

Responsable de stage (personne à contacter) : Cédric Gouy-Pailler, cedric.gouy-pailler@cea.fr

Laboratoire : CEA LIST, Laboratoire d'Outils pour l'Analyse de Données (LOAD), Saclay (91)

I Contexte

Les systèmes basés sur l'utilisation directe des signaux cérébraux pour le contrôle de périphériques électroniques (ordinateurs, neuroprothèses) s'appellent des *Interfaces Cerveau Machines* (ICM). Leur but initial était de fournir aux personnes souffrant de graves handicaps moteurs un moyen de communication ou de contrôle. Leurs applications couvrent aujourd'hui autant le domaine biomédical (réhabilitation, moyen de communication ou restauration de mouvement) que le jeu vidéo (casque Emotiv pour le grand public notamment).

La conception usuelle d'ICM comporte plusieurs étapes. On utilise généralement des systèmes d'enregistrement appelés « électroencéphalographie » (EEG) (réseau de capteurs de potentiels électriques) pour recueillir les signaux électriques provoqués par l'activation de certains neurones du cerveau. Ceux-ci sont ensuite traités puis analysés afin d'identifier parmi l'ensemble de l'activité cérébrale certains motifs connus. Ces motifs d'activation recherchés correspondent à des tâches mentales ou des états mentaux largement connus d'un point de vue neurophysiologique et donc relativement facilement identifiables parmi le flot de données recueilli.

On sait par exemple que l'imagination de mouvements de certaines parties du corps engendrent des signaux cérébraux spatialement distinguables (cortex contralatéral au membre). Néanmoins, la faible amplitude de ces signaux par rapport à l'ensemble de l'activité cérébrale nécessite l'utilisation de *filtres spatiaux* (Gouy-Pailler et al., 2010 et Rivet et al., 2009a) et de *classifieurs* spécifiques à chacun des sujets. Ils doivent être entraînés lors d'une phase initiale pendant laquelle on indique au sujet la tâche à réaliser. Cette étape a donc pour but la construction d'une base de données d'apprentissage pour la détermination des paramètres des algorithmes automatiques utilisés pour la reconnaissance lors de la phase d'utilisation du système.

La conception des ICM constitue un problème difficile en raison de plusieurs facteurs. Tout d'abord les enregistrements EEG sont souvent contaminés par des bruits divers (artefacts oculaires ou musculaires, bruits électromagnétiques) qui perturbent l'interprétation des signaux. De plus, les signaux EEG relatifs aux tâches considérées sont généralement noyés parmi l'ensemble de l'activité cérébrale enregistrée. Enfin le fait de travailler avec des sujets humains rajoute trois difficultés majeures : 1) il est impossible de s'assurer que le sujet respecte les consignes données, 2) l'acquisition de données coûte cher, *i.e.* les données d'apprentissage sont souvent réduites, 3) de nombreuses non-stationnarités peuvent exister, notamment celles dues au dysfonctionnement brutal d'une des électrodes (souvent en raison d'un mauvais contact avec la peau). Ces différentes difficultés font du traitement du signal un aspect primordial de la conception d'ICM. Parmi les étapes de traitement du signal couramment utilisées, le filtrage spatial est généralement conçu pour augmenter le rapport signal sur bruit. D'un côté, la qualité du filtrage augmente lorsque le nombre de capteurs est important, mais à l'opposé, la facilité d'installation du système ainsi que la probabilité de mauvais contact plaide en faveur d'un faible nombre de capteurs.

Le laboratoire LOAD au sein du CEA LIST

Situé à Saclay, en Île-de-France sud, le CEA LIST est un centre de recherche technologique dédié aux systèmes dans lesquels l'aspect logiciel a une place prépondérante. Les équipes du CEA LIST sont partenaires de nombreux laboratoires universitaires, de grandes écoles et d'autres organismes de recherche au travers de multiples projets de recherche collaboratifs. Elles font notamment partie intégrante de Digiteo, campus scientifique réunissant plus de 1200 chercheurs spécialistes des technologies de l'information (CEA, INRIA, CNRS, Supélec, École Polytechnique, Université Paris-Sud XI).

Au sein du CEA LIST, le Laboratoire d'Outils pour l'Analyse de Données (LOAD) regroupe des experts aux compétences variées et complémentaires, allant du traitement du signal et des images à l'apprentissage artificiel. Au fil des années, le laboratoire est notamment devenu spécialiste de la décomposition de signaux mono- et multi-dimensionnels. En utilisant par exemple des critères d'indépendance ou de parcimonie, de nombreux outils ont été proposés pour le débruitage, la déconvolution, l'identification de signaux faibles ou encore la détection précoce de motifs temporels dans les données. La volonté du laboratoire de proposer des solutions innovantes opérationnelles s'est également traduit par des associations à plusieurs organismes publics et privés, qui se concrétisent par des conventions communes (laboratoire commun, équipe de recherche commune, convention de partenariat), ou des projets collaboratifs.

Le LOAD développe des algorithmes de traitement du signal ou des images pour des applications diverses : biomédicales (spectrométrie, imagerie médicale, pharmacologie, interfaces cerveau-machines), sécurité (sismique, détection de molécules en phase gaz) ou encore industrielles (analyse de mouvements bi- ou tri-dimensionnels, par exemple pour l'écriture ou l'interaction gestuelle personne-système).

II Objectif du stage

Le stage se focalisera sur la conception de filtres spatiaux linéaires dans le contexte d'ICM basées sur deux classes distinctes (*e.g.* imagination d'un mouvement de la main gauche ou de la main droite). Ces méthodes consistent à choisir un ensemble de poids à appliquer aux électrodes afin de concentrer l'information qui nous intéresse sur un petit nombre de composantes. En

d'autres termes, elles permettent d'isoler les signaux cérébraux relatifs aux tâches sur un petit nombre de composantes à partir d'un grand nombre de signaux enregistrant l'activité générale du cerveau. Cette *réduction de dimensions* a pour but de faciliter l'apprentissage du classifieur en lui indiquant l'information pertinente à exploiter dans les données. En fonction de l'intérêt et des compétences du (de la) candidat(e), le stage pourra s'orienter selon l'une des deux pistes suivantes :

- les méthodes actuelles permettent de trouver les filtres spatiaux en utilisant un critère d'optimisation du rapport signal sur bruit. Cette approche permet des calculs quasi-instantanés mais l'objectif d'optimiser le rapport signal sur bruit peut ne pas être synonyme d'une bonne classification. Le (la) candidat(e) aura donc pour but de proposer des fonctions de coût capables de rendre directement compte du pouvoir de classification des filtres spatiaux. La détermination des filtres spatiaux se fera donc par l'optimisation de cette fonction de coût (optimisation dans un espace de variables continues). Les deux aspects de ce stage, d'une part proposer la bonne fonction de coût en comprenant les objectifs liés aux ICM, d'autre part son optimisation, seront d'importances comparables.
- Les recherches actuelles dans le domaine tentent de réduire au maximum le nombre de capteurs nécessaire à la conception des ICM. Cependant la plupart des approches séparent (ou alternent) l'étape de calcul des filtres spatiaux avec l'étape de sélection des meilleurs capteurs. Le (La) candidat(e) aura ici pour but de proposer des méthodes capable de réaliser les deux étapes conjointement à l'aide de décompositions matricielles pénalisées.

III Approche

En fonction de l'orientation choisie, le (la) candidat(e) pourra s'intéresser, dans le premier cas, aux méthodes d'optimisation (convexes et non-convexes) en grande dimension ; dans le deuxième cas, le (la) candidat(e) pourra s'inspirer, dans un premier temps, des travaux récents utilisant des approches itératives sur le nombre de capteurs (Rivet et al., 2009b), ils se concentrera ensuite sur l'adaptation des critères utilisés en ICM à l'aide des approches proposées dans Witten et al., 2009.

L'évaluation des méthodes se fera sur des données contrôlées ainsi que sur des données réelles.

IV Profil recherché

Stagiaire master recherche ou école d'ingénieur avec une formation solide en traitement du signal et/ou informatique. Connaissances d'un langage de programmation à base de script (R, matlab, python).

V Pistes bibliographiques

- Gouy-Pailler, C., Congedo, M., Brunner, C., Jutten, C. et Pfurtscheller, G. (2010). Non-stationary brain source separation for multi-class motor imagery. *IEEE Trans. Biomed. Eng.*, 57(2):469–478.
- Rivet, B., Souloumiac, A., Attina, V. et Gibert, G. (2009a). xDAWN algorithm to enhance evoked potentials: Application to brain computer interface. *IEEE Trans. Biomed. Eng.*, 56(8):2035–2043.
- Rivet, B., Souloumiac, A., Gibert, G., Attina, V. et Bertrand, O. (2009b). Sensors selection for P300 speller brain computer interface. Dans *Proc. European Symposium on Artificial Neural Networks (ESANN09)*.
- Witten, D. M., Tibshirani, R. et Hastie, T. (2009). A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis.. *Biostatistics*, 10(3):515–534.