI systems. Dans *Proceedings of the 12th European Symposium on Artificial Neural Networks ESANN*.
- Choudrey, R. A. (2002). *Variational Methods for Bayesian Independent Component Analysis*. Thèse de doctorat, University of Oxford.
- Dornhege, G., Blankertz, B., Curio, G. et Müller, K.-R. (2004). Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms. *IEEE Trans. Biomed. Eng.*, 51(6):993–1002.
- Galán, F., Nuttin, M., Lew, E., Ferrez, P. W. et Vanacker, G. et al. (2008). A brain-actuated wheelchair: Asynchronous and non-invasive brain-computer interfaces for continuous control of robots. *Clin. Neurophysiol.*, 119(9):2159–2169.
- Ghahramani, Z. et Beal, M. (2001). Propagation algorithms for variational Bayesian learning. *Adv. Neur. Inf. Process. Syst.*, 13:507–513.
- Jurcak, V., Tsuzuki, D. et Dan, I. (2007). 10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems. *Neuroimage*, 34(4):1600–1611.
- MacKay, D. J. C. (1997). Ensemble learning for hidden Markov models. Rapport technique, University of Cambridge, Cambridge, U.K..
- Mason, S., Kronegg, J., Huggins, J., Fatourech, M. et Schlögl, A. (2006). Evaluating the performance of self-paced brain computer interface technology. Rapport technique, Neil Squire Society, Vancouver, Canada.
- Mason, S. G. et Birch, G. E. (2000). A brain-controlled switch for asynchronous control applications. *IEEE Trans. Biomed. Eng.*, 47(10):1297–1307.
- Obermaier, B., Neuper, C., Guger, C. et Pfurtscheller, G. (2001). Information transfer rate in a five-classes brain-computer interface. *IEEE Trans. Neural Syst. Rehabil. Eng.*, 9(3):283–288.
- Rabiner, L. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proc. IEEE*, 77(2):257–286.
- Scherer, R., Lee, F., Schlogl, A., Leeb, R. et Bischof, H. et al. (2008). Toward self-paced brain-computer communication: navigation through virtual worlds. *IEEE Trans. Biomed. Eng.*, 55(2):675–682.
- Sykacek, P., Roberts, S., Stokes, M., Curran, E. et Gibbs, M. et al. (2003). Probabilistic methods in BCI research. *IEEE Trans. Neural Syst. Rehabil. Eng.*, 11(2):192–195.
- Sykacek, P., Roberts, S. J. et Stokes, M. (2004). Adaptive BCI based on variational bayesian Kalman filtering: an empirical evaluation. *IEEE Trans. Biomed. Eng.*, 51(5):719–727.
- Townsend, G., Graimann, B. et Pfurtscheller, G. (2004). Continuous EEG classification during motor imagery—simulation of an asynchronous BCI. *IEEE Trans. Neural Syst. Rehabil. Eng.*, 12(2):258–265.
- Tsui, C., Gan, J. et Roberts, S. (2009). A self-paced brain-computer interface for controlling a robot simulator: an online event labelling paradigm and an extended Kalman filter based algorithm for online training. *Med. Biol. Eng. Comput.*, 47(3):257–265.

Chapitre VI

VERS LA COADAPTATION CERVEAU-MACHINES



1	Introduction	135
2	Quantifier l'apprentissage du sujet	136
2.1	Quantifier la réalisation de chaque essai	137
2.2	Convergence de l'apprentissage des filtres spatiaux	140
2.3	Application aux interfaces asynchrones	141
3	Classification continue adaptative	143
3.1	Modéliser les probabilités en sortie du classifieur	143
3.2	Application aux interfaces asynchrones	145
4	Discussion et conclusion	146
5	Références du chapitre	148

LES INTERFACES CERVEAU-MACHINES (ICMs) impliquent une interaction entre le sujet humain et le système contrôlé. Or les systèmes actuels font la plupart du temps l'hypothèse relativement forte qui consiste à considérer que le sujet réalise les tâches demandées d'une manière constante (au sens où les zones cérébrales et les fréquences impliquées ne varient pas dans le temps). Ce chapitre propose des pistes nouvelles en vue d'une coadaptation entre le système et le sujet. Nous proposons dans un premier temps d'utiliser les informations fournies par la méthode des [chapitres III](#) et [V](#) afin de quantifier le degré d'apprentissage du sujet (comment est-il capable de maintenir la tâche qui lui est demandée ?). Nous nous intéressons dans un deuxième temps à contrôler le réapprentissage automatique du classifieur afin de prendre en compte les possibles changements et dérives dans la manière avec laquelle le sujet réalise les tâches mentales qui lui sont demandées.

1 Introduction

Lors de l'interaction entre les sujets et les systèmes ICMs, le sujet apprend à maîtriser son activité cérébrale afin de contrôler un système électronique. Or dans la plupart des systèmes, le sujet réalise quelques essais d'entraînement afin de fixer les paramètres de la méthode. Ces paramètres sont ensuite utilisés lors de la session d'évaluation (le sujet se sert effectivement du système). Or cette approche soulève deux problèmes : tout d'abord on suppose que le sujet réalise les tâches de manière parfaite lors de la session d'entraînement afin que les paramètres soient appris correctement ; de plus, cette approche suppose que les différences de réalisation des tâches par le sujet entre la session d'entraînement et d'évaluation soient minimales. Certaines approches robustes ont été envisagées afin de réduire l'effet des données aberrantes de la session d'entraînement et ainsi lutter contre le biais introduit par le premier problème ([Wang et Zheng, 2008](#) et [Blankertz et al., 2008](#)).

Dans le même but, d'autres équipes ont utilisé la détection de signaux EEG caractéristiques générés lors d'un mauvais retour envoyé au sujet suite à un essai (Buttfield et al., 2006) (le retour n'est pas en accord avec la tâche réalisée). Cependant, à notre connaissance, aucune proposition n'a été faite afin de détecter en temps réel la réalisation d'un mauvais essai par le sujet afin de le retirer de la base d'apprentissage. Or cet aspect est d'une importance primordiale dans les systèmes réels, pour lesquels il est impossible de vérifier que le sujet a bien réalisé la tâche qui lui a été demandée.

Le second problème (inconstance de la réalisation des tâches par le sujet) a été soulevé par plusieurs auteurs (Sykacek et al., 2004, Shenoy et al., 2006, Vidaurre et al., 2006 et Krauledat, 2008). Les raisons avancées pour expliquer cette inconstance varient selon les auteurs. Dans Sykacek et al. (2004), les auteurs avancent des raisons techniques (par exemple des changements d'impédance entre les électrodes et le scalp) mais également dues au sujet (effet d'apprentissage) qui rendent le classifieur inadapté. Dans Shenoy et al. (2006) et Krauledat (2008), une analyse très complète des caractéristiques de cette inconstance est réalisée. Les auteurs observent une dérive des distributions en sortie du classifieur qu'ils expliquent par une modification de l'activité générale du cerveau entre les sessions d'entraînement et de test. Enfin dans Vidaurre et al. (2006), l'influence du retour sensoriel est avancée, entraînant une adaptation du sujet et une modification de la manière de réaliser la tâche. Afin de lutter contre ces effets, une adaptation du classifieur est nécessaire. Celle-ci est réalisée en modifiant à la main les paramètres de la classification, en réentraînant un autre classifieur ou encore en modifiant la valeur de certains seuils (Pfurtscheller et al., 2000 et Wolpaw et al., 2000). Dans Shenoy et al. (2006) et Krauledat (2008), une approche automatique visant à adapter l'espace de représentation des caractéristiques est proposée. Enfin, dans Sykacek et al. (2004), les auteurs ont utilisé une formulation variationnelle d'un filtre de Kalman afin de mettre à jour le système de classification.

Nous proposons dans ce chapitre un moyen original de pallier ces deux problèmes. Tout d'abord, nous utilisons les modèles des chapitres III et V afin de quantifier la qualité d'un essai pour ainsi décider s'il doit ou non être inclus dans la base d'apprentissage (section 2). Dans un deuxième temps (section 3), nous proposons une approche simple afin de quantifier la dérive du classifieur et évaluer la qualité de la tâche réalisée sur chacune des fenêtres analysées. Ces deux mesures nous permettent de décider quand il est nécessaire de réapprendre le classifieur et également de tenir à jour une base de données d'essais dont la classe a été reconnue de manière univoque. Celle-ci permettra alors le réapprentissage du classifieur avec des données récentes prenant en compte les possibles inconstances de réalisation du sujet.

2 Quantifier l'apprentissage du sujet

Le sujet réalise-t-il bien la tâche qui lui est demandée ? Cette question récurrente dans la conception des ICMs a une importance considérable lors de la phase d'entraînement. En effet, si des essais mal réalisés sont inclus dans la base de données d'apprentissage, alors la qualité du système de discrimination en pâtira aussitôt. Or la réalisation d'un mauvais essai peut provenir de deux raisons distinctes : tout d'abord il se peut que le sujet se trompe dans la tâche à réaliser ; mais il est également possible que le sujet montre des signes de fatigue, conduisant à une réalisation approximative des tâches. Afin de quantifier ces phénomènes, nous proposons d'observer comment le sujet parvient à maintenir une tâche pendant un certain intervalle en utilisant le modèle de Markov du chapitre V et d'étudier la convergence des filtres spatiaux du chapitre III appris séquentiellement à l'aide des données d'entraînement. Ces mesures peuvent également être utilisées pour évaluer la quantité de données nécessaire au bon entraînement du système de discrimination. A notre

connaissance, aucune étude similaire n'a été entreprise dans ce sens pour la conception des ICMs.

2.1 Quantifier la réalisation de chaque essai

Nous nous plaçons dans le cadre de la phase d'apprentissage de la méthode décrite par la [figure III.5](#). Rappelons que cette méthode, à partir d'un ensemble de données d'apprentissage, fournit les meilleurs filtres spatiaux afin de discriminer entre différentes tâches ainsi qu'une application \mathcal{A} qui, à chaque position t par rapport au début de l'essai, fournit une probabilité pour que le sujet soit en train d'imaginer l'une des tâches qui lui a été demandée. Dans un premier temps, nous allons utiliser cette séquence de probabilités afin de quantifier la qualité de réalisation d'un essai par le sujet. Cette évaluation va bien entendu être réalisée de manière causale, c'est-à-dire que l'évaluation de la qualité de l'essai i n'utilise que les essais antérieurs $[1..i-1]$. Dans cette partie, nous supposerons que le paradigme n'implique que deux tâches mentales. La généralisation à plus de deux tâches mentales est immédiate. De plus, le classifieur ne sera entraîné qu'à l'aide d'une seule fenêtre temporelle de chaque essai (choisie au milieu de l'essai). $\mathcal{A}^{(i-1)}$ est donc constante sur tout l'intervalle de réalisation de l'essai.

Pour chaque nouvel essai i à notre disposition, nous allons utiliser les filtres spatiaux $W^{(i-1)}$ ainsi que le classifieur $\mathcal{A}^{(i-1)}$ appris à l'aide des données relatives aux essais antérieurs afin de classifier l'essai courant. Nous obtenons alors une séquence de probabilités (autant que de fenêtres considérées dans l'essai) à partir des caractéristiques $\chi_t^{(i)}$ de l'essai i (celles-ci sont calculées à partir des sources obtenues par projection selon $W^{(i-1)}$)

$$\mathcal{A}^{(i-1)} \left(\chi_t^{(i)} \right) = \begin{pmatrix} p_{1,t}^{(i)} \\ p_{2,t}^{(i)} \end{pmatrix} \quad (\text{VI.1})$$

où $p_{1,t}^{(i)}$ (respectivement $p_{2,t}^{(i)}$) représente la probabilité pour que le sujet soit dans l'état mental appelé 1 (respectivement 2).

Cette suite de probabilités pourrait alors être utilisée pour observer si l'essai appartient bien à la classe demandée au sujet. Cependant, ceci impliquerait d'observer l'ensemble de la suite des probabilités afin de déterminer, selon un critère à définir, si l'essai est bien classé. La quantité d'information est trop grande pour que cette suite de probabilités soit traitée de manière automatique. De plus nous voulons une mesure simple qui rende compte d'un effet de fatigue du sujet. Afin de satisfaire ces différents critères, nous proposons d'utiliser la méthode du filtrage variationnel présentée dans le [chapitre V](#). L'idée est de résumer la suite de probabilités en une unique valeur qui la caractérise. Pour ce faire, nous allons filtrer la suite de probabilités par la méthode variationnelle en partant d'un état initial non-informatif

$$\alpha_0 = \begin{pmatrix} 0,5 \\ 0,5 \end{pmatrix} \quad \beta_0 = \begin{pmatrix} 0,5 \\ 0,5 \end{pmatrix} \quad T_0 = \begin{pmatrix} 0,5 & 0,5 \\ 0,5 & 0,5 \end{pmatrix}. \quad (\text{VI.2})$$

Rappelons que la méthode du chapitre précédent infère, à partir d'une séquence de probabilités, la variable latente qui l'a générée (« forward » l_t et « forward-backward » l_{t-1}). Les variables l_t et l_{t-1} suivent une distribution multinomiale de paramètres respectifs α_t et β_t . Les paramètres α_0 et β_0 décrivent les distributions *a priori* de l_t et l_{t-1} . De plus, la matrice de transition T_t est apprise au cours du filtrage de la séquence. Celle-ci suit une distribution de Dirichlet de paramètre initial T_0 .

On obtient en sortie la suite des probabilités filtrées mais ce n'est cette fois pas cette suite qui nous intéresse. Elle ne permet pas de résumer l'information contenue dans la suite de manière satisfaisante. Nous nous concentrons plutôt sur la manière dont la suite des probabilités observées a modifié la matrice de transition estimée. Les paramètres ρ et

κ sont fixés respectivement à 2 et 10. Une faible valeur de κ est justifiée par le fait que nous voulons observer des modifications à l'aide d'une suite assez courte d'observations (typiquement une dizaine). Il ne faut cependant pas prendre κ trop faible au risque que la dernière valeur de la matrice de transition soit uniquement influencée par les dernières valeurs de la suite des probabilités et non par son ensemble. Cette méthode permet de définir le score

$$\zeta_T = T_f - T_0 \quad \zeta_T \in \mathbb{R}^{2 \times 2}, \quad (\text{VI.3})$$

où T_f est la matrice de transition estimée finale, après filtrage de la suite des probabilités observées (voir l'équation V.30). Dans la suite, nous noterons $[\zeta_T]_{i,j}$ l'élément de la ligne i et de la colonne j de la matrice score ζ_T . Cette mesure s'interprète alors de la manière suivante : plus $[\zeta_T]_{1,1}$ est proche de 0,5, plus la suite de probabilités contient des transitions de l'état 1 vers l'état 1 franches, par conséquent un grand nombre de probabilités $p_{1,t}^{(i)}$ grandes et de probabilités $p_{2,t}^{(i)}$ petites ; on peut interpréter de la même manière $[\zeta_T]_{2,2}$ en inversant le rôle des probabilités $p_{1,t}^{(i)}$ et $p_{2,t}^{(i)}$. L'interprétation des termes hors-diagonaux est moins évidente, ceux-ci permettent de quantifier la quantité de transitions observées dans la suite entre l'état 1 et l'état 2 (pour $[\zeta_T]_{2,1}$) et entre l'état 2 et l'état 1 (pour $[\zeta_T]_{1,2}$).

Ainsi, lorsqu'on demande au sujet de réaliser la tâche 1, on s'attend, si la tâche est bien réalisée, à ce que $[\zeta_T]_{1,2} + [\zeta_T]_{1,1}$ soit grand tandis que $[\zeta_T]_{2,1} + [\zeta_T]_{2,2}$ soit petit. Inversement, on s'attend à la situation opposée dans le cas où la tâche 2 est demandée. D'où finalement la mesure permettant de quantifier la qualité de réalisation de l'essai

$$\zeta = \left([\zeta_T]_{1,2} + [\zeta_T]_{1,1}\right) - \left([\zeta_T]_{2,1} + [\zeta_T]_{2,2}\right). \quad (\text{VI.4})$$

Plus $\zeta > 0$ est éloigné de zéro, moins l'essai présente d'ambiguïté et peut facilement être classé comme appartenant à la tâche 1. Lorsque $\zeta = 0$, on peut dire que la suite des probabilités ne présente pas de prédominance pour l'un ou l'autre des états. Enfin, plus $\zeta < 0$ est éloigné de zéro, plus il est facile de le classer comme appartenant à la classe 2. Notons enfin que, étant donnés les paramètres ρ et κ que nous avons choisis, notre mesure est bornée par $|\zeta| < 0,7$. Cette valeur a été trouvée expérimentalement en filtrant une suite de onze probabilités univoque, *i.e.* $\forall t \quad p_{1,t}^{(i)} = 1$ et $p_{2,t}^{(i)} = 0$; le nombre onze correspond au nombre de fenêtres que nous avons utilisé dans la suite, un nombre différent de fenêtres changerait la valeur de cette limite supérieure.

Afin d'illustrer le fonctionnement de cette mesure, les données de la quatrième compétition ont été utilisées (voir chapitre V). Pour chacun des sujets, nous n'avons considéré que les données d'entraînement. La première paire d'essais, constituée d'un essai de chaque classe, est utilisée pour initialiser le calcul des filtres spatiaux ainsi que les paramètres du classifieur. Pour chacun des 198 essais suivants, les filtres spatiaux ainsi que le classifieur entraînés avec les données antérieures servent à obtenir la suite des probabilités (classification de onze fenêtres de 1,5s successives se recouvrant à 80% décrivant l'intervalle théorique de réalisation de la tâche). Les onze paires de probabilités observées ($p_{1,t}^{(i)}$ et $p_{2,t}^{(i)}$) sont alors filtrées par la méthode variationnelle afin de donner un unique score ζ (défini par l'équation VI.4) pour l'essai, celui-ci reflétant la qualité de la réalisation de la tâche.

Les résultats de quatre sujets sont représentés sur la figure VI.1. Ils sont représentatifs de l'ensemble des sujets. Cette figure permet d'ores et déjà de faire plusieurs observations. Tout d'abord, on voit une nette différence entre le sujet artificiel d et les sujets réels a , f et g . Une excellente discrimination est obtenue pour le sujet artificiel à partir de la dixième paire d'essais et seuls deux essais semblent mal réalisés (peut-être des pièges introduits volontairement par les organisateurs de la compétition). Notons qu'une telle caractéristique n'est pas commune à l'ensemble des sujets artificiels et que l'un d'eux, le sujet c (non présenté ici), obtient une caractéristique qui ne peut pas être distinguée de celles des sujets réels. Le sujet g obtient des performances exceptionnellement bonnes,

on voit qu'une excellente discrimination est observée après seulement 5 paires d'essais. Celui-ci semble seulement avoir un léger « passage à vide » pour réaliser la tâche « main droite » entre les essais 35 et 53. Les courbes présentées pour les deux autres sujets réels (*a* et *f*) sont moins claires, néanmoins il semble encore une fois que le nombre d'essais nécessaires pour obtenir de bons scores soit très faible. Cependant ceux-ci semblent avoir réalisé un ensemble non négligeable d'essais médiocres. Notons enfin que le sujet *f* semble montrer des signes de fatigue lors de la réalisation de ces derniers essais. De meilleurs scores de classification auraient donc certainement pu être obtenus en arrêtant la session d'entraînement plus tôt et en fixant les paramètres avec uniquement les premiers essais.

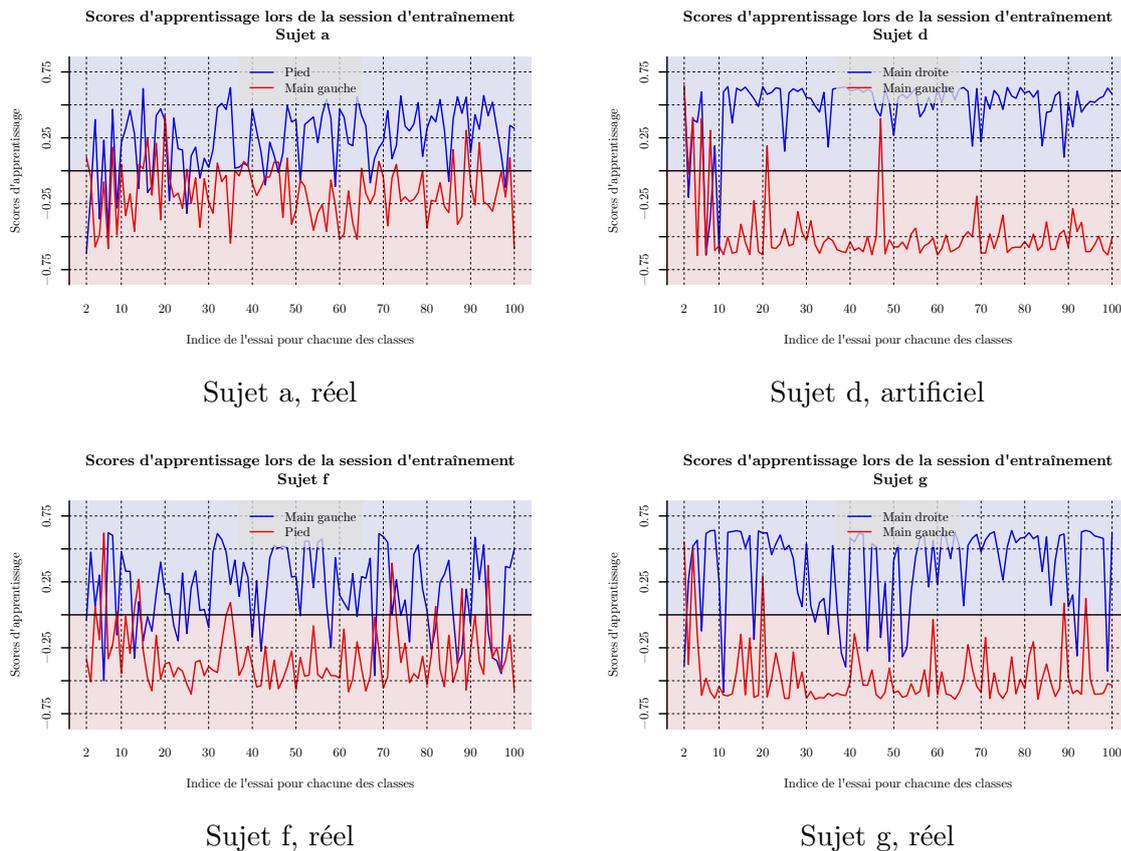


Figure VI.1 Scores d'apprentissage pour l'ensemble des essais de la session d'entraînement relatifs à quatre sujets (parmi sept). Nous rappelons que les sujets *a*, *f* et *g* sont des sujets réels alors que les données du sujet *d* ont été générées artificiellement. Sur chacun des quatre graphiques sont représentés les scores d'apprentissage de chaque essai. Les essais relatifs à chacune des classes sont représentés en utilisant une couleur différente (bleu pour la première tâche et rouge pour la seconde). La ligne horizontale matérialisée pour un score de zéro marque le seuil de décision pour déterminer si l'essai a été correctement appris (scores positifs pour la première tâche et négatifs pour la seconde). Ainsi, plus le score est proche de zéro, plus la réalisation de l'essai par le sujet a été ambiguë. A l'opposé, si la valeur absolue de score est importante, cela signifie que l'essai peut être discriminé sans difficulté en utilisant les essais précédents. Notons enfin que pour évaluer la qualité de l'essai i , seules les données antérieures $[1 \dots i - 1]$ ont été utilisées.

Ces observations permettent de proposer une méthode simple pour apprendre uniquement les essais bien réalisés. En effet, il semble que le nombre d'essais soit largement suffisant mais que certains d'entre eux sont de mauvaise qualité. On peut donc envisager de les enlever de la base d'apprentissage afin d'obtenir des filtres et un classifieur de meilleure qualité.

2.2 Convergence de l'apprentissage des filtres spatiaux

On peut également mieux contrôler la session d'apprentissage en observant la convergence des filtres spatiaux trouvés par diagonalisation conjointe au fur et à mesure que le nombre d'essais de la phase d'apprentissage augmente. Afin d'évaluer cette convergence, on calcule successivement les filtres spatiaux relatifs à la paire d'essai i (un pour chaque classe) en utilisant les données des essais $[1..i]$. Une difficulté consiste alors à mettre en relation les filtres correspondant à l'essai i et ceux correspondant à l'essai $i-1$. Cette étape est réalisée de manière à maximiser la corrélation totale entre chacune des paires. Une fois cette correspondance établie, on calcule le logarithme de la norme ℓ_2 de la différence entre les deux filtres des deux étapes successives.

La [figure VI.2](#) présente les résultats obtenus pour les huit meilleurs filtres appris successivement. Sur fond rosé, on a indiqué les filtres qui sont effectivement utilisés par la régression logistique à l'étape i afin de réaliser la classification (coefficients non nuls de la régression logistique). Les résultats présentés correspondent à un sujet artificiel d et un sujet réel f . Bien que l'ensemble des courbes ne puissent être présentées ici, et contrairement à la mesure présentées précédemment, ce critère permet bien de discriminer entre les sujets artificiels (seulement deux sources dont le critère de convergence décroît sont utilisées de manière constante lors de l'ajout de nouveaux essais) et les sujets réels (la convergence des sources est plus erratique et les sources utilisées varient lorsque les données s'accumulent). La remarque du paragraphe précédent concernant le sujet f qui semblait montrer des signes de fatigue à partir de la mesure ζ se confirme puisqu'un changement de la source utilisée survient à un état avancé de la session. Nous avons présenté uniquement ces deux sujets car il s'avérait en réalité relativement difficile de distinguer lequel était artificiel et lequel était réel en regardant uniquement leurs performances. Lors de la compétition, les organisateurs n'ont pas indiqué quels étaient les sujets réels et les sujets artificiels. Ils ont demandé aux compétiteurs d'évaluer ce point pour chacun des sujets. Or peu de réponses parfaites ont été données. En utilisant la convergence des filtres spatiaux et l'utilisation des sources par le classifieur, il est cependant relativement facile de décider entre ces deux types de sujets. Le critère présenté ici peut être interprété comme la dimension réelle des données (combien de sources sont nécessaires pour générer l'activité cérébrale des tâches demandées ?) car il permet de visualiser combien de sources (trouvées par diagonalisation conjointe) sont utilisées de manière consistante dans le temps. Ainsi, il est clair que les données du sujet d lors de la réalisation de la tâche mentale est de deux tout au long de la période d'apprentissage. Alors que cette dimension a été clairement de deux pour l'ensemble des sujets artificiels, elle était beaucoup plus instable pour les sujets réels et variaient entre 2 et 4. Notons néanmoins que le nombre de sources sélectionnées dépend de la valeur du paramètre de régularisation du classifieur. Dans le cadre de ce chapitre, une valeur constante égale à 0,2 a été choisie (en accord avec les observations du chapitre précédent, voir le [tableau V.2](#)).

Cette remarque concernant le nombre de sources cérébrales responsables de l'activité EEG ne permet cependant pas de limiter la dimension de W à cette dimension. En effet, les sources intéressantes ne sont pas forcément les mieux classées si l'on utilise le critère basé sur l'information mutuelle décrit dans [Grosse-Wentrup et Buss \(2008\)](#). Nous pensons donc que ce critère n'est pas forcément toujours très efficace pour classer les filtres spatiaux et qu'il vaut mieux surestimer le nombre de filtres nécessaires afin de ne pas omettre certains filtres intéressants. Notons enfin que l'observation de la convergence des filtres permet également, mais de manière moins claire que le critère précédent, de détecter des signes de fatigue du sujet.

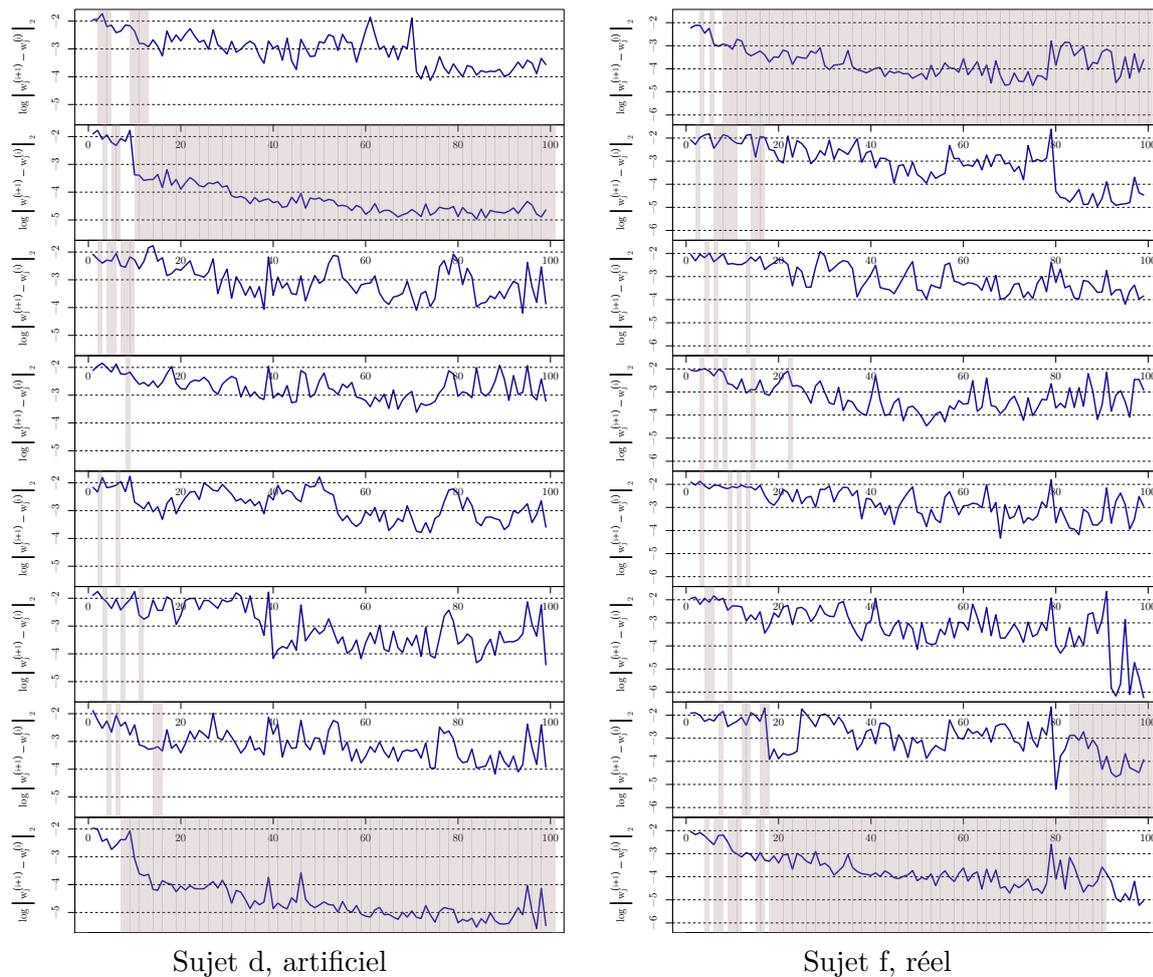


Figure VI.2 Étude de la convergence de l'apprentissage des filtres lors de la session d'entraînement en fonction du nombre d'essais de chacune des tâches réalisées par les sujets. Ces résultats sont relatifs aux sujets d et f. La convergence est évaluée comme le logarithme de la norme ℓ_2 entre les huit meilleurs filtres successifs trouvés par diagonalisation conjointe approchée. Afin de calculer les filtres pour l'essai i , seules les données relatives aux essais $[1 \dots i]$ sont utilisées. Dans le but d'éliminer les permutations dues à l'algorithme de diagonalisation, les filtres sont identifiés deux à deux avec ceux de l'étape précédente afin de maximiser la corrélation totale entre les filtres des deux étapes successives. Enfin, sur fond gris sont indiquées les sources réellement utilisées par le classifieur (dans au moins une des bandes de fréquence).

2.3 Application aux interfaces asynchrones

Comme suggérée précédemment, une application simple du score de qualité de chaque essai consiste à retirer de la base de données d'apprentissage les réalisations qui sont diagnostiquées comme mauvaises. Afin d'évaluer l'intérêt d'une telle procédure, nous considérons à nouveau les données relatives à la quatrième compétition BCI (voir [chapitre V](#) pour une description). Nous appliquons la méthode résumée sur la [figure V.4](#) sans utiliser la procédure de filtrage des probabilités par le modèle de Markov caché. Lors de la phase d'entraînement, deux déclinaisons de la méthode sont considérées :

- tous les essais, sans considération de qualité de réalisation, sont utilisés pour entraîner les filtres spatiaux et le classifieur ;
- les essais les moins bien réalisés, selon le critère défini par l'équation VI.4 sont retirés de la base de données d'apprentissage.

Les paramètres sont exactement les mêmes dans les deux cas et sont fixés selon le [tableau V.2](#). Les filtres spatiaux ainsi que le classifieur appris lors de la phase d'entraînement sont utilisés afin de classifier en continu l'EEG de la phase de test. L'état de repos n'est pas considéré, par conséquent le problème de classification est binaire puisqu'il se réduit au problème de décision entre deux classes distinctes. Afin de nous placer dans les conditions réelles d'une expérience, nous avons maintenu l'exigence de causalité de la procédure et avons décidé d'autoriser la possibilité d'enlever des essais mal classés seulement à partir de la quarantième réalisation (sur 200). Ce choix s'explique par la quantité insuffisante de données pour évaluer les scores des premiers essais. La [figure VI.3](#) présente les résultats relatifs aux deux déclinaisons précédentes pour les sujets *f* et *g*. Notons que la procédure n'a pas été appliquée aux sujets *d* et *e* dont le nombre d'essais mal réalisés est négligeable (un pour chaque sujet).

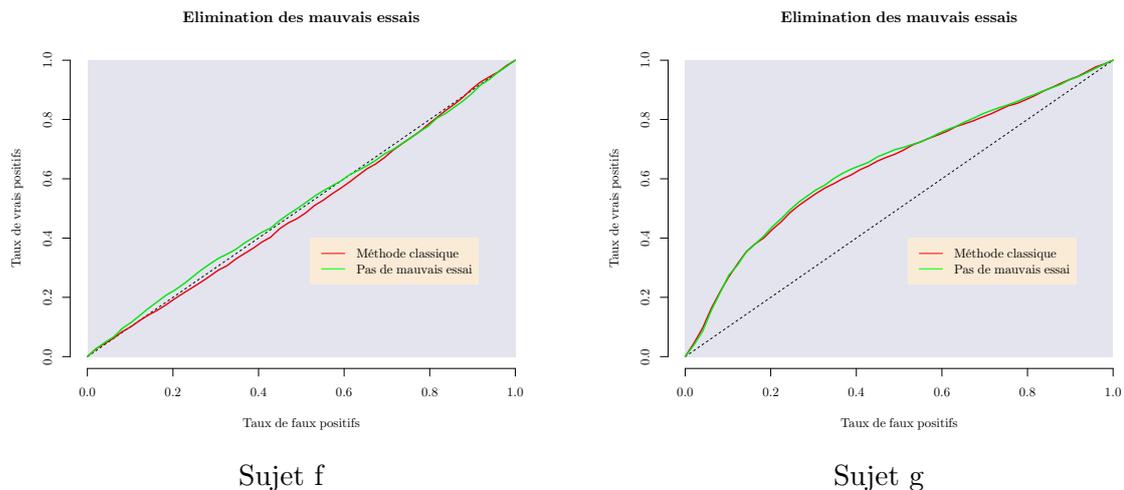


Figure VI.3 Influence de l'élimination des mauvais essais sur les performances. Seuls les résultats des sujets *f* et *g* sont présentés ici. Ces résultats sont présentés sous la forme de courbes ROC et rendent compte de manière synthétique des mesures présentées dans le chapitre précédent. Rappelons que les mesures utilisées permettent de quantifier la qualité de discrimination du système entre un état de repos et un état d'activité. Les courbes sont obtenues en faisant varier entre zéro et un la probabilité *a priori* pour que le sujet soit dans un état de repos. Pour chacun des sujets, en vert est représentée la courbe ROC relative à la méthode proposée dans ce chapitre et visant à retirer de l'ensemble d'entraînement les essais mal réalisés. La courbe ROC correspondant à la méthode consistant à utiliser l'ensemble des essais de la base d'apprentissage sans distinction de qualité est quant à elle représentée en rouge.

Les résultats des deux déclinaisons pour les sujets *a*, *b* et *c* sont équivalents. On aurait bien sûr attendu une amélioration des résultats avec la procédure visant à enlever les essais mal réalisés mais le fait que l'on n'observe pas de baisse de performances en enlevant plusieurs essais de la base d'apprentissage (respectivement 9, 17 et 7 essais pour les deux classes de ces trois sujets) est déjà un point positif important. Ainsi, cette observation confirme le fait que le nombre d'essais d'apprentissage peut dans la plupart des cas être largement réduit sans que les performances des systèmes s'en trouvent détériorées. Comme le montre la [figure VI.3](#), on observe pour les sujets *g* et *f* une légère amélioration des performances pour une large gamme de taux de faux positifs. Si l'amélioration semble très réduite sur les courbes, il faut prendre néanmoins en considération que ces courbes sont réalisées en interpolant la réponse sur chacune des fenêtres analysées au taux d'échantillonnage des données. Or dans ce cas, le taux d'échantillonnage est de 1 KHz et le nombre de points considérés dans les sessions d'évaluation est de plus de 2 millions. Par conséquent une

augmentation de 0,01 du taux de vrais positifs correspond en réalité à plusieurs milliers d'échantillons temporels.

Ces résultats valident donc l'approche visant à éliminer les essais mal réalisés de la base d'apprentissage. Nous avons proposé une mesure originale qui semble relativement adaptée à l'évaluation de la qualité de chacun des essais. D'une manière plus générale, nous pensons que les mesures visant à contrôler la qualité des données d'apprentissage ou encore la convergence des paramètres appris doivent contribuer à simplifier la conception d'ICMs en adaptant la longueur de la session d'entraînement aux performances du sujet.

3 Classification continue adaptative

Après avoir vu comment adapter la période d'entraînement en fonction des réalisations des essais par le sujet, nous allons maintenant aborder le deuxième problème soulevé dans l'introduction : la dérive de la qualité de classification au fur et à mesure de la session d'évaluation. En raison de la qualité et du très faible coût d'apprentissage de la régression logistique que nous utilisons, rien n'empêche de réapprendre en temps réel les paramètres du classifieur. Le problème de rendre la classification adaptative et robuste aux changements dans la réalisation des tâches par le sujet ou encore aux changements de l'activité de fond du cerveau se réduit donc à deux problèmes distincts :

- décider à quel moment on doit réapprendre le classifieur ;
- choisir les données qui vont être utilisées pour le réapprentissage afin que les non-stationnarités qui sont apparues soient prises en compte et compensées.

Ces deux points seront successivement abordés dans la suite de ce chapitre.

Remarque 3 (RÉGRESSION LINÉAIRE ADAPTATIVE). *Lorsque les classifieurs utilisés sont plus simples (la fonction de coût de l'apprentissage ne fait pas intervenir de terme de régularisation), alors il est possible de considérer le problème du réapprentissage de manière itérative en partant de la solution précédente et en la modifiant pour prendre en compte les nouvelles données. Cette approche est bien sûr très peu coûteuse en calculs (Sykacek et al., 2004). Une telle approche est par exemple adaptée au cas de l'analyse discriminante linéaire (Hastie et al., 2001) en considérant, lors de l'acquisition de chaque nouvelle donnée, une petite variation des séparations linéaires précédentes. Dans notre cas, le terme de régularisation employé dans la fonction de coût du classifieur force la plupart des coefficients à être nuls mais rend également délicat la modification des coefficients de la régression logistique de manière itérative. En raison du faible coût de la rapidité de l'apprentissage, nous préférons donc recommencer l'apprentissage à partir de zéro.*

3.1 Modéliser les probabilités en sortie du classifieur

Afin de contrôler les probabilités en sortie du classifieur, nous abordons à nouveau le problème dans un cadre Bayésien et en utilisant des approximations variationnelles pour parvenir à des équations de mise à jour simples et rapides à calculer.

Nous voulons modéliser les probabilités selon une loi simple qui se résume à quelques paramètres. Naturellement, quand on s'intéresse à des lois décrivant des probabilités, on pense tout d'abord à la distribution de Dirichlet. Dans un cadre Bayésien, on veut alors trouver la loi *a priori* conjuguée de la distribution de Dirichlet telle que la loi *a posteriori* sur les paramètres de la loi de Dirichlet suive cette même loi *a priori*. Or même si la distribution de Dirichlet appartient à la famille des lois exponentielles (Koopman, 1936), et que par conséquent, on sait trouver facilement une loi conjuguée, l'expression de la loi conjuguée n'est pas une forme connue de distribution (voir à ce propos la remarque avant

l'exercice 3.3.15 de Robert (2001)). Par conséquent, nous préférons adopter l'approche préconisée dans Gelman (1995), consistant à appliquer aux probabilités observées une transformation de manière à retomber sur une loi Gaussienne. A nouveau, nous écrivons les équations dans le cas de deux tâches mais celles-ci sont généralisables au cas d'un nombre de tâches supérieur à deux. La transformation considérée est

$$\forall i \in \{1, 2\} \quad p_{i,t} = \frac{e^{\phi_{i,t}}}{\sum_{j=1}^2 e^{\phi_{j,t}}}, \quad (\text{VI.5})$$

et par conséquent,

$$\phi_t = \phi_{2,t} - \phi_{1,t} = \ln \left(\frac{p_{2,t}}{p_{1,t}} \right). \quad (\text{VI.6})$$

La distribution de la nouvelle variable peut alors être approximée par une Gaussienne (Gelman, 1995). On a ainsi

$$\phi_t \sim \mathcal{N}(m_t, a_t) \quad (\text{VI.7})$$

où m_t est la moyenne de la loi Gaussienne et a_t est la variance. Or la loi Gaussienne appartient également à la famille des distributions exponentielles (Robert, 2001, proposition 3.3.13), on peut le voir en écrivant

$$\begin{aligned} f(\phi_t | m_t, a_t) &= \frac{1}{\sqrt{2\pi a_t}} \exp \left(-\frac{1}{2a_t} (\phi_t - m_t)^2 \right) \\ &= \exp \left(-\frac{\phi_t^2}{2a_t} + \frac{2m_t \phi_t}{2a_t} - \frac{m_t^2}{2a_t} - \ln(\sqrt{2\pi a_t}) \right) \\ &= \exp \left(-\frac{1}{2a_t} [1 \quad -m_t \quad -m_t \quad m_t^2] [\phi_t^2 \quad \phi_t \quad \phi_t \quad 1]^T - \frac{1}{2} \ln(2\pi a_t) \right) \end{aligned} \quad (\text{VI.8})$$

Une loi conjuguée de la forme précédente est alors donnée par la proposition citée précédemment de Robert (2001). On peut écrire cette loi conjuguée (Šmídl et Quinn, 2005)

$$f(m_t, a_t | V_t, \nu_t) \propto \exp \left(-\frac{1}{2a_t} [1 \quad -m_t]^T V_t [1 \quad -m_t] - \nu_t \ln(2\pi a_t) \right). \quad (\text{VI.9})$$

Celle-ci peut être reconnue comme une loi normale-inverse-gamma¹ de paramètres $V_t \in \mathbb{R}^{2 \times 2}$ et $\nu \in \mathbb{R}$. On sait alors (Šmídl et Quinn, 2005) exprimer de manière simple, pour toute nouvelle donnée ϕ_t , la mise à jour des statistiques suffisantes décrivant $f(m_t, a_t | V_t, \nu_t)$

$$V_t = V_{t-1} + \begin{pmatrix} \phi_t^2 & \phi_t \\ \phi_t & 1 \end{pmatrix} \quad \text{et} \quad \nu_t = \nu_{t-1} + 1. \quad (\text{VI.10})$$

De plus, on sait, en utilisant les moments théoriques des marginales de la loi normale-inverse-gamma, qu'une estimation de la moyenne de ϕ_t est

$$\widehat{m}_t = \frac{[V_t]_{2,1}}{[V_t]_{2,2}} \quad (\text{VI.11})$$

et une estimation de la variance à l'instant t est donnée par

$$\widehat{a}_t = \frac{[V_t]_{1,1} - [V_t]_{2,1}^2}{\nu_t - 5} \quad (\text{VI.12})$$

¹ On aurait pu, en simplifiant les expressions initiales, aboutir à une loi normale-inverse- χ^2 . Cette dernière n'est cependant qu'un cas particulier de la loi normale-inverse-gamma. La loi normale-inverse-gamma exprime le fait que les marginales par rapport respectivement à la moyenne et à la variance sont normale et inverse-gamma.

Pourquoi avoir utilisé ce cadre Bayésien pour aboutir à de simples calculs de moyenne VI.11 et de variance VI.12 ? Si ces expressions fonctionnent bien dans le cas d'un processus stationnaire, elles sont inadaptées dans le cas de l'estimation de la moyenne et de la variance d'un processus non-stationnaire. Dans ce cas, une solution consiste à faire appel au concept de facteur d'oubli qui introduit une perturbation à l'aide d'une distribution auxiliaire. Or si l'introduction du facteur d'oubli a été faite à l'aide d'une heuristique dans (Peterka, 1981), le cadre Bayésien a permis de le formaliser (Kulhavý et Zarrop, 1993). Ainsi, pour un facteur d'oubli $\psi \in [0; 1]$ fixé, les équations VI.10 deviennent (Šmídl et Quinn, 2005)

$$V_t = \psi V_{t-1} + \begin{pmatrix} \phi_t^2 & \phi_t \\ \phi_t & 1 \end{pmatrix} + (1 - \psi)V_0 \quad \text{et} \quad \nu_t = \psi \nu_{t-1} + 1 + (1 - \psi)\nu_0. \quad (\text{VI.13})$$

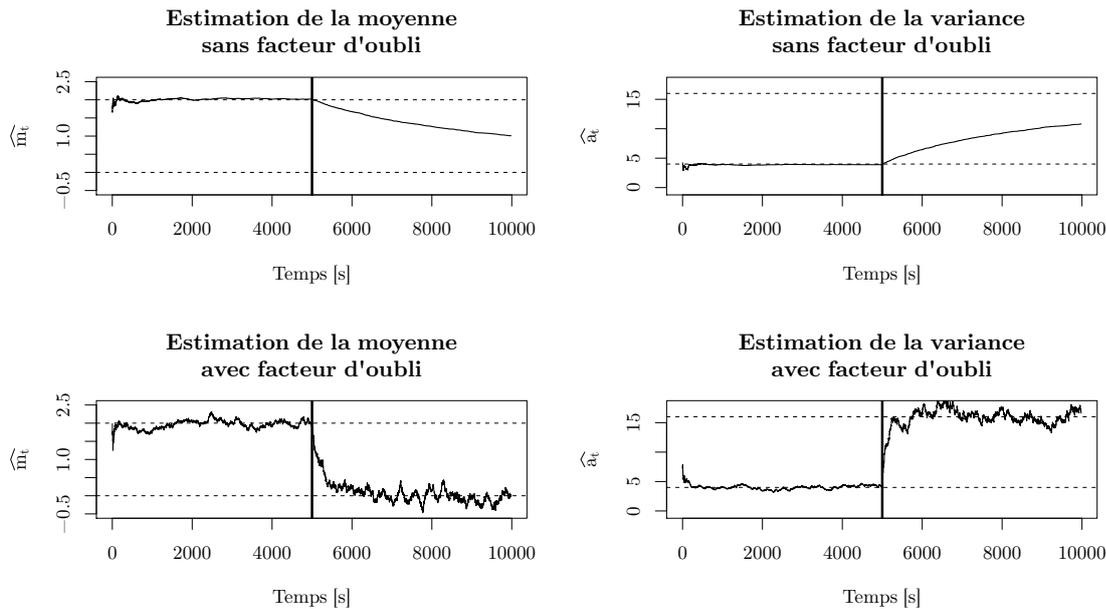


Figure VI.4 Illustration de l'estimation de la moyenne et de la variance d'un processus Gaussien comportant un saut à $t = 5000$. Le facteur d'oubli est $\psi = 0,995$. La distribution de perturbation est telle que $V_0 = 0_{2 \times 2}$ et $\nu_0 = 1$. Pendant les premières 5000 secondes, $\phi_t \sim \mathcal{N}(2, 4)$ et ensuite $\phi_t \sim \mathcal{N}(0, 16)$.

On se sert alors de l'estimation des paramètres de la loi de ϕ_t pour évaluer la dérive et effectuer le réapprentissage. Dans le cas d'un paradigme équilibré, la moyenne \widehat{m}_t doit être relativement proche de zéro. L'idée est de fixer un seuil qui, une fois franchi par la valeur absolue de la moyenne, déclenche un réapprentissage du classifieur. De plus, on entretient un ensemble de données (de taille constante) dont on est sûr de la classification en utilisant la variance estimée $\widehat{\nu}_t$. Ce jeu de données est initialisé avec les caractéristiques de la session d'entraînement et, au fur et à mesure de la session d'évaluation, on remplace les caractéristiques des essais les plus anciens par les caractéristiques des fenêtres classées de manière univoque. Comme on connaît la loi estimée de ϕ_t , les seuils qui permettent de définir une classification « univoque » sont définis de manière simple à partir du nombre d'essais moyen que l'on désire inclure dans la base d'apprentissage et de la fonction cumulative de la Gaussienne de moyenne \widehat{m}_t et de variance $\widehat{\nu}_t$.

3.2 Application aux interfaces asynchrones

Cette procédure adaptative a été appliquée aux données de la troisième compétition BCI (voir chapitre V pour une description). Nous avons évalué la procédure adaptative

présentée dans ce chapitre pour la classification en continu des états mentaux. Contrairement au chapitre précédent, nous utilisons la méthode décrite sur la [figure III.5](#) (les filtres spatiaux trouvés ne sont pas spécifiques aux différentes bandes de fréquence). L'état de repos n'est pas appris explicitement, c'est un paramètre qui varie afin de construire les courbes ROC. Nous comparons quatre déclinaisons de la méthode :

- la matrice de confusion est calculée de manière directe à partir des probabilités brutes en sortie du classifieur ; aucune procédure adaptative n'est utilisée lors de la session d'évaluation ; cette méthode est appelée « non adaptative » par la suite ;
- les décisions sont obtenues après filtrage par la méthode présentée dans le [chapitre V](#) ; aucune procédure adaptative n'est utilisée lors de la session d'évaluation ; cette méthode est appelée « vbhmm + non adaptative » par la suite ;
- on n'utilise pas le lissage par modèle de Markov mais la procédure adaptative est employée. L'ensemble d'apprentissage est mise à jour lorsque $|\phi_t - m_t| > 2\sqrt{a_t}$ (soit environ 10 % des données en moyenne). Enfin, le taux d'oubli est fixé à 0,995 afin de ne pas engendrer trop de variations lorsque le sujet réalise plusieurs fois de suite la même tâche.
- on utilise à la fois le lissage des probabilités par le modèle de Markov et la procédure adaptative (avec les mêmes paramètres que précédemment) pour le réapprentissage du classifieur.

Les autres paramètres de la méthode sont fixés de la même manière que dans le [chapitre V](#). A partir des matrices de confusion construites en faisant varier la probabilité *a priori* pour que le sujet soit au repos, les mesures de précision et d'inertie telles que définies dans la [figure V.6](#) sont calculées pour chacune des méthodes. Les résultats sont rapportés dans la [figure VI.5](#).

Cette figure montre clairement l'intérêt d'une procédure adaptative visant à réapprendre périodiquement les paramètres du classifieur. On y voit également que la combinaison entre la procédure adaptative et le filtrage des probabilités par le modèle de Markov fournit les meilleurs résultats. De manière générale, la procédure adaptative permet d'améliorer significativement les résultats obtenus par rapport à son homologue non-adaptatif.

Illustrant le fonctionnement de chacune des déclinaisons présentées, la [figure VI.6](#) montre un échantillon des décisions obtenues à l'aide de chacune des méthodes pendant un intervalle de temps relativement important (200s). On y voit clairement l'intérêt de la procédure adaptative qui rééquilibre la sortie du classifieur (traces 4 et 5 par rapport à 2 et 3). Enfin, cette figure illustre à nouveau l'intérêt du filtrage des probabilités par le modèle de Markov puisqu'on voit sur les traces 3 et 5 une bien meilleure stabilité des états mentaux.

4 Discussion et conclusion

Ce chapitre nous a permis de présenter quelques idées simples pour la conception d'interfaces cerveau-machines plus robustes et plus efficaces. Les idées préliminaires qui ont été présentées utilisent d'une manière originale les méthodes développées dans l'ensemble de cette thèse et les problèmes abordés sont d'une importance primordiale en vue de l'utilisation des ICMs dans la vie réelle.

Nous nous sommes tout d'abord intéressés au problème de la qualité de réalisation des tâches par les sujets afin d'optimiser la session d'entraînement. Ce problème a été très peu exploré dans la littérature. Or la réalisation d'un mauvais essai par le sujet peut détériorer de manière importante la qualité des filtres spatiaux trouvés ainsi que le classifieur. Nous

Diagnostic inertie/précision du système

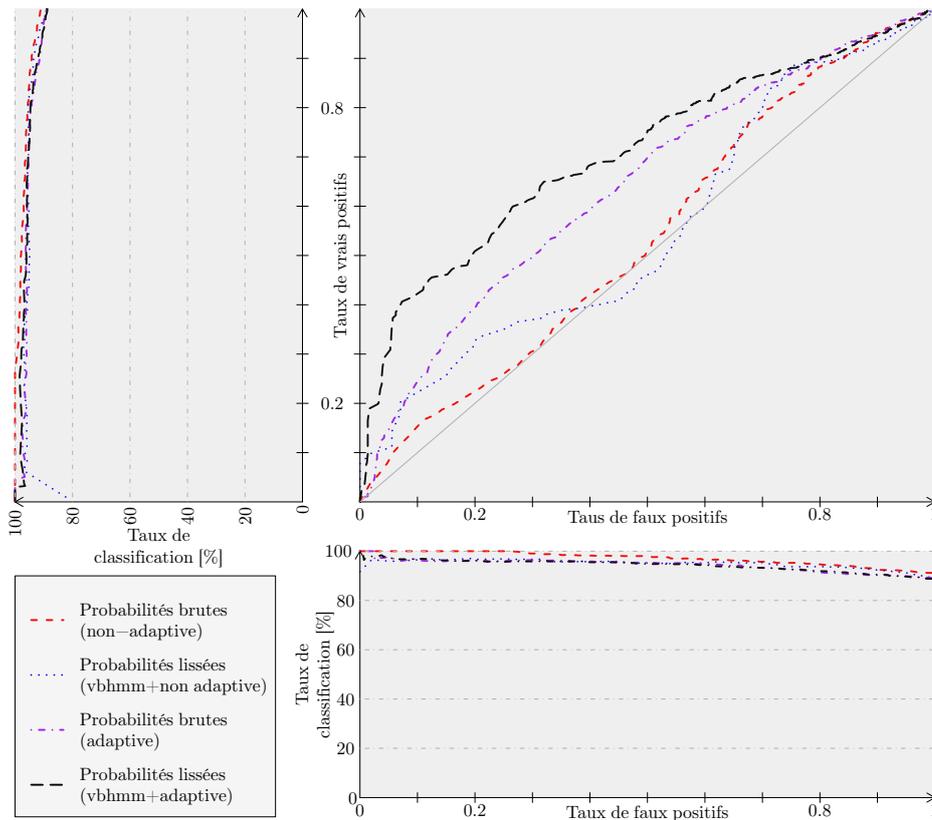


Figure VI.5 En haut à droite, courbes ROC correspondant aux quatre déclinaisons proposées. Les courbes sont obtenues en faisant varier entre zéro et un la probabilité *a priori* pour que le sujet soit en activité. En haut à gauche et en bas à droite, courbes de précision des méthodes en fonction respectivement du taux de vrais positifs et du taux de faux positifs.

avons proposé un moyen relativement simple de traiter ce problème en tentant de résumer la qualité de réalisation de chaque essai à l'aide d'un unique score. Pour ce faire, nous avons utilisé de manière originale les informations fournies par le filtrage des probabilités à l'aide du modèle présenté dans le [chapitre V](#). Ce score est basé sur l'observation de la modification de l'estimation de la matrice de transition lors du filtrage des probabilités observées. Nous proposons d'utiliser ce score afin d'éliminer les essais mal réalisés par le sujet mais également afin de pouvoir détecter un état de fatigue ou de lassitude du sujet lorsqu'il ne parvient plus à maintenir la tâche demandée pendant un intervalle de temps suffisamment long.

Enfin, nous avons proposé une façon simple et originale pour rendre les algorithmes adaptatifs. Comme largement développé dans [Krauledat \(2008\)](#), ce problème est fondamental dans un système réel puisqu'il est courant d'observer une dérive des scores de classifications liée aux changements de l'activité de fond du sujet ou encore à l'apprentissage du sujet qui modifie sa manière de réaliser la tâche en observant et en s'adaptant aux réponses du système. La méthode présentée est basée sur l'estimation continue en temps réel de la distribution de probabilité d'une variable gaussienne obtenue par une transformation des probabilités observées initiales.

Les deux idées proposées dans ce chapitre abordent des problèmes cruciaux rencontrés avec les données réelles. Les deux méthodes présentées dans ce chapitre peuvent bien sûr être couplées, rendant ainsi le système plus robuste aux données aberrantes de la phase d'entraînement et sachant s'adapter à l'inconstance du sujet.

Classification continue, FPR=0.25
R=repos, L=main gauche, F=pied

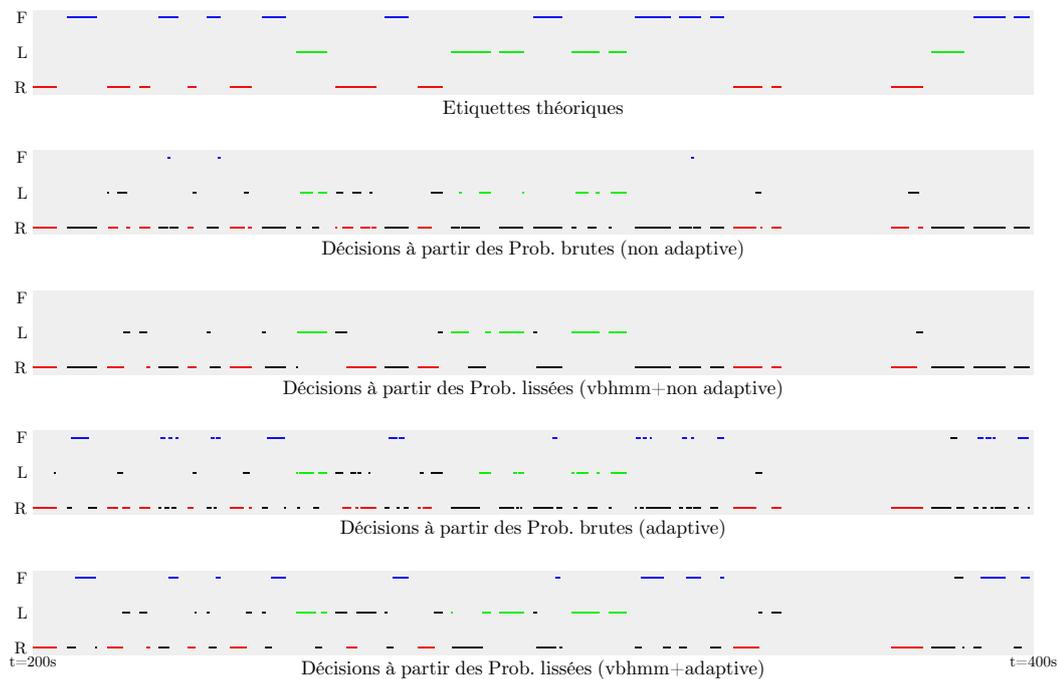


Figure VI.6 Classification continue pour chacune des quatre déclinaisons pendant une période d'une durée de 200s. Les tâches théoriquement réalisées par le sujet, repos en rouge, main gauche en vert et pied en bleu, sont représentées dans la première trace (haut). Les quatre suivantes représentent les décisions à partir des quatre déclinaisons considérées dans cette partie. Dans les quatre traces du bas, les échantillons mal classés sont en noir alors que les points bien classés sont de la couleur de leur classe respective.

5 Références du chapitre

- Šmídl, V. et Quinn, A. (2005). *The Variational Bayes Method in Signal Processing (Signals and Communication Technology)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.
- Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M. et Müller, K.-R. (2008). Optimizing spatial filters for robust EEG Single-Trial analysis. *IEEE Signal Process. Mag.*, 25(1):41–56.
- Buttfield, A., Ferrez, P. W. et Millán, J. d. R. (2006). Towards a robust BCI: error potentials and online learning. *IEEE Trans. Neural Syst. Rehabil. Eng.*, 14(2):164–168.
- Gelman, A. (1995). Methods of moments using Monte Carlo simulation. *J. Comput. Graph. Stat.*, 4(1):36–54.
- Grosse-Wentrup, M. et Buss, M. (2008). Multiclass common spatial patterns and information theoretic feature extraction. *IEEE Trans. Biomed. Eng.*, 55(8):1991–2000.
- Hastie, T., Tibshirani, R. et Friedman, J. H. (2001). *The Elements of Statistical Learning*. Springer.
- Koopman, B. O. (1936). On distributions admitting a sufficient statistic. *Trans. Am. Math. Soc.*, 39(3):399–409.
- Krauledat, M. (2008). *Analysis of Nonstationarities in EEG Signals for Improving Brain-Computer Interface Performance*. Thèse de doctorat, Technical University of Berlin.
- Kulhavý, R. et Zarrop, M. B. (1993). On a general concept of forgetting. *International Journal of Control*, 58(4):905–924.

- Peterka, V. (1981). *Trends and progress in system Identification*, chapitre Bayesian approach to system identification, pages 239–304. Pergamon Press, Oxford.
- Pfurtscheller, G., Neuper, C., Guger, C., Harkam, W. et Ramoser, H. et al. (2000). Current trends in Graz Brain-Computer Interface (BCI) research. *IEEE Trans. Rehabil. Eng.*, 8(2):216–219.
- Robert, C. (2001). *The Bayesian Choice*. Springer, 2nd édition.
- Shenoy, P., Krauledat, M., Blankertz, B., Rao, R. P. N. et Müller, K.-R. (2006). Towards adaptive classification for BCI. *J. Neural Eng.*, 3(1):R13–R23.
- Sykacek, P., Roberts, S. J. et Stokes, M. (2004). Adaptive BCI based on variational bayesian Kalman filtering: an empirical evaluation. *IEEE Trans. Biomed. Eng.*, 51(5):719–727.
- Vidaurre, C., Schlögl, A., Cabeza, R., Scherer, R. et Pfurtscheller, G. (2006). A fully on-line adaptive BCI. *IEEE Trans. Biomed. Eng.*, 53(6):1214–1219.
- Wang, H. et Zheng, W. (2008). Local temporal common spatial patterns for robust single-trial EEG classification. *IEEE Trans. Neural Syst. Rehabil. Eng.*, 16(2):131–139.
- Wolpaw, J. R., McFarland, D. J. et Vaughan, T. M. (2000). Brain-computer interface research at the wadsworth center. *IEEE Trans. Rehabil. Eng.*, 8(2):222–226.

Chapitre VII

CONCLUSION ET PERSPECTIVES



Nous concluons à présent ce manuscrit en revenant sur les éléments nouveaux et les principaux résultats de cette thèse. Les limites de chaque approche sont mises en valeur afin de proposer de futurs développements ou des pistes de recherche.

1 Filtrage spatial pour les ICMs

Nous avons proposé deux méthodes distinctes pour le filtrage spatial des données EEG. La première est destinée à faire apparaître les sources cérébrales responsables de la réalisation des tâches mentales proposées aux sujets lors d'une expérience d'ICM. Nous avons montré qu'une des méthodes la plus utilisée dans le contexte des interfaces cerveau-machines (« Common Spatial Pattern ») pouvait être interprétée facilement dans le cadre de l'extraction de sources non-stationnaires. Cette formulation nous a conduit à proposer une méthode de filtrage spatial neurophysiologiquement adaptée, dans la mesure où elle cherche à prendre en compte les activations successives de différentes sources cérébrales lors de la réalisation d'une tâche mentale. Nous avons montré que cette méthode était significativement plus performante que la méthode du « Common Spatial Pattern » et qu'elle surpassait, pour la plupart des sujets, une méthode analogue ne prenant pas en compte les successions d'activations cérébrales. En comparant nos résultats avec ceux de la dernière compétition internationale du domaine, nous avons montré que notre méthode nous permettait de nous classer parmi les trois meilleurs compétiteurs. L'inconvénient de notre méthode réside cependant dans le fait qu'elle suppose que les successions d'activation des sources cérébrales se déroulent toujours de la même manière (même enchaînement) et au même moment au cours de la réalisation d'une tâche mentale. Cette hypothèse forte peut constituer une limite importante de la méthode puisqu'on sait que la variabilité entre les différents essais est souvent importante dans le cas de tâches mentales telles que l'imagerie motrice.

Pouvons-nous espérer améliorer encore significativement les résultats de classification à l'aide de méthodes de filtrage spatial ? Cette question est primordiale afin d'orienter les futures recherches sur les méthodes de traitement du signal pour les ICMs. Au cours de ces trois années, j'ai plusieurs fois envisagé (au fur et à mesure de l'augmentation des capacités de calcul à ma disposition) de réaliser l'expérience numérique suivante : en considérant un jeu de données classique d'une expérience ICM synchrone, comportant un nombre de capteurs raisonnable (de l'ordre de 20), et étant donnée une méthode fixe de calcul des caractéristiques et de classification, trouver la loi de probabilité du score de classification en choisissant un ensemble de cardinal constant de filtres spatiaux dont les coefficients sont aléatoires. On pourrait alors reporter dans la distribution correspondante

la position de chacune des méthodes de filtrage spatial linéaire et estimer la nécessité d'améliorer les méthodes existantes. Malheureusement, si l'on choisit un ensemble de six filtres spatiaux, la dimension de l'espace des paramètres est de l'ordre de cent... c'est pourquoi l'idée est restée jusqu'à présent irréalisable. On peut néanmoins penser, en observant les améliorations actuellement obtenues à l'aide de méthodes de filtrage spatial, que le temps passé à optimiser les filtres commence à devenir trop important par rapport à l'enjeu. On peut par contre envisager d'utiliser des méthodes de filtrage non-linéaires. Bien que l'emploi de telles méthodes ne semble pas justifié par les connaissances sur la propagation physique des signaux cérébraux, la question de l'intérêt de méthode de filtrage non-linéaire reste relativement ouverte. Une des principales limitations qui a favorisé l'emploi de méthodes linéaires est l'exigence de temps réel et la nécessité d'avoir des algorithmes pouvant être appliqués de manière simple et rapide.

Nous avons également proposé une méthode de débruitage linéaire très efficace pour enlever les artefacts oculaires des enregistrements EEG. Si l'intérêt de cette méthode n'a pas été très significatif pour augmenter les scores de classification des données en notre possession, nous pensons néanmoins qu'une telle méthode, simple, efficace et pouvant être utilisée de manière quasi-automatique, peut s'avérer très utile dans les conditions d'enregistrement de la vie réelle. Nous pensons par exemple que l'intégration des systèmes ICMs en tant qu'élément d'un système multimodal impliquant plusieurs modalités perceptives (e.g. retours proprioceptifs, périphériques haptiques, stimulations visuelles) devra impliquer une méthode de débruitage très efficace afin que chacune des modalités ne vienne pas perturber les mesures EEG.

Enfin, la méthode de débruitage que nous avons proposée peut s'appliquer à d'autres types d'artefacts ainsi que d'autres systèmes d'acquisition. On pourra par exemple considérer le débruitage des signaux cérébraux contaminés par des artefacts cardiaques.

2 Utilisation de la dynamique des états mentaux

Nous avons proposé une méthode de filtrage des probabilités en sortie du classifieur afin de prendre en compte le fait que les états mentaux des sujets doivent être relativement stables d'une fenêtre d'analyse à l'autre. Ce filtrage suppose que les états mentaux du sujet peuvent être modélisés par une chaîne de Markov cachée, il a été réalisé dans un cadre variationnel Bayésien. En utilisant deux jeux de données provenant de deux compétitions distinctes, nous avons montré qu'un tel modèle pouvait, pour la plupart des sujets, conduire à des améliorations de performances de classification pour un paradigme nécessitant la classification en continu des états mentaux. Ces améliorations ont été modestes pour certains d'entre eux et plus importantes pour d'autres. Néanmoins, nous avons montré que, même si seules quelques fenêtres temporelles étaient mieux classées à l'aide du filtrage, l'amélioration ergonomique qui en résulte peut s'avérer très importante dans certains systèmes, empêchant la réponse d'osciller entre deux décisions.

Nous avons également proposé une utilisation originale de la modélisation de la dynamique des états mentaux afin de quantifier la réalisation de chaque essai par le sujet lors de la phase d'apprentissage. Cette méthode peut être utilisée afin d'aborder plusieurs problèmes : tout d'abord nous avons montré que le suivi de la réalisation des essais lors de la phase d'apprentissage permettait d'adapter facilement la longueur de la session d'apprentissage aux capacités du sujet ; de plus, il est possible d'utiliser notre mesure afin de détecter un état de fatigue ou une lassitude du sujet ; enfin, il est possible d'enlever de la base de données d'apprentissage les essais qui sont mal réalisés. Nous avons ainsi montré que, dans certains cas, l'utilisation d'une base d'apprentissage épurée pouvait conduire à une amélioration des performances lors de la phase d'évaluation.

Nous avons également proposé une méthode adaptative afin de mettre à jour la méthode de classification en fonction des variations de réalisation des tâches par le sujet.

Nous avons montré que cette approche améliorerait significativement les performances de classification en continu lors de la phase d'évaluation. Néanmoins, notre approche a été utilisée dans un cas simple en utilisant l'*a priori* selon lequel la distribution des probabilités pour que le sujet soit dans chacune des tâches devait être symétrique. Si cet *a priori* est effectivement utilisable dans le cas d'un paradigme visant à déplacer un curseur sur un écran, il existe des paradigmes pour lesquels il n'y a aucune raison pour que l'ensemble des tâches soient équiprobables (« P300 speller » par exemple). D'autres procédures adaptatives devront alors être envisagées afin de prendre en compte les inconstances de réalisation des tâches par les sujets.

Nous avons utilisé une manière très simple de rendre le système adaptatif puisque le classifieur est entièrement réappris. De futurs développements pourront rendre cette phase plus intelligente en modifiant légèrement de manière itérative les paramètres du classifieur. De plus, nous n'avons pas considéré la possibilité de réapprendre les filtres spatiaux en plus du système de classification. Ceci pourrait être fait selon les mêmes idées que celles développées dans cette thèse. L'une des difficultés d'une telle approche réside dans la constitution d'un jeu de données bien étiqueté pour chacune des tâches (on ne connaît en effet pas la tâche réelle que le sujet a voulu réaliser lors de la phase d'évaluation).

3 Vers des modèles dynamiques plus précis

Afin de pallier les possibles variabilités de la réalisation des tâches qui ne pourraient pas être prises en compte par la méthode de filtrage linéaire que nous avons proposée, de futurs travaux pourront envisager des modèles dynamiques plus précis des activations successives des sources cérébrales lors de la réalisation d'une tâche mentale. Dans un premier temps, cette approche devra définir clairement des « micro-états » mentaux, atomes élémentaires de la réalisation d'une tâche. Un atome élémentaire devra être caractérisé par un ensemble de filtres spatiaux et une fréquence spécifique (une source cérébrale et une fréquence de fonctionnement privilégiée). Si différentes zones cérébrales sont connues pour être impliquées dans la réalisation d'une tâche d'imagination de mouvement, il est néanmoins difficile d'en déduire une source cérébrale (et un filtrage associé) de manière univoque pour chaque individu. C'est pourquoi nous pensons qu'une première phase préliminaire d'analyse des sources cérébrales potentiellement impliquées dans la réalisation d'une tâche est nécessaire.

Nous avons entrepris certains travaux dans ce sens et avons exploré la possibilité d'utiliser des méthodes d'apprentissage semi-supervisées pour trouver des intervalles de temps pendant lesquels les activités cérébrales sont similaires. L'idée est de regrouper les intervalles à la fois selon leur ressemblance (méthode non-supervisée) et la tâche que le sujet était théoriquement en train de réaliser (méthode supervisée). Afin de mesurer la ressemblance, nous avons considéré pour l'instant la distance cepstrale à partir de la modélisation autorégressive sur des fenêtres courtes (de l'ordre de une seconde) ainsi que la divergence de Kullback-Leibler symétrisée entre les spectres sur chacune des fenêtres (considérés alors comme une densité de probabilité). L'idée de ces mesures est de construire une matrice de distances entre les spectres de chacune des fenêtres considérées. A partir de la matrice de distances, on a alors utilisé des algorithmes de « clustering » afin de regrouper les intervalles similaires. Dans un deuxième temps, l'idée est d'analyser les regroupements en observant les appartenances des intervalles lors de la réalisation par le sujet de tous les essais d'une même classe.

Cette approche n'a pour l'instant pas permis de mettre en évidence une quelconque régularité qui pourrait être utilisée afin de construire un modèle dynamique fin des activations successives des sources cérébrales. Cependant, cette démarche pourrait s'avérer très utile afin de discriminer un état de repos par rapport à un état d'activité. Cette piste pourrait donc être prometteuse dans le cadre des ICMs. Notons cependant qu'une

approche semblable à partir de chaînes de Markov a été utilisée à de multiples reprises sans grand succès. Ces approches n'utilisaient cependant pas de modèles *a priori* et la dynamique était apprise uniquement en fonction des données. On peut espérer obtenir des résultats plus probants en utilisant une approche basée sur des *a priori* plus importants.

4 Conclusion générale

Les travaux présentés dans cette thèse ont été résolument tournés vers les données réelles. Nous avons tout au long du manuscrit tenté de proposer des solutions simples à de réels problèmes rencontrés lors de la conception des systèmes ICMs. Cette approche a été à la fois riche d'enseignements et quelquefois frustrante. Les résultats que nous avons obtenus ont en effet été rarement univoques et aucune des méthodes proposées ne peut se targuer de surpasser toutes les autres pour l'ensemble des sujets analysés. Ainsi, si les méthodes d'apprentissage et de traitement du signal sont primordiales pour la conception des ICMs, les données réelles nous rappellent que nous sommes encore loin de comprendre toutes les finesses du fonctionnement du cerveau.

Nous pensons qu'une analyse approfondie des erreurs réalisées par le système pourrait être très instructive afin de comprendre pourquoi certaines méthodes s'avèrent excellentes pour certains sujets et beaucoup moins performantes pour d'autres. Dans l'attente de telles explications, les nouvelles méthodes de traitement du signal peuvent potentiellement permettre de diminuer le nombre de personnes que l'on considère comme incapables de contrôler un tel système (on les appelle « BCI illiterate » en anglais). En développant des méthodes innovantes pour l'analyse des signaux cérébraux, nous espérons donc avoir œuvré dans ce but.

Si l'aspect « traitement du signal » est primordial dans la conception de systèmes ICMs, il faut néanmoins garder à l'esprit que le développement de tels systèmes s'est toujours déroulé aux frontières d'un ensemble de compétences complémentaires. Les travaux réalisés dans cette thèse nous en ont convaincus. En exprimant la volonté de vouloir rendre les systèmes ICMs accessibles à un vaste public, certaines compagnies semblent vouloir aujourd'hui orienter les ICMs vers une nouvelle ère de leur existence, une ère résolument tournée vers l'utilisation des systèmes dans des conditions d'enregistrements extrêmes. Les défis soulevés par de tels desseins seront certainement plus pratiques et soulèveront de nouvelles problématiques qui commencent à être abordées actuellement dans la littérature (diminution du temps d'apprentissage, simplicité d'emploi du système d'acquisition, intégration avec d'autres modalités sensorielles, ergonomie du retour perceptif). Nous espérons que les méthodes proposées dans cette thèse sauront répondre à certaines d'entre elles, et, plus généralement, que les démarches utilisées et les questions soulevées sauront faire naître de nouvelles idées fructueuses.

Annexe A

CLASSIFICATION EN GRANDE DIMENSION



Nous l'avons vu tout au long de cette thèse, la classification est un élément crucial de la conception des interfaces cerveau-machines (ICMs). Les conditions d'apprentissage imposées sont souvent extrêmes puisque la dimension de l'espace des variables a souvent été supérieure au nombre d'essais. Après avoir essayé plusieurs techniques de classification, notre choix s'est arrêté sur la régression logistique régularisée décrite dans [Friedman et al. \(2008\)](#). Nous allons ici décrire les grandes lignes du fonctionnement de cette technique et expliquer en quoi elle est particulièrement adaptée au problème de la classification pour les ICMs.

1 Fonction de coût pour l'apprentissage

Considérons le vecteur de prédiction $X \in \mathbb{R}^p$ (régresseur) ainsi que la variable catégorielle à prédire G pouvant prendre un ensemble fini de K valeurs. Dans le cas où $K > 2$, on peut généraliser le modèle de régression logistique ([Hastie et al., 2001](#)) par un modèle « multi-logit » en se ramenant à $K - 1$ problèmes binaires

$$\forall l \in 1, \dots, K - 1 \quad \ln \frac{\mathbb{P}(G = l | x)}{\mathbb{P}(G = K | x)} = \beta_{0l} + x^T \beta_l \quad (\text{A.1})$$

où $x \in \mathbb{R}^p$ est une réalisation du vecteur aléatoire X . $\beta_{0l} \in \mathbb{R}$ est le terme d'intersection du modèle et $\beta_l \in \mathbb{R}^p$ représente donc le vecteur des paramètres du modèle. On peut aussi écrire de manière symétrique (au prix d'une augmentation du nombre de paramètres qui est compensée par les conditions imposées par la manipulation de probabilités)

$$\mathbb{P}(G = l | x) = \frac{e^{\beta_{0l} + x^T \beta_l}}{\sum_{k=1}^K e^{\beta_{0k} + x^T \beta_k}} \quad (\text{A.2})$$

La log-vraisemblance de ce modèle, étant donné un ensemble de réalisations de X que l'on note $\{x_1, \dots, x_N\}$, doit alors être maximisée. En incluant le terme de régularisation, elle s'écrit sous la forme

$$\max_{\{\beta_{0l}, \beta_l\}_{l \in [1..K]} \in \mathbb{R}^{K \times (p+1)}} \left[\frac{1}{N} \sum_{i=1}^N \ln p_{g_i}(x_i) - \lambda \sum_{l=1}^K P_\alpha(\beta_l) \right], \quad (\text{A.3})$$

où $p_{g_i}(x_i) = \mathbb{P}(G = l | x_i)$ et $P_\alpha(\beta_l)$ représente le terme de régularisation du modèle défini par

$$P_\alpha(\beta_l) = (1 - \alpha) \frac{1}{2} \|\beta_l\|_{\ell_2}^2 + \alpha \|\beta_l\|_{\ell_1}. \quad (\text{A.4})$$

Cette forme de régularisation (appelée elastinet) a été introduite dans [Zou et Hastie \(2005\)](#) afin de résoudre le problème de l'apprentissage dans le cas de variables très corrélées. C'est une pondération de la pénalisation appelée « Ridge regression » ([Hoerl et Kennard, 1970](#)) qui utilise la norme ℓ_2 afin de compenser les problèmes liés aux variables non-orthogonales, et la méthode « lasso » ([Tibshirani, 1996](#)) qui effectue la sélection de variables en forçant un grand nombre de coefficients à rester nuls. En pratique, tout au long de cette thèse nous avons utilisé un coefficient $\alpha = 0.95$ rendant la méthode proche de la régularisation « lasso » pure.

L'algorithme proposé dans [Friedman et al. \(2008\)](#) permet d'obtenir les coefficients β non pas pour une unique valeur de λ , mais pour un chemin entier, à partir d'une valeur λ initiale, en utilisant une méthode appelée « coordinate descent ». Ainsi, l'optimisation par rapport à la valeur du paramètre λ peut se faire de manière très efficace par validation croisée sans que la complexité de calcul explose.

2 Illustration du fonctionnement

Afin d'illustrer le fonctionnement du classifieur sur des données réelles, nous considérons les données relatives aux quatre sujets réels de la quatrième « BCI competition, dataset 1 » (voir [chapitre V](#) pour une description plus détaillée). En utilisant uniquement les données des sessions d'entraînement, nous avons entraîné le classifieur pour un chemin de valeurs de λ (voir le [chapitre V](#) pour les détails de la méthode). L'algorithme part de la valeur de λ la plus grande et ajoute au fur et à mesure les variables en diminuant la valeur de λ . Ce mécanisme est illustré sur la [figure A.1](#) dans laquelle on voit le nombre de coefficients non nuls augmenter quand les valeurs de λ diminuent. Nous avons de plus indiqué sur la figure la bande de fréquence correspondant à chacun des coefficients en utilisant un code de couleur qui est indiqué sur la figure en bas à droite. Ce graphique permet donc d'analyser d'une nouvelle manière l'importance de chacune des bandes de fréquences utilisées pour réaliser la discrimination. Confirmant les observations réalisées tout au long de cette thèse, on voit que les bandes de fréquence qui apparaissent en premier (sur la gauche de chacun des graphiques) sont relatives à la bande α (alpha, 8 à 13 Hz). D'un point de vue du temps de calcul, notons que l'entraînement à l'aide des 200 essais de la session d'entraînement pour l'ensemble des valeurs de λ considérées se fait de manière immédiate (moins de 100 ms sur un processeur cadencé à 3 GHz).

3 Quantité de données d'apprentissage et régularisation

Nous nous sommes également posé la question suivante : comment varie la valeur optimale du paramètre de régularisation lorsque la taille du jeu de données d'apprentissage varie ? Cette question peut en effet avoir de l'importance dans une procédure consistant à utiliser environ 90 % des données pour trouver, par validation croisée, la valeur optimale du paramètre, puis, dans un deuxième temps, à entraîner en fait le classifieur à l'aide de 100 % des données pour utiliser l'ensemble des informations afin de fixer les paramètres pour la session d'évaluation.

Afin de quantifier ce effet, nous avons réalisé l'étude suivante : en utilisant les données de la troisième « BCI Competition, dataset IIIa », nous avons appliqué la méthode du [chapitre III](#) en faisant varier le nombre d'essais utilisés pour l'apprentissage. Une estimation de l'erreur est alors calculée par validation croisée pour plusieurs valeurs de λ . On relève alors pour chaque valeur du nombre d'essais de la base de données d'entraînement, la valeur de λ qui permet d'obtenir le taux de classification en validation croisée le plus élevé. Les résultats sont rapportés sur la [figure A.2](#). En noir, les valeurs optimales pour chaque nombre d'essais considéré sont reliées par une ligne droite et marquées par des \times .

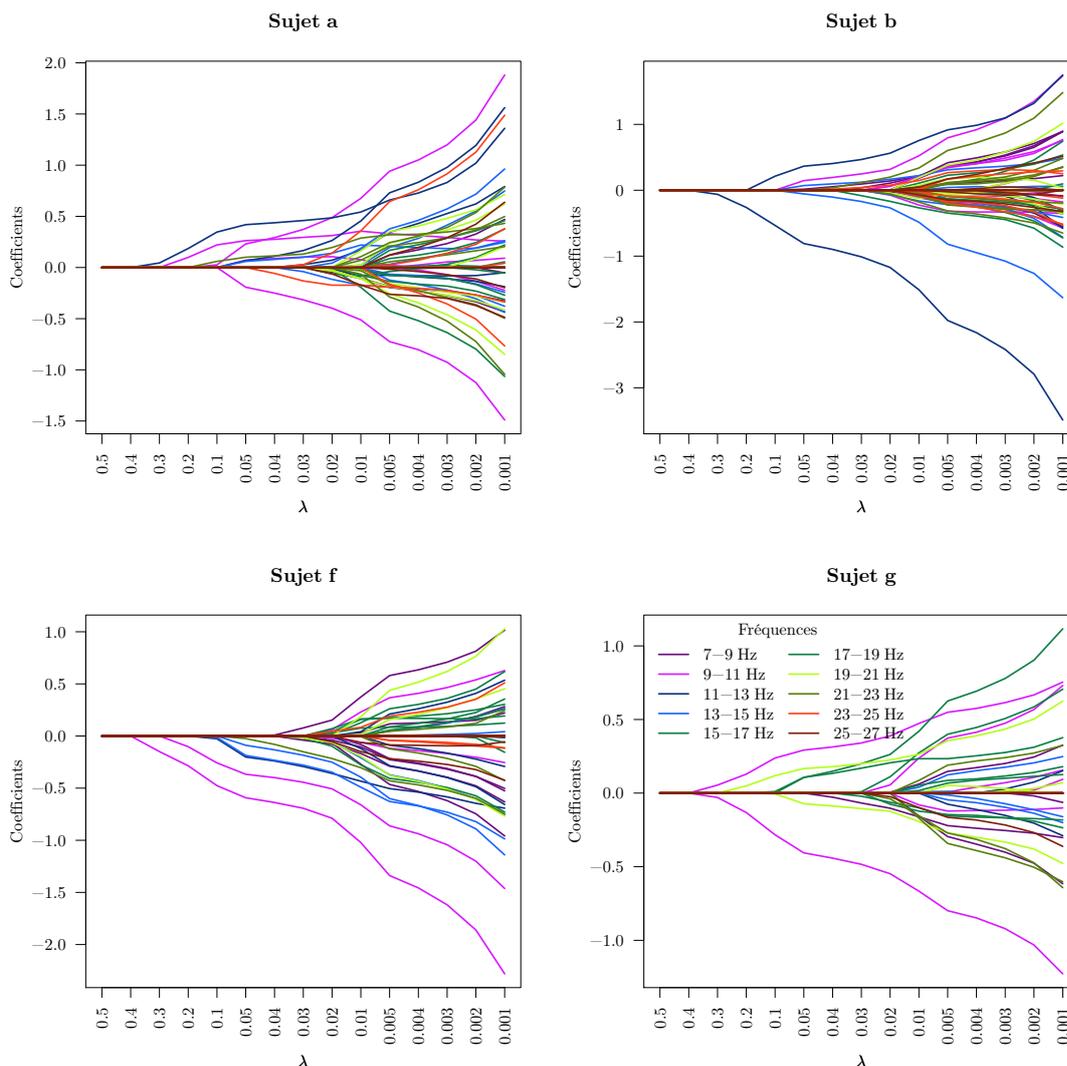


Figure A.1 Influence de la valeur du paramètre de régularisation sur l’entraînement du classifieur pour les quatre sujets réels de la quatrième compétition BCI. En abscisse sont indiquées les valeurs de λ et en ordonnée les valeurs des coefficients du modèle β . La couleur indique la bande de fréquence concernée.

De plus, sur chacune des figures, on a tracé les lignes de niveau représentant le score de classification obtenu par rapport au score maximal.

De manière prévisible, on s’aperçoit que la régularisation doit être de plus en plus drastique lorsque le nombre d’essais est faible. L’effet est relativement minime pour des jeu de données de tailles comparables mais on pourra garder à l’esprit ce phénomène afin de le compenser en sous-estimant la valeur du paramètre de régularisation trouvé à l’aide de 90 % des données s’il doit ensuite être utilisé pour l’ensemble des données.

4 Conclusion

En conclusion, nous avons expliqué les raisons qui nous ont fait préférer cette technique de classification parmi l’ensemble des techniques disponibles. Aux arguments mentionnés dans ce chapitre (rapidité, apprentissage), on peut ajouter que la classification de nouveaux essais est particulièrement simple puisqu’elle consiste en une projection linéaire des caractéristiques selon le vecteur défini par les coefficients β . De plus, la méthode de régularisation elasticnet a permis d’obtenir des scores de classification comparables (voire

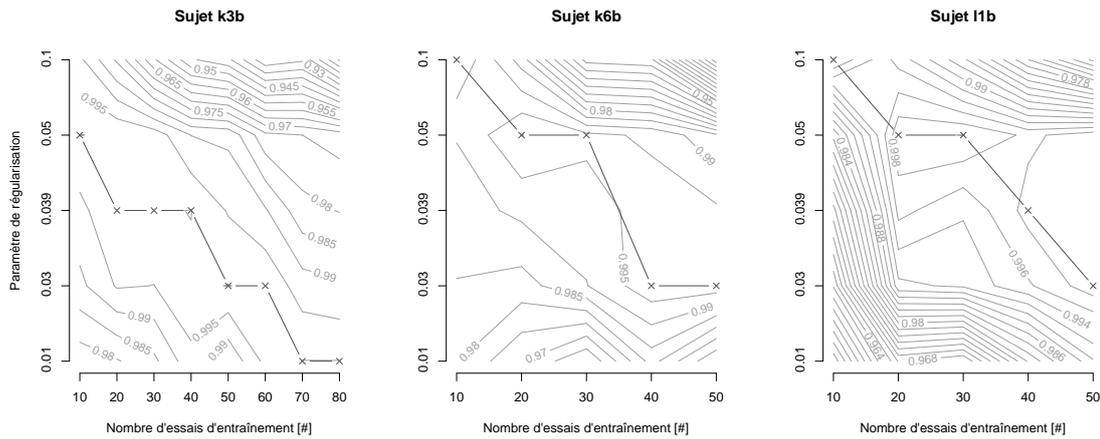


Figure A.2 Taux de classification par rapport au maximum obtenu pour chacun des couples de valeurs (taille des données d'apprentissage, valeur de la régularisation). Les valeurs de régularisation qui permettent d'obtenir le taux de classification maximum pour chaque taille du jeu de données d'apprentissage sont reliées et marquées par une croix (×).

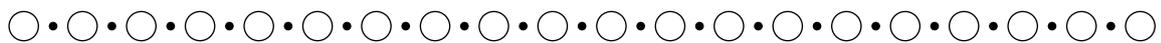
meilleurs) à plusieurs des méthodes de classification les plus performantes dans [Zou et Hastie \(2005\)](#). Enfin, on obtient une probabilité d'appartenance de chacun des essais aux classes ce qui permet de quantifier facilement la valeur de la décision.

5 Références du chapitre

- Friedman, J., Hastie, T. et Tibshirani, R. (2008). Regularized paths for generalized linear models via coordinate descent. The R package glmnet is available from CRAN.
- Hastie, T., Tibshirani, R. et Friedman, J. H. (2001). *The Elements of Statistical Learning*. Springer.
- Hoerl, A. E. et Kennard, R. W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B.*, 58(1):267–288.
- Zou, H. et Hastie, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. B.*, 67(2):301–320.

Annexe B

LA DIAGONALISATION JOINTE EN PRATIQUE



Au cours de cette thèse, nous avons fait une utilisation intensive d'algorithmes de diagonalisation conjointe. Ceci nous a conduit à développer une bibliothèque libre s'intégrant au logiciel R¹. Cette bibliothèque a pour but de fournir un cadre commun et homogène à l'ensemble des algorithmes de diagonalisation conjointe afin que leur utilisation puisse en être simplifiée.

Cette bibliothèque s'appelle « jointDiag », elle est librement téléchargeable sur le site de R à l'adresse <http://cran.r-project.org/web/packages/jointDiag/index.html>. Les sources ainsi qu'un manuel d'utilisation comportant de nombreux exemples sont accessibles au même endroit. En tant qu'extension officielle, la bibliothèque peut également être installée facilement depuis le logiciel R à l'aide des commandes

```
## installing jointDiag from official repository
install.packages("jointDiag")

## loading the package
library(jointDiag)
## then you can look at the examples provided
example(uwedge) # or jedi or ffdiag...
```

R

Différents algorithmes sont fournis par la bibliothèque, la [figure B.1](#) présente les résultats d'une simulation très simple qui permet de comparer les algorithmes en fonction de la qualité du résultat et de la vitesse de convergence. On peut observer une nette supériorité de l'algorithme développé par le Dr. Antoine Souloumiac appelé jedi ([Souloumiac, 2009](#)). L'ensemble des travaux de cette thèse a cependant utilisé jadiag ([Pham, 2001](#) et [Pham et Cardoso, 2001](#)). Les algorithmes jedi et uwedge ont en effet été publiés en 2009.

D'un point de vue de la complexité de calcul, chacun de ces algorithmes est extrêmement efficace. Par exemple, la diagonalisation conjointe d'une dizaine de matrices de taille 20×20 prend moins de 10 ms (version codée en C dans la bibliothèque). Notons néanmoins que les plus rapides lorsque la taille des matrices ou le nombre des matrices augmentent sont jadiag, uwedge et jedi.

¹ Logiciel Open-Source initialement dédié aux calculs statistiques qui peut être téléchargé sur <http://www.r-project.org/>

Convergence des algorithmes de diagonalisation conjointe

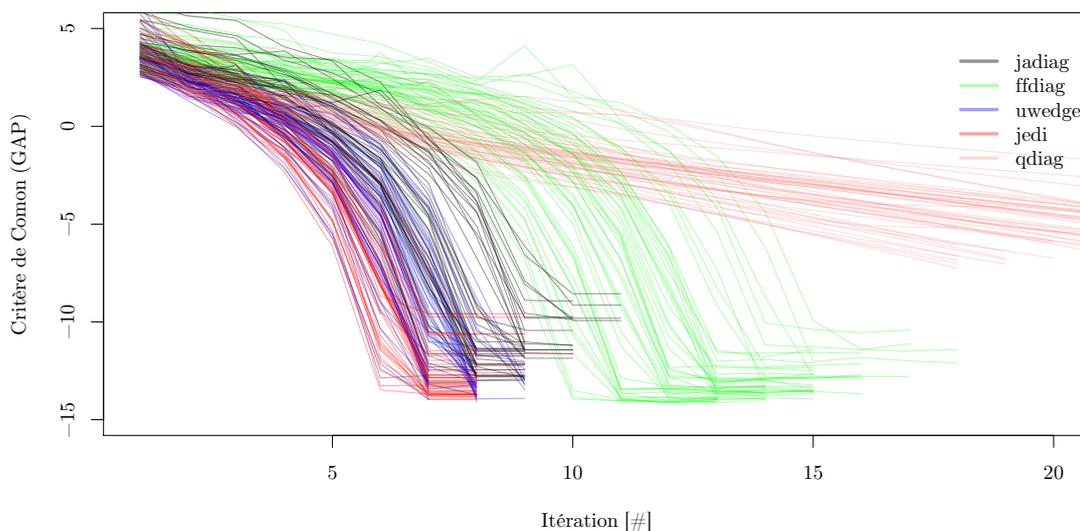


Figure B.1 Convergence de plusieurs algorithmes de diagonalisation conjointe fournis dans la bibliothèque jointDiag. On a effectué 50 diagonalisations successives pour chacun des algorithmes en analysant, après chaque itération, la qualité du résultat d'après un critère défini par Pierre Comon (GAP) dans Comon (1994). Pour chaque simulation, on a généré artificiellement 30 matrices de taille 10×10 parfaitement diagonalisables (cas orthogonal). Chaque courbe tracée correspond à une simulation. Le critère d'arrêt global de chaque algorithme a été choisi de manière à ne pas stopper les itérations avant la convergence. Les algorithmes utilisés sont jdiag (Pham, 2001 et Pham et Cardoso, 2001), ffdiag (Ziehe et al., 2004), uwedge (Tichavsky et Yeredor, 2009), jedi (Souloumiac, 2009) et qdiag (Vollgraf et Obermayer, 2006).

- Comon, P. (1994). Independent component analysis: a new concept?. *Signal Process.*, 36(1994):287–314.
- Pham, D.-T. (2001). Blind separation of instantaneous mixture of sources via the gaussian mutual information criterion. *Signal Process.*, 81(4):855–870.
- Pham, D.-T. et Cardoso, J.-F. (2001). Blind separation of instantaneous mixtures of nonstationary sources. *IEEE Trans. Signal Process.*, 49(9):1837–1848.
- Souloumiac, A. (2009). Non-orthogonal joint diagonalization by combining Givens and hyperbolic rotations. *IEEE Trans. Signal Process.*, 57(6):2222–2231.
- Tichavsky, P. et Yeredor, A. (2009). Fast approximate joint diagonalization incorporating weight matrices. *IEEE Trans. Signal Process.*, 57(3):878–891.
- Vollgraf, R. et Obermayer, K. (2006). Quadratic optimization for simultaneous matrix diagonalization. *IEEE Trans. Signal Process.*, 54(9):3270–3278.
- Ziehe, A., Laskov, P., Nolte, G. et Müller, K.-R. (2004). A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind source separation. *J. Mach. Learn. Res.*, 5(Jul):777–800.

Annexe C

NOTE SUR R ET LA PARALLÉLISATION DES TÂCHES (À MOINDRE COÛT)



L'aspect pratique de la complexité des tâches engendrées par les méthodes mises en œuvre est souvent négligé. Les considérations présentées ici visent à mettre en exergue trois niveaux principaux d'optimisation afin de rendre les calculs moins longs à exécuter :

- Utilisation de bibliothèques optimisées pour le calcul algébrique ;
- Parallélisation pour l'utilisation de tous les cœurs d'une même machine ;
- Parallélisation multi-machines et multi-processeurs.

Les premier et second niveaux d'optimisation peuvent se recouvrir car certaines bibliothèques de calcul algébrique sont intrinsèquement optimisées pour utiliser la totalité des cœurs disponibles sur une même machine. Ces versions sont dites « multi-thread » par opposition aux versions « mono-thread » qui ne génèrent qu'un seul flux d'opérations à la fois. Il existe enfin des manières de considérer ces trois niveaux d'optimisation à la fois en utilisant par exemple des versions de la bibliothèque MKL (Math Kernel Library) d'Intel basée sur une implémentation de MPI (Message Passing Interface), mais nous ne détaillerons pas ce point. Les procédures techniques permettant de mettre en place les différents cadres proposés ici ne seront pas décrites, nous choisirons seulement quelques exemples pour avoir une idée des ordres de grandeur que peuvent faire gagner ces différentes procédures. Notons enfin que ces méthodes sont des moyens très simples (du point de vue du travail supplémentaire à fournir) d'obtenir des gains relativement importants. Il est évidemment possible de parvenir à des gains plus conséquents mais le coût nécessaire pour réaliser les optimisations s'avère alors prohibitif (e.g. nous ne considérerons pas les optimisations du code assembleur généré par les compilateurs).

Les méthodes abordées dans cette note sont relatives au langage R¹ mais peuvent facilement être étendues à d'autres langage de script du même type. La première partie ne concerne néanmoins pas Matlab qui utilise une bibliothèque de calcul optimisée par défaut.

1 Utilisation d'une bibliothèque de calcul algébrique optimisée

En utilisant de manière efficace les architectures des processeurs actuels, on peut diminuer de manière substantielle le temps d'exécution d'un algorithme. Les techniques permettant de parvenir à un tel résultat relèvent la plupart du temps de l'optimisation au

¹ Logiciel Open-Source initialement dédié aux calculs statistiques qui peut être téléchargé sur <http://www.r-project.org/>

niveau assembleur. Si de telles optimisations sont très fastidieuses et nécessitent une excellente connaissance de l'architecture du processeur, il est néanmoins possible d'en profiter à moindre coût dans le cas de l'algèbre linéaire. Dans les années 1970, la librairie BLAS a été développée en FORTRAN dans le but de fournir aux utilisateurs les briques de base de l'algèbre linéaire nécessaires à beaucoup d'algorithmes (opérations scalaire-vecteur, matrice-vecteur et matrice-matrice de base). Depuis ces années, les squelettes de base de l'ensemble des fonctions de cette bibliothèque ont été conservés, permettant à n'importe quelle compagnie de développer son propre code interchangeable avec le code initial dans la mesure où le squelette de chaque fonction est respecté. Parmi les implémentations qui ont vu le jour depuis, on peut citer ATLAS (Automatically Tuned Linear Algebra Software) qui s'adapte à chaque architecture de processeur en compilant et testant, à l'installation, les versions optimisées les plus adéquates ; MKL (Math Kernel Library) est développée par INTEL pour les processeurs de cette compagnie (fournie sous licence gratuite pour un usage non-commercial) ; ACML (AMD Core Math Library) développée par les ingénieurs d'AMD pour leurs propres processeurs ; ou encore GotoBLAS. Considérons alors l'exemple de code R suivant et observons les différences entre une version de R liée à la bibliothèque BLAS fournie par R et une version de R liée à la bibliothèque d'INTEL.

```
R
## multiplication matricielle en grande dimension
do_mult <- fonction(N) {
  x <- matrix(rnorm(N^2), N, N)
  system.time(x %*% x)
}
```

L'exemple précédent est bien sûr choisi à dessein car la dimension de la matrice est trop importante pour qu'elle puisse être stockée entièrement dans le cache de niveau L1. Les optimisations deviennent donc primordiales dans ce cas de figure. Les deux courbes correspondant aux implémentations les moins performantes de la [figure C.1](#) correspondent respectivement aux versions de base et avec la bibliothèque MKL.

Cette figure montre à quel point deux implémentations concurrentes d'une même fonction de base peuvent engendrer des différences énormes. Ainsi, pour une taille de matrice de 4000×4000 , la configuration de base réalise l'opération en 283 s alors que la configuration utilisant MKL réalise cette même opération en 13 secondes.

2 Utilisation de tous les cœurs d'une même machine

Supposons à présent que nous disposions d'une machine dotée de plusieurs unités de traitement (quatre à huit cœurs sont relativement courant aujourd'hui), il est possible de mettre à profit l'ensemble de ces unités encore une fois sans beaucoup d'effort. Avec R, deux méthodes très simples permettent d'y parvenir.

La première consiste, comme souvent, à mettre à profit le travail existant fait par d'autres. Ainsi, comme nous l'avons déjà sous-entendu, il existe une version de la bibliothèque de calcul algébrique MKL d'Intel qui permet d'utiliser sans effort les calculs en parallèle sur plusieurs cœurs d'une même machine. Bien sûr l'optimisation se restreint ici au cas d'un ensemble d'algorithmes relativement réduit mais sur des opérations telles que des multiplications de matrices, on peut observer des gains très intéressants. La [figure C.1](#) montre les performances de la bibliothèque en version multithread ($N = 4$) par rapport aux deux méthodes présentées jusqu'alors. Dans ce cas précis, on observe un facteur proche de quatre entre la version mono et la version multithread, le temps mis pour effectuer une

Temps d'exécution d'une multiplication de matrices

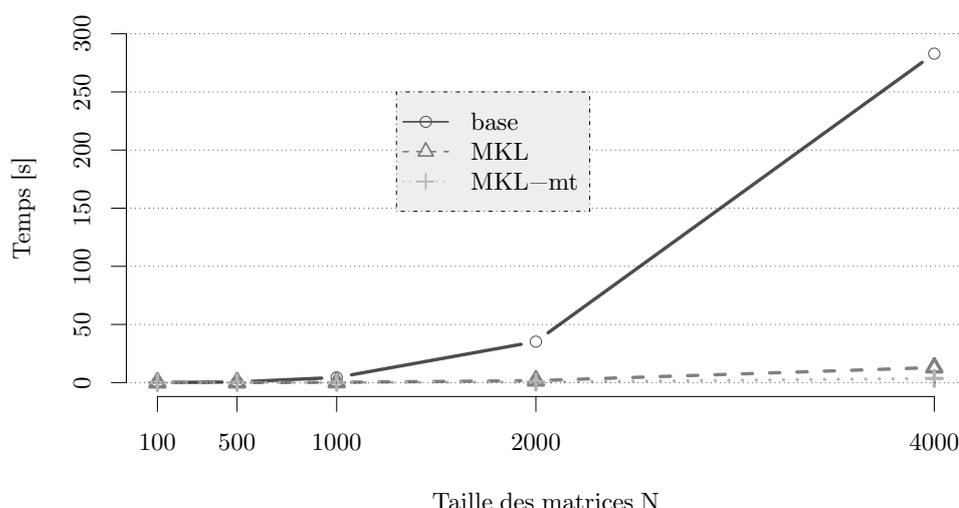


Figure C.1 Temps de calcul pour une multiplication matricielle en fonction de la taille des matrices. Les trois configurations comparées sont : le logiciel R avec la bibliothèque BLAS interne, R avec la version monothread de la bibliothèque MKL et enfin R avec la version multithread de la bibliothèque MKL. Ces tests ont été réalisés sur une machine disposant de quatre cœurs Intel XEON cadencés à 2,66 GHz. La taille du cache mémoire est de 6 MO. Toutes les simulations ont été réalisées sur la même machine mais les deux premières configurations (base et MKL) n'utilisent qu'un cœur alors que la dernière (MKL-mt) met à profit les quatre cœurs.

multiplication matricielle avec des matrices de taille 4000×4000 est alors réduit à environ $3,5 \text{ s}^1$.

Comment faire alors pour mettre à profit les différentes unités de calcul d'une même machine dans des cas de traitements plus complexes ? Nous ne traiterons ici qu'un exemple de cas le plus simple, *i.e.* qu'on veut effectuer différents traitements indépendants. Il existe plusieurs manières avec R de réaliser ceci sans trop de difficultés, nous présentons ici l'une d'entre elles basée sur une extension nommée `multicore`. Prenons l'exemple typique en EEG : un banc de filtres qui doit réaliser le filtrage d'un nombre important de capteurs (typiquement entre 20 et 256). Chaque filtre doit agir sur chaque capteur de manière complètement indépendante. Considérons l'exemple de code suivant permettant de réaliser cette opération de manière séquentielle.

¹ Pour information, Matlab (The MathWorks™) permet d'obtenir des performances similaires à R dans le cas où R utilise MKL (version monothread). Ceci est normal puisque Matlab utilise également cette même bibliothèque. Il suffit de cocher une case dans l'interface de Matlab pour autoriser le multithread et ainsi obtenir les mêmes performances que dans le cas de R utilisant MKL-multithread.

R

```

## filtrage d'un ensemble de signaux par un banc de filtres
#   x les données à filtrer
#   Fs la fréquence d'échantillonnage
#   lowFreqs l'ensemble des fréquences de coupure
basses
#   highFreqs l'ensemble des fréquences de coupure
hautes
#   Lfilt la longueur des filtres de Butterworth
mvfiltfilt <- function(x,Fs,Lfilt=5,lowFreqs=c(5,5),
  highFreqs=c(35,8)) {
  nFreqs <- length(lowFreqs)
  nChans <- dim(x)[2]
  # création des filtres
  myfilter <- lapply(1:nFreqs,function(i) butter(
    Lfilt,c(lowFreqs[i]/(Fs/2),highFreqs[i]/(Fs/2)),type="pass"))
  idx <- 0:(nFreqs*nChans-1)
  aux <- lapply(idx,function(i) filtfilt(
    myfilter[[1+(i %/% (nChans))]],
    x[,1+(i %/% (nChans))]))
  res <- array(unlist(aux), dim=c(dim(x),nFreqs))
  return(res)
}

```

Cette version séquentielle filtre chaque colonne de \mathbf{x} par l'ensemble des filtres du banc de filtres et fournit en résultat un tableau multidimensionnel dont les deux premières dimensions font la même taille que \mathbf{x} et dont la dernière dimension comporte autant d'entrées que le nombre de filtres dans le banc de filtres.

En modifiant ce programme afin qu'il utilise toutes les capacités d'une machine à quatre cœurs par exemple, on peut considérer la version parallèle qui suit. La fonction `lapply`, qui s'apparentait à une simple boucle `for` dans la version séquentielle a été remplacée par la fonction `mcapply`. Cette dernière est une version multicœurs d'une boucle `for` qui distribue les tâches entre les unités de calcul disponibles.

R

```

## filtrage d'un ensemble de signaux par un banc de filtres
mvfiltfilt_par <- function(x,Fs,Lfilt=5,lowFreqs=c(5,5),
  highFreqs=c(35,8)) {
  nFreqs <- length(lowFreqs)
  nChans <- dim(x)[2]
  myfilter <- lapply(1:nFreqs,function(i) butter(
    Lfilt,c(lowFreqs[i]/(Fs/2),highFreqs[i]/(Fs/2)),type="pass"))
  idx <- 0:(nFreqs*nChans-1)
  aux <- mcapply(idx,function(i) filtfilt(
    myfilter[[1+(i %/% (nChans))]],
    x[,1+(i %/% (nChans))]),mc.cores=4)
  res <- array(unlist(aux), dim=c(dim(x),nFreqs))
  return(res)
}

```

On teste alors ces deux approches avec des données de taille relativement importante (800000 points temporels \times 12 capteurs) mais courante dans les analyses en ICM. On

obtient alors les performances du [tableau C.1](#). Notons qu'on n'obtient pas un facteur quatre entre les deux versions en raison du temps mis pour rassembler les données entre les quatre processus dans le cas multi-cœurs.

	1 cœur	4 cœurs
Temps de calcul [s]	126	64
Mémoire RAM maximum utilisée [Goctets]	2,5	15

Tableau C.1 Performances en termes de temps de calcul et de mémoire RAM utilisée par deux versions différentes d'un même traitement. La version 1 cœur correspond à l'application de la fonction `mvfiltfilt` tandis que la versions 4 cœurs correspond à `mvfiltfilt_par`. Dans cette simulation, on a filtré des données multivariées de taille 800000×12 et la taille du banc de filtres était de 14. Les processeurs étaient les mêmes dans les deux cas (Intel XEON 5160 cadencés à 3 GHz).

Si l'utilisation des quatre cœurs disponibles sur la machine permet bien de réduire considérablement le temps de calcul, on voit que la contrepartie de cette procédure est qu'une instance de R est clonée pour chaque unité de traitement. Par conséquent, si on utilise initialement 2,5 Giga Octets de RAM pour un seul processus, R va ajouter quatre instances et ainsi passer à une utilisation de mémoire insoutenable assez rapidement quand le nombre de capteurs augmente.

3 Parallélisation multi-machines et multi-processeurs

Afin de remédier aux limites imposées par la mémoire RAM, il faut alors considérer la maîtrise plus fine des données partagées entre le processus maître et les processus esclaves. Or dans le cas précédent, on peut observer que les processus esclaves ne doivent traiter les données que d'un seul capteur à la fois. On peut donc facilement améliorer l'utilisation de la mémoire vive en transmettant aux processus esclaves, non pas les données, mais uniquement une fonction qui, sur chaque processus esclave, chargera uniquement les données correspondant au capteur qui doit être traité. Afin de rendre cette méthode relativement efficace, il est important que les données soient enregistrées au format binaire sur un espace disque accessible à tous les processus esclaves. Le chargement des données se fait alors rapidement et il est facile de restreindre le chargement à un capteur seulement en effectuant des déplacements dans le fichier binaire.

Il existe une bibliothèque sous R permettant de gérer simplement les transferts de données entre les processus maître et esclaves, il s'agit de `Rmpi`¹. Supposons alors que l'on dispose d'un ensemble de machines, le programme suivant permet de mettre à profit chaque processeur de chacune d'entre elles.

¹ `Rmpi` implémente une interface simple entre la bibliothèque pour langage C nommée MPI (Message Passing Interface) et R.

R

```

## filtrage d'un ensemble de signaux par un banc de filtres
# myfun charge les données d'un capteur
mvfiltfilt_mpi <- fonction(myfun,Fs,chans,Lfilt=5,lowFreqs=c(5,5),
  highFreqs=c(35,8)) {
  nFreqs <- length(lowFreqs)
  nChans <- dim(x)[2]
  myfilter <- lapply(1:nFreqs,function(i) butter(
    Lfilt,c(lowFreqs[i]/(Fs/2),highFreqs[i]/(Fs/2)),type="pass"))
  # envoi des variables utiles aux processus esclaves
  mpi.bcast.Robj2slave(myfun)
  mpi.bcast.Robj2slave(myfilter)
  mpi.bcast.Robj2slave(nChans)
  mpi.bcast.Robj2slave(chans)
  idx <- 0:(nFreqs*nChans-1)
  aux <- mpi.applyLB(idx,function(i) filtfilt(
    myfilter[[1+(i %/% (nChans))]],
    myfun(chans[1+(i %/% (nChans))]]))
  res <- array(unlist(aux), dim=c(dim(x),nFreqs))
  return(res)
}

```

Notons que les performances obtenues (tableau C.2) correspondent à un facteur 4 gagné entre les deux versions. Ce facteur relativement faible s'explique par deux raisons principales : d'une part le traitement considéré ici implique de lourds échanges de données lors du retour du résultat par les processus esclaves, ceci s'avère très coûteux dans le cas de l'utilisation de multiples machines et dépend de la qualité du réseau qui les connecte ; d'autre part, les 22 cœurs utilisés représentent un cluster hétérogène et chacun d'entre eux est loin de pouvoir fournir les mêmes performances que le cœur utilisé dans le cas de la fonction `mvfiltfilt`.

	1 cœur	22 cœurs
Temps de calcul [s]	230	80
Mémoire RAM maximum utilisée [Goctets]	4	<1

Tableau C.2 Performances en termes de temps de calcul et de mémoire RAM utilisée par deux versions différentes d'un même traitement. La version 1 cœur correspond à l'application de la fonction `mvfiltfilt` tandis que la versions 22 cœurs correspond à `mvfiltfilt_mpi`. Dans cette simulation, on a filtré des données multivariées de taille 800000×22 et la taille du banc de filtres était de 14.

4 Conclusion

Les considérations pratiques abordées ici sont loin d'être anecdotiques en raison de la lourdeur des traitements qu'impliquent les interfaces cerveau-machines. Il n'est pas rare que certains d'entre eux nécessitent plusieurs jours de calcul dans le cas de multiples validations croisées. La mise en place de méthodes relativement simples pour diminuer ce temps de calcul s'avère donc souvent primordiale si plusieurs valeurs de paramètres doivent être testées.

PUBLICATION DE L'AUTEUR



Je profite de cette dernière partie pour remercier la Délégation Générale pour l'Armement (DGA) qui a financé cette thèse et a donc permis aux différentes publications qui suivent de voir le jour.

Les travaux présentés dans cette thèse ont fait l'objet d'articles dans des revues internationales et ont été présentés lors de plusieurs congrès nationaux et internationaux (avec comités de lecture et publications des actes).

Les travaux du [chapitre III](#) (réalisés en collaboration avec le Prof. Pfurtscheller et son équipe) ont donné lieu à des présentations lors du congrès EUSIPCO 08 et du BCI Workshop 08. De plus, un article de revue a été accepté dans *IEEE Transactions on Biomedical Engineering*. Un article plus orienté sur les aspects neurophysiologiques relatifs à ce chapitre ont été publiés dans *Clinical Neurophysiology* (M. Congedo est le premier auteur).

La méthode présentée dans le [chapitre IV](#) a pour l'instant été présentée lors du congrès international ICA 09.

Les travaux des [chapitres V](#) et [VI](#) ont été présentés lors de ESANN 09.

Enfin, certains travaux relatifs à l'utilisation des graphes de connectivité pour l'analyse des signaux cérébraux ont été réalisés en collaboration avec différents membres du laboratoire. Une partie (relative à l'application pour les interfaces cerveau-machines) a été présentée lors des congrès EMBC 07 et GRETSI 07. Deux d'entre eux, plus théoriques, seront présentés en septembre prochains (ECCS 09 et GRETSI 09).

— Publication dans des journaux internationaux —

Congedo, M., Gouy-Pailler, C. et Jutten, C. (2008a). On the blind source separation of human electroencephalogram by approximate joint diagonalization of second order statistics. *Clin. Neurophysiol.*, 119(12):2677–2686.

Gouy-Pailler, C., Congedo, M., Brunner, C., Jutten, C. et Pfurtscheller, G. (2009). Non-stationary brain source separation for multi-class motor imagery. *IEEE Trans. Biomed. Eng.*. Accepté pour future publication.

— Publications dans des conférences avec actes et comités de lecture —

Amblard, P.-O., Achard, S., Lemoine, J. et Gouy-Pailler, C. (2009). Complex networks and connectivity, application to fMRI. Dans *European Conference on Complex Systems (ECCS 09)*.

Congedo, M., Jutten, C., Sameni, R. et Gouy-Pailler, C. (2008b). A new general weighted least-squares algorithm for approximate joint diagonalization. Dans *Proceedings of the 4th International BCI Workshop*.

Gouy-Pailler, C., Achard, S., Rivet, B., Jutten, C. et Maby, E. et al. (2007). Topographical dynamics of brain connections for the design of asynchronous brain-computer interfaces. Dans *Proc. Int. Conf. IEEE Engineering in Medicine and Biology Society (IEEE EMBC 07)*, pages 2520–2523, Lyon, France.

Gouy-Pailler, C., Congedo, M., Jutten, C., Brunner, C. et Pfurtscheller, G. (2008). Model-based source separation for multi-class motor imagery. Dans *Proceedings of the 16th European Signal Processing Conference (EUSIPCO-2008)*, EURASIP.

- Gouy-Pailler, C., Mattout, J., Congedo, M. et Jutten, C. (2009). Uncued brain-computer interfaces: a variational hidden markov model of mental state dynamics. Dans *Proceedings of the 17th European Symposium on Artificial Neural Networks (ESANN 09)*.
- Gouy-Pailler, C., Rivet, B., Achard, S., Souloumiac, A. et Jutten, C. et al. (2007). Théorie des graphes et dynamique des connexions cérébrales pour la conception d'interfaces cerveau-machines asynchrones. Dans *XXI ème colloque GRETSI*.
- Gouy-Pailler, C., Sameni, R., Congedo, M. et Jutten, C. (2009). Iterative subspace decomposition for ocular artifact removal from eeg recordings. Dans *Proceedings of the 8th International Conference on Independent Component Analysis and Blind Source Separation (ICA 09)*, pages 419–426, Paraty, Brasil.
- Gouy-Pailler, C., Congedo, M., Brunner, C., Jutten, C. et Pfurtscheller, G. (2008). Multi-class independent common spatial patterns: Exploiting energy variations of brain sources. Dans *Proceedings of the 4th International BCI Workshop*.
- Lemoine, J., Gouy-Pailler, C., Achard, S. et Amblard, P.-O. (2009). Recherche de la connectivité de réseaux complexes. application en fmri. Dans *XXII ème colloque GRETSI*.

Les Interfaces Cerveau-Machines (ICMs) visent à établir un moyen de communication direct entre le cerveau humain et un système électronique. Si d'importantes avancées ont d'ores et déjà été réalisées depuis une vingtaine d'années dans le domaine, les systèmes existants s'avèrent lents et fastidieux, imposant aux utilisateurs le moment auquel ils doivent envoyer leur commande (systèmes synchrones). Nos travaux se sont concentrés sur le développement de systèmes asynchrones basés sur l'électroencéphalographie (EEG) : ce n'est pas l'ordinateur qui impose le moment de réalisation de la tâche mentale mais le sujet qui a la possibilité d'envoyer sa commande lorsque qu'il le désire. Ceci impose que le système soit capable d'analyser en temps réel l'activité cérébrale du sujet et de reconnaître la tâche mentale recherchée parmi les activités cérébrales mesurées. Nos travaux proposent des méthodes pour les problématiques suivantes :

- permettre à l'utilisateur d'envoyer sa commande lorsqu'il le désire ;
- augmenter les capacités du système à reconnaître correctement une tâche mentale ;
- rendre le système robuste à certains changements qui peuvent s'opérer lors de l'utilisation du système, i.e., adapter le système aux possibles changements du sujet dans la manière dont celui-ci accomplit sa tâche mentale.

Les méthodes proposées utilisent de manière conjointe les caractéristiques spatiales (zones cérébrales impliquées), fréquentielles (bandes de fréquences) et temporelles (succession des activations) des tâches mentales afin de pallier le faible rapport signal sur interférences. Différents jeux de données sont utilisés afin de valider ces approches.

Brain-Computer Interfaces (BCIs) aim at establishing a direct communication pathway between the human brain and an electronic device. In the past twenty years, many breakthroughs have been made in the field, but current systems remain slow and difficult to use. They indeed often constrain users to send their command at precise moments (synchronous systems). Our research work is devoted to asynchronous systems based on electroencephalography (EEG): subjects are allowed to send their commands whenever they want. This entails brain activity of the subject to be analyzed in real time to detect the task-related brain activity among the whole brain activity. Our research work focuses on three main issues raised by asynchronous BCIs:

- to allow users to send their command whenever they want;
- to improve the ability of the system to correctly recognize a particular mental task;
- to increase the robustness of the system when coping with possible variations in the way the subject performs mental tasks.

The proposed methods aim at jointly using the spatial (cerebral areas implied), frequential (frequency bands involved) and temporal (succession of activations) features of the different mental tasks in order to compensate the low signal to interference ratio. The approaches are evaluated and compared to state-of-the-art results using datasets from the successive BCI Competitions.