



HAL
open science

Détection de la fatigue mentale à partir de données électrophysiologiques

François Laurent

► **To cite this version:**

François Laurent. Détection de la fatigue mentale à partir de données électrophysiologiques. Neurosciences [q-bio.NC]. Université Pierre et Marie Curie - Paris VI, 2010. Français. NNT: . tel-00634776

HAL Id: tel-00634776

<https://theses.hal.science/tel-00634776>

Submitted on 23 Oct 2011

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.

**THESE DE DOCTORAT DE
L'UNIVERSITE PIERRE ET MARIE CURIE**

Spécialité

Neurosciences
(Cerveau, Cognition, Comportement)

Présentée par

M. François LAURENT

Pour obtenir le grade de

DOCTEUR de l'UNIVERSITÉ PIERRE ET MARIE CURIE

Sujet de la thèse :

Détection de la fatigue mentale à partir de données électrophysiologiques
soutenue le 6 janvier 2010

devant le jury composé de :

Mme Geneviève FLORENCE Directrice de thèse

M. Jacques MARTINERIE Directeur de thèse

M. Régis MOLLARD Rapporteur

M. Jean-Philippe LACHAUX Rapporteur

M. Alain MUZET Examineur

M. Stéphane CHARPIER Examineur

M. Didier BAZALGETTE Invité

Résumé

Trois expériences ont été réalisées pour évaluer la faisabilité d'un détecteur de fatigue mentale qui exploiterait des signaux électrophysiologiques pour discriminer des fenêtres de temps associées à un état « fatigué » de fenêtres associées à un état « pas fatigué ».

Dans la première expérience, les sujets réalisaient une tâche de *task switching* pendant environ 2 h. Deux segments de 20 min ont été retenus pour chaque sujet, le premier a été associé à un état « fatigué », l'autre à un état « pas fatigué », puis ces segments ont été découpés en fenêtres de temps glissantes sur lesquelles les signaux électrophysiologiques (EOG, ECG et EEG) ont été quantifiés. L'usage d'un SVM linéaire (appris sujet par sujet) pour la classification des fenêtres de temps a permis plusieurs observations : le taux de bonnes classifications augmentaient avec la taille des fenêtres de temps (de 4 à 30 s) mais pas assez pour que l'ITR augmentât lui aussi avec cette taille des fenêtres ; les quantifications de l'EOG se plaçaient derrière celles de l'ECG, elles-mêmes supplantées par les amplitudes moyennes de l'EEG qui elles aussi étaient dépassées par les PLV de l'EEG en termes de taux de bonnes classifications avec (pour les PLV) des taux approchant les 100 % ; l'examen des poids SVM permet de constater que les quantifications d'EOG et d'ECG contribuaient à la classification lorsque toutes les quantifications des différentes modalités étaient utilisées ensemble, en utilisant des fenêtres de temps de 30 s ; les amplitudes moyennes de l'EEG augmentaient en 7-13 Hz et diminuaient en 13-18 Hz ; en 3-7 Hz apparaissait une augmentation en FCz évoquant la négativité d'erreur.

Dans une deuxième expérience avec une tâche de compatibilité spatiale, nous avons discriminé des fenêtres de 20 s entre un niveau de bonnes performances (temps moyen de réaction courts) et un niveau de moins bonnes performances, en quantifiant ces fenêtres à l'aide des amplitudes moyennes, des PLV ou des cohérences de l'EEG, au niveau du scalp et au niveau de sources reconstruites sur le cortex cérébral. Nous nous attendions à des meilleurs taux de bonnes classifications avec les sources reconstruites qu'avec les données de scalp, quelque soit la quantification, mais ce ne fût pas le cas. Nous avons également réalisé une simulation reprenant l'essentiel des analyses réalisées précédemment et en provoquant deux phénomènes, l'un à travers l'amplitude de signaux utiles à la classification, l'autre jouant sur le couplage entre deux signaux utiles. Ces simulations ont montré d'une part la complémentarité des mesures de synchronie locale, telles que l'amplitude moyenne, avec les mesures de synchronie à distance telles que la cohérence ou la PLV, et d'autre part l'apport de la reconstruction de sources en termes de taux de bonnes classifications et de localisation des motifs discriminants. Cette incompatibilité entre les résultats sur données simulées et ceux sur les données EEG réelles est interprétée ici comme la conséquence de la non-prise en compte, dans le processus de reconstruction, de sources extra-corticales comme les sources cérébrales sous-corticales et les sources extra-craniennes (yeux, muscles) qui pourraient avoir eu une contribution.

La troisième étude portait sur une troisième expérience en contexte écologique cette fois-ci, impliquant les élèves-pilotes dans un simulateur de vol. La simulation durait près de 3 h et les signaux EEG ainsi que trois paramètres de vol (cap, altitude, vitesse) ont été enregistrés. Nous avons proposé une mesure comportementale quantifiant la variabilité des paramètres de vol, par fenêtres de 10 s, pour suivre la performance des pilotes avec une résolution temporelle adaptée à nos analyses. Nous avons construit deux classes, à savoir un groupe de fenêtres de temps de 10 s associé aux bonnes performances (variabilité faible) et un second groupe de fenêtres associé aux moins bonnes performances. Plutôt que d'utiliser une procédure de validation croisée, nous avons évalué les classifieurs appris sur des données postérieures aux données d'apprentissage, de manière à évaluer aussi la capacité de la procédure de classification à déterminer la tendance générale des performances du pilote. Les taux de bonnes classifications se maintenaient aux environs de 70 % tandis que la procédure permettait effectivement de déterminer si les performances du pilotes étaient plutôt à la hausse, à la basse ou stables, ce pour 12 sujets sur 13. La classification ne se détériorait pas avec le temps écoulé entre la fin des données utilisées pour l'apprentissage et le début des données utilisées pour l'évaluation du classifieur appris.

Les corrélats physiologiques exhibés ne correspondaient pas d'une expérience à une autre ce qui laissait penser que soit les différentes manières de construire les classes ne permettaient pas d'isoler un même différentiel de fatigue mentale, soit la fatigue mentale telle qu'observée à travers les techniques électrophysiologiques employées dépendait de la tâche et/ou des sujets. En revanche, la faisabilité d'une discrimination d'états mentaux proches de réalités opérationnelles a été établie.

Table des matières

0.1.Contexte.....	5
0.2.Remerciements.....	5
1 – Introduction.....	9
1.1.Sujet.....	9
1.2.Objectif.....	9
1.3.Structure du document.....	11
2 – La fatigue mentale.....	12
2.1.Le concept de fatigue.....	12
2.2.Les trois aspects de la fatigue.....	13
2.3.Classification des états de fatigue.....	16
2.4.Techniques expérimentales pour l'étude de la fatigue mentale.....	20
3 – La physiologie.....	23
3.1.Pourquoi l'électrophysiologie.....	23
3.2.Principes.....	24
3.3.Fatigue mentale et système nerveux central.....	27
3.4.Quantifications.....	29
4 – Analyse de données.....	33
4.1.Vue générale de la chaîne des traitements.....	33
4.2.Classes et classification.....	35
4.3.Validation et mesures de pouvoir discriminant.....	40
4.4.Support d'interprétation.....	43
5 – Axes de recherche.....	45
5.1.L'approche générale.....	45
5.2.Objectifs des études.....	46
5.3.Matériel, prétraitements et quantifications.....	48
5.4.Classification, validation et interprétation.....	54
5.5.Résumé : classification multi-modalités.....	57
5.6.Résumé : reconstruction de sources.....	58
5.7.Résumé : simulation de vol.....	60
6 – Étude 1 : classification multi-modalités.....	63
6.1.L'expérience.....	63

6.2.Déroulement.....	64
6.3.Données comportementales et subjectives.....	64
6.4.Données physiologiques – Classification.....	67
6.5.Contribution des variables à la classification.....	69
6.6.Discussion.....	72
7 – Étude 2 : reconstruction de sources.....	79
7.1.Introduction.....	79
7.2.La simulation.....	80
7.3.Résultats de la simulation.....	85
7.4.L'expérience réelle.....	89
7.5.Analyse du comportement.....	90
7.6.Analyse des données EEG.....	91
7.7.Discussion.....	91
8 – Étude 3 : simulation de vol.....	95
8.1.Introduction.....	95
8.2.Méthodes.....	98
8.3.Résultats.....	105
8.4.Discussion.....	108
9 – Conclusion et perspectives.....	113
9.1.Retour sur les principaux résultats.....	113
9.2.Quelles données physiologiques ?.....	113
9.3.La fatigue mentale comme objet d'étude.....	116
9.4.Le potentiel de l'analyse de données par apprentissage.....	118
10 – Références.....	119

0.1. Contexte

Cette thèse s'est déroulée de octobre 2006 à décembre 2009 au Centre de Recherche de l'Institut du Cerveau et de la Moelle épinière (CRICM), à l'hôpital de la Pitié Salpêtrière à Paris. Ce laboratoire s'appelait initialement Laboratoire de Neurosciences Cognitives et Imagerie Cérébrale, et répondait également au nom de LENA. Les projets de recherche développés dans ce laboratoire concernent donc le fonctionnement du cerveau, principalement humain, aux échelles d'observation que permettent d'atteindre les techniques d'imagerie cérébrale actuelles. Depuis de nombreuses années, au laboratoire, ces techniques ont été l'électroencéphalographie (EEG) et la magnétoencéphalographie (MEG). Plus récemment, l'imagerie par résonance magnétique nucléaire (IRM) a fait son entrée dans nos locaux. Cette thèse est restée fidèle à l'une des deux techniques de prédilection du laboratoire : l'EEG.

Le CRICM est une unité du Centre National de la Recherche Scientifique (CNRS) et de l'Institut National de la Santé et de la Recherche Médicale (INSERM). C'est donc un laboratoire civil. Ce travail a été réalisé aussi avec la participation d'un second laboratoire, le récent Institut de Recherche Biomédicale des Armées (IRBA), qui, comme son nom l'indique, est d'affiliation militaire. Ce laboratoire a également changé de nom au cours de la thèse, et était connu auparavant comme l'Institut de Médecine Aérospatiale du Service de Santé des Armées (IMASSA). Il est implanté en région parisienne, sur la base aérienne de Brétigny-sur-Orge (Essonne).

Ce cadre en partie militaire s'explique par une collaboration de longue date entre les deux directeurs de cette thèse, Geneviève Florence (Vétérinaire en chef, à l'IRBA) et Jacques Martinerie (Ingénieur de recherche hors classe, au CRICM), et par le rôle de la Délégation Générale pour l'Armement (DGA). Cette dernière a financé cette thèse en accordant une bourse sur 3 ans dont j'ai pu bénéficier.

0.2. Remerciements

Avant d'aller plus loin, il est temps pour moi de présenter mes plus sincères remerciements à un certain nombre de personnes. En effet, la thèse est une aventure enrichissante à bien des égards, et on soupçonne mal, avant de s'y engager, les trésors que peuvent offrir des personnes dont les parcours intellectuel et personnel sont très différents les uns des autres. C'est aussi, pour le thésard, une étape dans le parcours académique, c'est-à-dire un étage de plus dans l'ascension intellectuelle que représente l'éducation. Certes on ne s'arrête jamais d'apprendre, mais à bien y regarder, la thèse représente une transition importante, accompagnée d'inévitables moments forts. Les hommes et les femmes qui participent à l'aventure de la thèse en gravitant dans l'environnement du thésard, marquent bien souvent ce dernier au fer rouge. Puisque cette section est dédiée à des personnes dont l'influence s'entend ici par rapport à moi, je suis amené à parler de moi aussi. Ces remerciements prennent donc par moment des allures d'histoire personnelle.

Quelques années avant de débiter la thèse, j'avais réalisé que je ne pensais pas en mots. Les concepts se laissaient manipuler dans mon esprit sans que la cartographie du langage n'intervinsse. J'avais à cette époque décidé de travailler sur ce point, pour de multiples raisons. Au début de la thèse, j'avais encore de grandes difficultés à improviser des phrases bien formées dans le temps de la réflexion nécessaire à chacune de ces phrases. Les doctorants et post-doctorants alors présents dans l'équipe, à savoir Frédérique Amor, Aurélie Campagne, Michel Besserve et Benoît Crépon, avaient chacun une vivacité d'esprit et une habileté à manier les mots que j'admirais. Ils ont représenté des modèles pour moi ; je leur dois donc beaucoup, car j'ai grandement évolué grâce à leur exemple et cette évolution est une de mes plus importantes réalisations personnelles du temps de la thèse.

Puisque je parlais du début de la thèse, j'en profite pour exprimer mes remerciements aux personnes, pour une part anonyme, qui m'ont donné une chance, celle de réaliser cette thèse. Geneviève Florence et Jacques Martinerie sont de ceux là, évidemment. Peut-être suis-je trop enclin à me dévaluer, toujours est-il que la chance a visiblement joué en ma faveur, car ni mon dossier, ni mes compétences réelles au début de la thèse, ne me faisaient arriver à la cheville des étudiants vétérans de l'équipe. Je le sentais aussi dans leur culture générale, dans les souvenirs de lycée et de classe préparatoire bien plus frais dans leurs esprits que dans le mien alors que le temps écoulé était en leur défaveur. Je crois qu'à l'évolution que représente naturellement la thèse, s'est ajouté une sorte de coup d'accélérateur dont j'ai grandement bénéficié.

L'aventure humaine ne s'est pas limitée à ces quelques personnes. Je me dois d'attribuer une mention particulière à Mario Valderrama, le seul doctorant de ma « génération », la personne qui aura le plus solidement ancré son image dans mon esprit, mon compagnon de thèse. D'autres personnes, au passage moins long voire beaucoup plus bref, m'ont laissé aussi un patrimoine inattendu. Je pense par exemple aux personnes à l'aide de qui j'ai franchi mes appréhensions à parler dans une langue étrangère : Stavros Nikolopoulos pour l'anglais et Fabrizio De Vico Fallani pour l'italien.

Enfin, viennent mes collègues de travail. Il y a eu Miguel Valencia et Mario Chavez qui, par ailleurs, ont beaucoup contribué à l'ambiance joviale dans l'équipe. Il y a eu Mathias Guillard, de l'IRBA. Mathias était impliqué dans les deux expériences auxquelles j'ai participé, et à ce sujet, je remercie également Florence Bouchet. Il y a eu aussi, comme collaboratrice, Line Garnero, notre défunte directrice de laboratoire. Je lui rends ainsi un très bref hommage car ma thèse, et d'autant plus la période de la rédaction, ont été marquées par quelques tristes événements auxquels je ne souhaite pas attribuer plus de place dans ces pages.

Il me reste à témoigner ma gratitude envers mes encadrants qui, loin d'avoir eu à m'évaluer à la manière d'enseignants (je ne pense pas aller à l'encontre d'une quelconque éthique en les remerciant ici), ont construit le cadre dans lequel ma progression scientifique a pu se faire. Geneviève avait la rigueur nécessaire pour me cadrer, justement, sur plus d'un plan. Je parlais de l'importance qu'avait

pour moi l'habileté à s'exprimer. Alors que mon usage du français évoluait vers plus de complexité, Geneviève a par exemple veillé à ce que je n'aie pas trop loin dans cette complexité, tandis que Jacques avait d'excellents conseils en matière de présentation orale. Je retiens de Jacques principalement sa technique managériale. En effet, il s'est montré directif pendant la première année de ma thèse, puis il s'est présenté plutôt comme un conseiller avisé lors de la deuxième année et enfin, en troisième année, il cantonnait son rôle à celui d'interlocuteur attentif. Il y a peu, Jacques a quitté sa veste de chef d'équipe. Les anciens de l'équipe s'étaient manifestés en grand nombre pour exprimer leurs appréciations très positives de l'exercice de Jacques en cette qualité. Je pense être bien placé à présent pour pouvoir confirmer tout le bien qui a en été dit et écrit.

J'ajoute un grand « merci » en direction des personnes que je n'ai pas citées, en particulier les autres membres passés ou présents de ce qui était l'équipe, celles aussi qui m'ont soutenu dans les bas, « professionnels » ou personnel, que j'ai pu traverser.

1 – Introduction

1.1. *Sujet*

Cette thèse a pour objet la détection en temps-réel de la fatigue mentale à partir de mesures électrophysiologiques.

La physiologie est l'étude du fonctionnement des organes. Les organes concernés par cette thèse sont le cerveau, mais aussi les yeux et le cœur. Pour observer l'activité de ces organes, nous avons appliqué des électrodes sur la peau des participants à nos expériences. Cela nous a permis de suivre l'activité électrique liée au fonctionnement de ces organes, pour ensuite chercher des manifestations de la fatigue dans cette activité.

La fatigue mentale, de part l'origine du financement qui a permis la réalisation de cette thèse, est celle des pilotes d'aéronefs dans l'aviation militaire. Cependant, la fatigue des pilotes d'aéronefs ayant très probablement des points communs avec la fatigue rencontrée dans un certain nombre d'autres métiers et activités, nous nous sommes par conséquent intéressés à la fatigue mentale dans un entendement très général.

Enfin, l'aspect « détection en temps réel » sous-entend ici que des outils d'analyse d'une certaine catégorie ont été employés pour mener la recherche des manifestations de la fatigue dans l'activité électrophysiologique. C'est le chaînon qui relie fatigue mentale et électrophysiologie.

Ainsi, trois domaines de compétence ont été nécessaires pour réaliser ce travail : l'électrophysiologie et la quantification des activités électriques enregistrées, la fatigue mentale et la conception d'expérience la faisant intervenir et l'analyse de données.

1.2. *Objectif*

L'objectif premier de tels travaux est de prévenir les conséquences les plus dramatiques de la fatigue mentale. En effet, la fatigue est par exemple connue pour avoir eu une contribution dans les catastrophes industrielles de Three Mile Island, Tchernobyl et Bhopal, mais aussi dans des drames tels que ceux du pétrolier Exxon Valdez et de nombre d'avions de transport qui se sont écrasés (Akerstedt, 1995). La fatigue est aussi reconnue comme une cause importante d'accidents de la route dans de nombreux pays (Akerstedt, 2000). De même, le milieu médical et en particulier hospitalier est sujet à des incidents fréquents, des erreurs réalisées par le personnel et imputables à la fatigue (Mion & Ricouard, 2007).

Pour prévenir ces conséquences de la fatigue, il faut pouvoir détecter cette dernière.

La détection d'un état mental implique la recherche en continu (d'où l'aspect « temps-réel ») de marqueurs de cet état dans l'activité électrophysiologique mesurée. Les principaux efforts scientifiques déployés jusque là se sont concentrés légitimement sur des marqueurs communs à une majorité d'individus par exemple.

Le contexte de détection permet une orientation différente de la démarche employée pour étudier la fatigue, par rapport aux approches classiques de description de quelques marqueurs principaux.

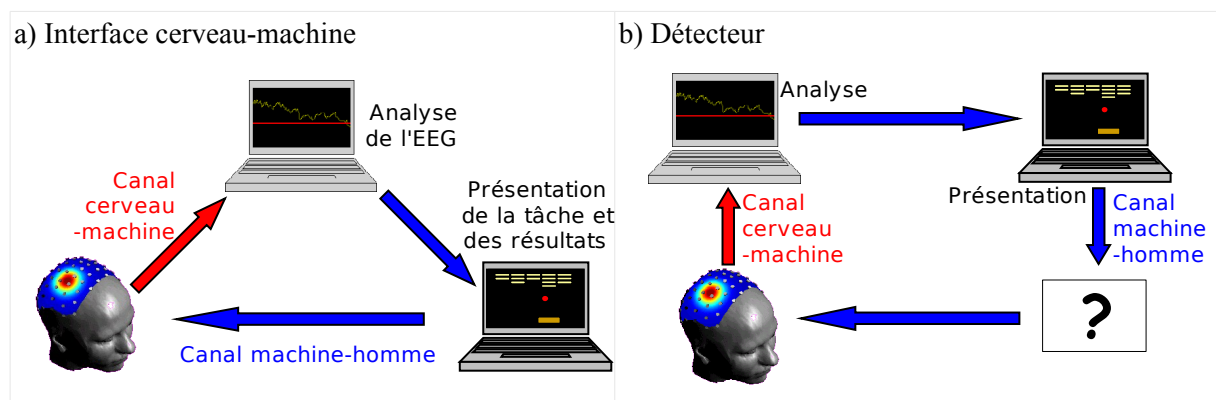


Fig. 1 – Principe d'une interface cerveau-machine (a). Un sujet réalise une tâche par l'intermédiaire d'une interface homme-machine classique, éventuellement à sens unique (machine-homme). Parallèlement, l'activité cérébrale du sujet est enregistrée et analysée en corrélant les motifs d'activité cérébrale à des réponses du sujet à la tâche, le système ayant une connaissance au moins partielle de la tâche. La tâche est ensuite adaptée à l'information utile que le système pense avoir extraite.

b) Dans un contexte de détection, la rétroaction qui consiste à modifier la tâche dynamiquement n'est pas nécessairement employée, d'autant plus que la tâche est imposée et qu'elle est prioritaire sur le fonctionnement du détecteur. Dans la pratique, un décideur s'intercalera entre la sortie du système de détection et le sujet.

En effet, à l'instar des Interfaces Cerveau-Machine (ICM, lire Wolpaw *et al.* (2002) pour une introduction), on considère un système dans lequel l'individu, dont on souhaite observer la fatigue, est placé au centre. L'outil de détection opère sur cet individu seul ; par conséquent, les marqueurs de la fatigue peuvent être propres à cet individu. De plus, l'objectif principal est celui de la détection, et non plus de la description.

Cela implique en revanche l'emploi d'outils d'analyse automatique. L'objectif n'est pas nécessairement de trouver un marqueur parmi des événements précis, localisés dans l'espace voire dans le temps, mais plutôt d'appliquer une procédure qui produit un détecteur de fatigue avec une fiabilité que l'on peut estimer avant la mise en production. Cette généralisation de l'approche présente deux avantages.

Elle permet en premier lieu d'éviter la relative pauvreté d'un critère de détection trop simple. En effet, on peut se permettre de rechercher des motifs parmi un grand nombre de variables car il n'est plus nécessaire de décrire ces motifs en détail. Cela est rendu possible par l'emploi d'un algorithme qui recherche ces motifs, puis les exploite. Ensuite, le problème de la variabilité inter-individuelle est contourné.

En outre, bien que l'objectif principal n'est pas porté sur la description des marqueurs de la fatigue, les algorithmes de détection s'appuient sur de tels marqueurs, qu'ils construisent, et offrent donc chacun leur propre représentation de ces marqueurs. On peut donc dépouiller et interpréter les résultats à la « lumière » du fonctionnement de l'algorithme. Par exemple, on peut chercher la partie de ces marqueurs commune à tous les individus qui ont participé à l'expérience.

Ainsi, bien que l'intérêt premier de la démarche de détection soit applicatif, on peut l'exploiter à des fins scientifiques, c'est-à-dire avec une finalité de description du mécanisme d'intérêt.

Une telle approche amène de nouvelles difficultés méthodologiques et cette thèse n'a pu faire l'économie d'un développement tant en discours qu'en expérimentations au sujet de l'usage adéquat de certaines techniques d'analyse.

1.3. Structure du document

Trois études sont décrites dans cette thèse. Une première étude porte sur l'intérêt de l'association d'informations apportées par trois techniques d'enregistrement électrophysiologiques différentes. Une deuxième étude évalue l'apport de la reconstruction des sources corticales de l'activité électroencéphalographique. Enfin, une troisième étude explore des données acquises en environnement « réel », à savoir une simulation de vol.

Les trois parties qui suivent cette introduction, successivement **La fatigue mentale** (chapitre 2), **L'électrophysiologie** (chapitre 3) et **L'analyse de données** (chapitre 4), présente un état de l'art des domaines de compétence cités.

Le chapitre 5 énonce les hypothèses de travail en les faisant suivre de points de méthode communs aux trois études détaillées ensuite au travers des chapitres 6 (**Classification multi-modalités**), 7 (**Reconstruction de sources**) et 8 (**Simulation de vol**). Les chapitres relatifs aux études sont très développés et représentent un volume de lecture important, si bien que le chapitre 5 donne également un résumé de ces études.

Enfin le chapitre 9 est celui de la conclusion. Sont présentées quelques-unes des perspectives offertes à une éventuelle poursuite des travaux.

2 – La fatigue mentale

2.1. Le concept de fatigue

Le terme « fatigue » est issu du langage courant et fait référence à un état subjectif. Il accompagne l'individu tout au long de sa vie, dès le plus jeune âge, par exemple lorsque l'enfant découvre l'impérialité du nyctémère (le cycle jour-nuit) sur son rythme de vie, jusqu'au troisième âge, où le corps et l'esprit s'affaiblissent, en passant par la majeure partie de la vie où une activité scolaire, professionnelle ou relative au ménage (tâches domestiques mais aussi, pour un parent, élever ses enfants, etc) est nécessaire à l'intégration sociale.

La notion de fatigue est de plus un concept vaste, qui par conséquent est susceptible de revêtir différentes réalités, selon le vécu et l'environnement social de chaque individu.

En effet, comme toute notion subjective relative à un état interne, la notion de fatigue se construit au gré des corrélations avec d'éventuels « symptômes » plus aisément identifiables, et est donc partiellement dépendante de l'activité.

Ainsi, chez certaines personnes dont les rythmes circadiens sont peu stables (quelle qu'en soit la raison) ou décalés (travail posté par exemple) ou celles dont l'activité professionnelle est souvent monotone (citons la conduite routière), la notion de fatigue mentale se confond avec celle d'hypovigilance, et la fatigue est réduite par l'expérience à un état précurseur du sommeil.

Chez d'autres personnes, par exemple celles pour qui la fatigue est plus souvent physique que mentale, l'hypovigilance est un concept bien distinct de la fatigue mentale car alors, la fatigue, qu'elle soit physique ou mentale, est en général la conséquence d'un excès d'activité, d'un effort relativement intense, et moins, à l'inverse, d'une activité prolongée à faible charge de travail, ou sujette à un contexte circadien peu favorable.

La fatigue est également souvent perçue comme un état anormal. Prenons l'exemple vécu d'une personne qui répond à la question coutumière « ça va, pas trop fatigué ? » par « non, non ; je m'endors un peu mais je ne suis pas fatigué(e) ». Il y a en fait deux sujets de réflexion ici.

En premier lieu, on lit dans la réponse que la fatigue est perçue comme différente du début d'hypovigilance relaté par l'affirmation « je m'endors un peu », puisque la seconde personne déclare dans le même temps : « je ne suis pas fatigué ». Dans la situation précise d'où est extrait cet échange, la seconde personne effectuait en fait une tâche de pilotage. Or, le fait de s'endormir représente un problème dans ce genre de tâches. On remarque également que la première personne, celle qui a posé la question, a mis en opposition l'état « ça va » avec l'état « fatigué ». Cet état fatigué a donc été présenté d'emblée comme un état anormal. La seconde personne, souhaitant probablement se montrer rassurante sur son état, a implicitement répondu « ça va » ou du moins « ça va encore, à ce stade ».

L'état fatigué, comme repris par la deuxième personne, a probablement été pensé, cette fois-ci, comme un état plus anormal, plus extrême, que celui de « [s'endormir] un peu », c'est-à-dire peut-être un stade plus avancé de ce qui pourrait être un même état, qualitativement parlant.

Sur cet exemple, on peut légitimement se demander quel est le poids de l'implicite dans le langage usuel, lorsqu'on parle de fatigue. Les langues, dont les cultures associées sont à fort contexte historique, comme le français (de France), sont connues pour faire un usage très conséquent de l'implicite. D'autres cultures, comme celle des États-Unis, font porter la préférence sur une explicitation plus systématique de l'information échangée, même lors d'une simple conversation. Toutes les notions subjectives sont donc susceptibles de plus ou moins « souffrir » de biais propres au contexte de l'échange langagier qui permet à un individu de rapporter son niveau de fatigue. De plus, l'effort de recherche scientifique portant sur la fatigue mentale a été rapporté jusque là dans la langue anglaise, et celle-ci offre un vocabulaire que l'on peine parfois à transposer au français.

Face à la multiplicité des situations et sensations associées à la fatigue, toute étude sur le sujet semble donc condamnée à restreindre son champ d'application, et à choisir éventuellement un autre terme pour éviter la confusion. Le terme « fatigue mentale » a en revanche l'avantage de souligner l'origine cérébrale de ce phénomène, mieux que d'autres termes qui apparaissent plus tard dans ce texte comme celui de « performance ». Il laisse mieux imaginer l'intérêt des techniques d'imagerie cérébrale pour l'étude de cet état.

Faisons donc un rapide tour des définitions de la fatigue dans la littérature scientifique.

2.2. Les trois aspects de la fatigue

a) Trois quarts de siècle auparavant...

Dès 1937, A. Bills présentait la fatigue sous trois angles différents qui restent pleinement pertinents encore de nos jours. Citons quelques extraits de son essai intitulé *Fatigue in mental work* (travail mental et fatigue) :

« [The term fatigue] can refer to at least three entirely different, and, to a certain extent, independent things. By derivation it means “weariness” or “feelings of exhaustion”; in other words, the subjective experience of the worker. But as used by the physiologist, it usually means that change in the condition of the cells or organs which have undergone excessive activity resulting in a loss of power. And from the point of view of output or product, it means that decrement in the quantity or quality of the product which results from the continuity of the work. We might call these 1, the subjective; 2, the organic, and 3, the product definitions of fatigue. [...] Subjective fatigue is measured by the change in feeling-tone described as loss of readiness for work, or increasing desire to terminate work. [...] Organic fatigue, on the other

hand, is measured in terms of quantitative change in metabolism. [...] Objective fatigue, or decrement in output, [...] » (Bills, 1937)

[Le terme fatigue] peut correspondre à au moins trois choses complètement différentes et, dans une certaine mesure, indépendantes. Par dérivation, il évoque une « usure » ou des « sensations d'épuisement », en d'autres termes l'expérience subjective de l'individu. Mais, de l'usage des physiologistes, il désigne en général un changement dans la condition de cellules ou organes qui présentent une activité excessive, ce qui entraîne une baisse de leurs capacités dans leurs fonctions. Et du point de vue du travail produit, ou « sortie », il désigne une diminution, en quantité ou en qualité, du produit qui résulte de la continuité du travail. On désignera chacune de ces définitions de la fatigue par les qualificatifs suivants : fatigues 1) subjective, 2) organique et 3) retranscrite dans le travail produit. [...] La fatigue subjective est mesurée par le changement, dans le ressenti de l'individu, décrit comme une diminution de la disponibilité vis-à-vis du travail, ou un désir croissant de terminer ce travail. [...] La fatigue organique, d'un autre côté, est mesurée en termes de changements quantitatifs du métabolisme. [...] La fatigue objective, ou dégradation du travail produit, [...]

Reprenons les points évoqués ci-dessus. La fatigue dite objective (3) regroupe les effets de la fatigue sur les actions réalisées par l'individu lors d'une tâche. La dernière phrase de l'extrait cité a été éludée car l'auteur entre dans des détails assez précis alors qu'ici, il s'agit de simplement retenir que ce type de fatigue se qualifie ou se quantifie par l'observation du comportement de l'individu. Ce terme (« comportement ») revient donc fréquemment dans ce manuscrit. Plus généralement, on peut assimiler ce versant de la fatigue à la perte progressive de fonction qui vient en conséquence d'une activité normale prolongée ou répétée.

À l'usage, on conçoit un protocole expérimental dans lequel les individus participants réalisent une tâche avec des objectifs fixés. La fatigue objective peut alors être définie comme certains changements dans une grandeur liée à ces objectifs. Les variations de cette grandeur ont en général l'avantage d'être aisément mesurables et interprétables. On peut par exemple choisir des objectifs exprimés directement par des critères quantitatifs de performance, tels que le maintien de temps de réponse courts et de taux d'erreurs faibles. On considère ensuite qu'il y a fatigue mentale au delà d'une valeur seuil à définir pour chacune de ces variables.

La fatigue organique ou physiologique (2), quant à elle, rassemble toutes les manifestations dans le corps humain qui sont liées à la fatigue. Si le comportement ou le ressenti d'un individu peut changer avec la fatigue, et puisque l'organisme est le siège du comportement comme de l'expérience subjective, alors ces changements doivent nécessairement se retrouver dans l'activité des organes. À la tâche du physiologiste de trouver et de décoder les manifestations de la fatigue dans l'organisme.

Enfin, la fatigue subjective (1) représente peut-être la définition la plus légitime de la fatigue mentale. Un individu sera considéré comme fatigué s'il se sent fatigué.

Autorisons-nous une mauvaise caricature : le décideur s'intéressera à la performance ou à la prévention des risques, et donc à la fatigue objective ; le scientifique voudra expliquer les mécanismes de la fatigue mentale et à cette fin, se penchera sur une description physiologique de la fatigue ; enfin, l'individu, celui qui est susceptible d'éprouver de la fatigue, retiendra la (sa propre) définition subjective de la fatigue. Alors quelle approche privilégier ?

b) Quelques définitions empruntées à la littérature contemporaine

Avant de discuter à nouveau des objectifs généraux de cette thèse et certains des choix faits quant à son orientation, présentons les définitions les plus reprises dans la littérature scientifique contemporaine. L'auteur de la définition suivante s'intéresse dans ses recherches à la fatigue au volant. Il décrit la fatigue mentale comme :

« [...] *a subjectively experienced disinclination to continue performing the task at hand.* »
(Brown, 1994)

[...] un désengagement de la tâche, lorsque celle-ci doit être prolongée jusqu'à son terme, accompagné d'une sensation subjective caractéristique.

On retrouve l'expression « répulsion mentale à l'égard du travail » dans le dictionnaire de l'Office québécois de la langue française, et lorsque l'on consulte le dictionnaire Merriam Webster sur le sens du mot « *disinclination* », on trouve en réponse « *slight aversion* » (légère aversion). Les termes « réticence » et « lassitude » reflètent cette même part de la fatigue telle que définie par Brown.

La seconde définition présentée ici a été publiée dans un journal médical :

« *The awareness of a decreased capacity for physical and/or mental activity due to an imbalance in the availability, utilization, and/or restoration of resources needed to perform activity.* » (Aronson *et al.*, 1999)

La perception consciente d'une diminution de la capacité à mener une activité physique ou mentale, due à un déséquilibre entre la disponibilité, l'utilisation, et/ou le renouvellement des ressources nécessaires à cette activité.

Les deux définitions citées mettent au premier plan l'expérience subjective. En effet, la première définition place cet aspect (« *subjectively experienced* ») en première position dans la phrase et le mot « *disinclination* », dont la fatigue serait une forme, représente en lui-même une certaine charge émotionnelle. La seconde, construite comme une vraie définition où le terme parent plus général vient en premier, commence donc par « *awareness* ». Cependant, les deux définitions mentionnent une

activité ou une tâche. Il n'y a qu'un pas pour penser aux effets de la fatigue sur cette activité et par dérivation à la fatigue objective. Enfin, dans la deuxième définition, l'auteur a choisi d'explicitier le rôle de ressources. La formulation fait penser en premier lieu aux ressources énergétiques, mais peut-être peut-on y inclure d'autres types de ressources (psychologique, etc). Toujours est-il qu'une mention est faite au mécanisme de la fatigue et donc à la physiologie.

On remarque également que la fatigue est présentée comme un état sous-optimal ou un phénomène associé à un tel état, avec des mots comme « *disinclination* » et « *decreased* ».

Bien que nous aurons à reprendre ces dernières définitions, la communauté scientifique semble encore insatisfaite par les propositions de définition émises à ce jour. Le fond du débat se résume à une problématique de classification des états de fatigue et à la recherche d'un cœur commun à tous les types de fatigue que l'on peut recenser. Certains plaident en faveur d'une unique fatigue globale, d'autres favorisent une caractérisation dite multidimensionnelle de la fatigue (Shen *et al.*, 2006). Pour mieux comprendre la complexité de cette tâche, la section suivante s'attache à inventorier les différents types de fatigue.

2.3. Classification des états de fatigue

a) Au delà du monde du vivant

La paysage sémantique du terme « fatigue » est ponctué de dichotomies. On trouve par exemple, en élargissant le champ de recherche, une dichotomie entre le monde du vivant et celui de l'industrie. Dans l'industrie en effet, le terme « fatigue » est employé et prend même de très nombreux usages, allant des fatigues mécanique, thermique et à la corrosion, à la fatigue d'un gisement, en passant par la fatigue d'une substance luminescente. On retrouve dans les définitions de ces diverses fatigues la notion de perte progressive de fonction, en qualité ou en quantité. Un fort lien de parenté ressort avec le concept d'usure.

Quelques points communs avec la fatigue du monde du vivant sont toutefois à évoqués. On retrouve les critères de sous optimalité de l'état fatigué, de l'installation progressive de cet état, et de l'origine liée à une action répétée ou soutenue. En revanche, l'objet de la fatigue n'est pas la même, ce qui a pour conséquence évidente qu'on ne trouve pas d'équivalent de la fatigue subjective dans ce domaine. De plus, un caractère irréversible de la fatigue du monde de l'industrie, bien que pas systématiquement cité dans les nombreuses définitions, semble être prépondérant, et c'est là une différence majeure avec le monde du vivant qui a une dynamique bien plus complexe.

b) Au delà du mental

Si l'on considère à présent la fatigue chez l'Homme, on trouve de nouvelles dichotomies dans la littérature, comme par exemple la mise en opposition entre fatigue mentale et fatigue physique. La fatigue physique fait elle-même l'objet d'une dichotomie, avec la fatigue périphérique qui touche essentiellement le muscle squelettique et la fatigue centrale qui affecte le système nerveux central suite à l'activité physique (Gandevia, 2001). La fatigue physique, à l'échelle d'un muscle, est une conséquence de contractions répétées ou soutenues, avec un épuisement des ressources énergétiques et une accumulation de résidus du métabolisme. À l'opposé, la fatigue mentale est supposée siéger dans le système nerveux central.

Bien que cette dichotomie doive très probablement son origine aux différents types de professions et de loisirs dans la société, on voit là déjà une contribution de la physiologie à cette classification de la fatigue. En effet, bien que la sensation subjective de fatigue soit en partie un processus cognitif et donc mental, elle représente une sorte de signal d'alarme dont un organe est à l'origine. Si l'organe fatigué est un muscle, on parle de fatigue physique ; si cet organe est le système nerveux central, on parle de fatigue mentale.

Cette classification présente en revanche des limites. Par exemple, la fatigue centrale, associée à l'activité physique, est responsable du fait qu'un sportif doit déployer une certaine volonté pour pouvoir mener à bien un exercice.

Des liens de cause à effet ont été mis en évidence entre la fatigue mentale et la fatigue physique, la première induisant une augmentation de la « fatigabilité » lors d'un exercice physique (Marcora *et al.*, 2009), la seconde favorisant le sommeil et la perte de motivation après l'exercice (Meeusen *et al.*, 2006) ou affectant les performances cognitives immédiatement après l'exercice (Lorist *et al.*, 2002).

En outre, à cette dissociation entre fatigue physique et fatigue mentale, on pourrait ajouter la fatigue sensorielle, ou fatigue des organes sensoriels, même si là encore, il n'est pas toujours évident de définir quelles sont les parts respectives d'un organe sensoriel et du cortex sensoriel primaire associé, dans la cartographie de cette fatigue.

c) Spécifique ou non à la tâche

En ce qui concerne la fatigue mentale liée à la réalisation d'une tâche, on peut imaginer plusieurs fatigues.

En premier lieu, on peut par exemple penser qu'il existe dans le système nerveux central une fatigue dont le mécanisme serait assez similaire à celui de la fatigue physique, locale à la chaîne de traitements de l'information recrutée pour la réalisation de la tâche, faisant suite à une activité excessive à l'intérieur de cette chaîne, dont le siège serait par exemple les cellules cérébrales. On ne fait cependant qu'imaginer une telle fatigue car, malgré les nombreux mécanismes candidats, les

études de l'activité cérébrale à l'échelle de la cellule ne se font que sur certains animaux auxquels on peine à faire faire une tâche cognitive jusqu'à l'épuisement. En revanche, un autre mécanisme de fatigue mentale mieux étudié est l'hypovigilance ou plus généralement la transition de l'état de veille à celui de sommeil. Ce deuxième type de fatigue mentale ne serait pas ou peu spécifique de la tâche.

Enfin, la sensation de fatigue est un troisième phénomène à considérer en tant que tel. En effet, c'est une sensation qui apparaît avec la perception d'une moindre capacité à mener une tâche à bien. Son intérêt réside probablement dans l'inhibition des comportements trop dépensiers par rapport au contexte, ce qui la fait intervenir dans de nombreuses situations différentes, aussi bien sous l'action de rythmes circadiens que suite à une activité longue et intense, physique ou mentale. Boksem et Tops ont proposé une théorie de la mesure subconsciente de la dépense énergétique que requerrait la réalisation réussie d'une tâche, par rapport au gain attendu (Boksem & Tops, 2008). Il s'agirait d'une composante subconsciente du système de prise de décisions. Si la mesure penche en faveur d'une inaction, alors que l'individu maintient sa décision consciente de continuer la tâche, la motivation chute et la lassitude s'installe.

D'une certaine manière, on est tenté d'étendre la fatigue subjective au concept de fatigue cognitive, le qualificatif « cognitif » étant propre ici¹ aux processus mentaux de « haut niveau ». En effet, la fatigue cognitive se rapporte à un épuisement de « ressources cognitives » plutôt qu'énergétiques, par exemple motivationnelles, attentionnelles et d'autres encore, si l'on peut parler de ressources.

Les mécanismes cognitifs de la fatigue, comme celui décrit par Boksem et Tops, pourraient être impliqués dans des situations très variées, y compris la fatigue physique ou celle liée à la douleur, comme dans le cas des troubles musculo-squelettiques par exemple (Fishbain *et al.*, 2004). Il y a sensation subjective de fatigue dès lors que l'inconfort est modérée et donc tolérée, mais en même temps limitant pour l'individu, ce dernier devant fournir un effort, si bien qu'il favorise une lassitude, un désengagement de la tâche propre à la fatigue.

d) Vers une neuropsychologie de la fatigue mentale ?

Pour la fatigue mentale, une autre dichotomie concerne le caractère normal ou pathologique de cette fatigue, c'est-à-dire encore son caractère aigu ou chronique. De nombreuses maladies sont connues pour entraîner une fatigue, telles que la plupart des infections respiratoires (Bartels, 2009), les troubles du sommeil, les cancers, la polyarthrite rhumatoïde, ainsi que les maladies neuroinflammatoires

¹ La cognition est considérée ici comme l'ensemble des processus mentaux dont les fonctions sont dites de haut niveau, telles que la conscience, l'attention, la mémoire, le langage, etc. En cela, les sciences cognitives – terme quelque peu « fourre-tout » pour certains – n'ont rien d'une super-catégorie des neurosciences, les neurosciences traitant de la fonction des cellules nerveuses, seules ou en communauté. La cognition n'englobe pas par exemple la totalité des fonctions nécessaires aux activités sensorielle et motrice. Certaines cartographies du cortex cérébral, basées sur celle de Brodmann, dissocient même trois catégories de fonctions (et de régions) : sensorielles, perceptuo-gnosiques et cognitives.

(sclérose en plaques) et neurodégénératives (Alzheimer, Parkinson) (lire Chaudhuri & Behan, 2000). On parle également de syndrome de fatigue chronique lorsque la fatigue est un symptôme chronique, qu'elle dure depuis au moins six mois, sans que ne soient décelés d'autres symptômes laissant penser à une maladie mieux connue. De manière générale, la fatigue mentale de cause anormale se traduit essentiellement par des plaintes du patient, si bien que les milieux médicaux semblent supporter majoritairement l'approche subjective de la fatigue.

De l'autre côté, les activités fatigantes qualifiées de normales sont par exemple des tâches de conduite routière, de pilotage, de supervision, ou tout travail de nuit. En ce qui concerne cette fatigue mentale normale, les facteurs inducteurs sont la période de la journée (ou de la nuit), la durée écoulée depuis la fin du précédent sommeil, la durée du précédent sommeil, la durée depuis laquelle la tâche a commencé, la monotonie de la tâche, la chaleur, le bruit, la luminosité ambiante (quand elle est faible), le confort ou l'inconfort, et probablement d'autres encore.

e) Concepts voisins

La fatigue mentale se traduit par des erreurs de jugement, de perception et de mémorisation, et un ralentissement des réactions.

Pour mieux se représenter les tenants et aboutissants de la fatigue mentale en tant que variable, on peut considérer certains concepts voisins.

L'un d'eux est la vigilance, c'est-à-dire le niveau d'éveil, ou niveau d'activation générale. On décele l'hypovigilance (un niveau bas de vigilance) par des signes comportementaux de somnolence (paupières lourdes, tête dodelinante) ou par le biais de l'électroencéphalographie. L'hypovigilance est en général accompagnée d'une sensation de fatigue, donc on peut considérer que l'hypovigilance est un mécanisme de fatigue mentale, mais on doit garder à l'esprit que la fatigue mentale ne se réduit pas à l'hypovigilance.

Un autre est l'attention. Le maintien de l'attention ou attention soutenue est nécessaire pour la bonne réalisation de la plupart des tâches. La fatigue est connue pour diminuer l'attention, ce qui provoque les erreurs que l'on souhaite empêcher.

L'effort mental, ensuite, est un mécanisme par lequel un individu en proie à un début de fatigue peut compenser sa disponibilité réduite pour la tâche, et empêcher ainsi, pour un temps, la chute de performance ou les erreurs que devraient entraîner la fatigue.

La motivation quant à elle est souvent nécessaire pour qu'il y ait effort. De plus, une motivation modérément élevée permet de gommer l'influence des cycles ultradiens et de minimiser celle des rythmes circadiens (Guerrien & Mansy-Danney, 2003).

La charge de travail est un facteur important également, car si elle est trop faible, le travailleur risque de s'endormir, et si elle est trop élevée, outre le risque de mauvaises performances dû à la difficulté des

tâches à forte charge de travail, le travailleur risque d'être amené à l'épuisement. Il semblerait de plus qu'il y ait un niveau de la charge de travail qui maximiserait la motivation dite intrinsèque (Danner & Lonky, 1981). Il y a motivation intrinsèque lorsque le sujet a plaisir à réaliser une tâche ou qu'il ou elle s'est approprié(e) les objectifs de cette tâche.

2.4. Techniques expérimentales pour l'étude de la fatigue mentale

Cette section porte sur les stratégies pour « construire » expérimentalement des états de fatigue.

a) Le temps passé à la tâche

On retrouve fréquemment le temps passé à la tâche comme variable d'étude (par exemple Lorist *et al.*, 2000 ; Trejo *et al.*, 2005). L'idée est simple : on demande au sujet de réaliser une même tâche pendant une durée suffisamment longue pour qu'il montre des signes de fatigue. On peut ensuite comparer les observations entre le début et la fin, ou rechercher une tendance linéaire dans l'ensemble des indicateurs, tendance que l'on est alors tenté de rapprocher avec le temps. C'est une approche qui permet d'observer un effet de la fatigue, mais en prenant le risque d'une superposition d'autres effets, tels que celui de l'automatisation ou des rythmes ultradiens ou circadiens. La fatigue elle-même est influencée par ces rythmes, mais si l'on veut un détecteur de fatigue qui fonctionne à tout moment de la journée, il faut soit intégrer la variable de l'heure de la journée et en estimer la fiabilité dans la détection de la fatigue mentale, soit ne conserver que la part invariante de la fatigue due à la tâche.

b) La privation de sommeil

Dans la littérature, on dénombre une majorité d'études ayant fait usage de la privation de sommeil pour induire une fatigue mentale. L'effet de la privation de sommeil s'additionne avec celui de la tâche réalisée, les sujets n'étant jamais complètement inactifs. Cependant, bien que ce genre de protocoles permette l'observation de différences plus manifestes entre l'état « pas fatigué » et l'état « fatigué », leur intérêt dans le contexte de la détection de la fatigue mentale lors d'une tâche de longue durée n'est pas certain. En effet, on peut douter que, par exemple, la fatigue du pilote soit systématiquement comparable à l'état de privation de sommeil. Comme évoqué précédemment, la fatigue mentale ne se réduit pas à l'hypovigilance (Shen *et al.*, 2006). De plus, les protocoles de privation de sommeil nécessitent des moyens « lourds » comme un environnement médicalisé et du personnel, avec des contraintes ajoutées à ces moyens lorsqu'on sort des horaires de travail normaux.

De manière parente à la privation de sommeil, on s'intéresse aussi aux effets du décalage horaire, du travail de nuit et du travail posté.

c) La mesure subjective

Une mesure de la fatigue en environnement contrôlé permet également de « construire » des états « pas fatigué » et « fatigué » sans chercher nécessairement à induire une fatigue mentale d'une manière spécifique. On laisse simplement le sujet réaliser une tâche en une ou plusieurs sessions et on retient les périodes de temps au cours desquelles les mesures de fatigue ont présenté un niveau de valeurs correspondant à l'un ou l'autre des états recherchés.

On peut ainsi s'appuyer sur des mesures subjectives, par exemple en donnant au sujet un questionnaire à remplir. Pour un suivi de la fatigue avec une résolution temporelle inférieure à l'heure, les questionnaires utilisés sont très simples et consistent en général en une échelle analogique ou en points. Une échelle analogique est une représentation graphique des niveaux possibles de fatigue sous la forme d'un segment dont chaque extrémité correspond à un extrême de fatigue ou d'absence de fatigue. Une échelle en points se présente sous la forme d'une liste de niveaux de fatigue pour chacun desquels une description est fournie. Pour ce second type de questionnaires, le sujet choisit un niveau parmi ceux qui sont proposés. Si ces niveaux sont arrangés dans l'ordre, ce type d'échelle est susceptible d'être rapidement détourné en échelle analogique. En effet, les descriptions sont parfois peu évocatrices pour les sujets et lorsqu'on administre un questionnaire à intervalle régulier, la notion d'historique ou encore de valeur relative n'échappe à aucun esprit, d'où une tendance chez les sujets à utiliser les échelles en points ordonnés comme des échelles analogiques. Malgré cela, les descriptions offrent des points de repère qui améliorent la reproductibilité intra- et inter-sujet des valeurs absolues relevées, ce qui peut présenter un intérêt lorsqu'il est méthodologiquement hasardeux de recaler des séries de mesures entre elles (par exemple entre différents sujets ou sessions).

d) La mesure objective

Enfin, certaines mesures comportementales, tout comme les mesures subjectives, permettent de saisir un état de fatigue lorsque celui-ci se manifeste en environnement contrôlé. On utilise en général une mesure comportementale qui donne une estimation du niveau de performance du sujet. Cette stratégie de la mesure de la fatigue objective est exploitée par une communauté scientifique totalement dissociée de celle traitant de la fatigue mentale liée au temps passé à la tâche, cette dernière communauté étant de bien plus petite taille. En effet, la performance n'est pas nécessairement corrélée à la fatigue subjective, par exemple (Brown, 1994).

3 – La physiologie

3.1. Pourquoi l'électrophysiologie

L'activité des organes d'un individu est une fenêtre d'observation sur l'état interne de cet individu. Sans connaître le détail du mécanisme physiologique de la fatigue mentale, on sait qu'il en existe un, car la fatigue mentale est un processus mental et prend donc naissance dans l'organisme, avec le cerveau au premier rang des organes d'intérêt. La question est de savoir si on peut observer ce mécanisme à l'aide de tel ou tel outil d'observation.

Les techniques électrophysiologiques se présentent en premières candidates de part leur faible coût et leur faible encombrement, ainsi que leur caractère non invasif. Elles sont en particulier transportables « sur le terrain » dans le contexte du pilotage d'aéronefs. Ces qualités ouvrent ces techniques à un large éventail d'applications possibles. Un autre avantage de ces techniques est leur excellente résolution temporelle, de l'ordre de la milliseconde, qui permet un suivi des activités à une échelle temporelle voisine de celle des trains de potentiels d'action par exemple.

Si l'on en vient au cerveau, à l'observation de son activité, relativement peu de techniques sont disponibles. L'électroencéphalographie (EEG) mesure l'activité électrique du cerveau, principalement des neurones du cortex cérébral, via des électrodes posées sur le scalp (le cuir chevelu). La magnétoencéphalographie (MEG) mesure l'activité magnétique du cerveau, mais nécessite un appareil plus encombrant et coûteux. Outre ces deux techniques électrophysiologiques, citons l'imagerie fonctionnelle par résonance magnétique nucléaire (IRMf) et la tomographie par émission de positons (TEP). Ces techniques sont appréciées pour leur relative fine résolution spatiale, mais souffrent d'une médiocre résolution temporelle, de l'ordre du dixième de seconde. Des hypothèses sur les mécanismes physiologiques de la fatigue sont présentées plus loin dans ce chapitre, mais avouons-le tout de suite, il est difficile de savoir pour le moment quelle technique est la plus adaptée à l'étude de la fatigue mentale. Le principe qui propose que l'on favorise la solution la plus simple nous fait choisir l'EEG.

Par ailleurs, le cerveau n'est pas le seul candidat parmi les organes susceptibles de modifier leur activité avec la fatigue mentale. Le champ d'investigation est large. Les pratiques expérimentales font qu'il est fréquent d'associer l'électrooculographie (EOG) et l'électrocardiographie (ECG) à l'EEG, par l'ajout d'une simple paire d'électrodes par modalité supplémentaire, ce afin d'aider au dépistage de segments d'enregistrement EEG contaminé par des activités indésirables.

3.2. Principes

a) L'électroencéphalographie

Le cerveau est constitué de deux grands types de cellules, les neurones et les cellules gliales. Seuls les neurones ont une activité électrique susceptible de contribuer significativement au signal mesuré par des électrodes placées sur le scalp (l'EEG). Les neurones communiquent en émettant des potentiels d'action, mais on considère actuellement que ce sont les courants dont les dendrites sont le siège qui participent le plus au signal que l'on mesure. Ces courants sont fonctions des potentiels postsynaptiques excitateurs et inhibiteurs, donc de l'activité de communication des neurones entre eux. Ces courants sont appelés courants primaires. Les courants secondaires circulent dans le milieu extracellulaire et sont la conséquence des courants primaires et des mécanismes cellulaires qui servent à l'établissement ou au maintien du potentiel de membrane. Les courants secondaires se propagent des dendrites au corps cellulaire du neurone considéré, mais aussi aux neurones voisins.

Les deux types de courants sont à l'origine d'ondes électromagnétiques qui traversent différents milieux, tissu cérébral, liquide céphalo-rachidien, crâne, peau, jusqu'à atteindre les électrodes qui elles mesurent chacune une différence de potentiel avec une électrode commune appelée référence.

On utilise en général de quelques-unes à une soixantaine d'électrodes. Celles-ci sont positionnées sur le scalp suivant une convention appelée « système 10-20 » (Jasper, 1958), si bien qu'il peut être utile de désigner les électrodes sur lesquelles des phénomènes particuliers ont été observés.

Le signal EEG n'est pas en totalité d'origine cérébrale. Les muscles péricrâniens et les globes oculaires causent par leur activité des perturbations parfois de très forte amplitude relative.

b) L'électrooculographie

L'œil et la paupière sont les organes dont l'activité est mesurée par l'EOG. On enregistre leur activité principalement pour mesurer la contamination des segments d'EEG par ces activités. D'un point de vue électrique, le globe oculaire peut être assimilé à un dipôle électrique, du fait de la différence de potentiel entre la cornée et la rétine. Les mouvements des yeux étant des rotations de ce dipôle, ils modifient la distribution du potentiel en de nombreux points alentours. Pour détecter les mouvements oculaires, on place deux électrodes, une en dessous et une au dessus de l'œil, entre lesquelles on mesure la différence de potentiel. Un tel montage permet d'enregistrer les mouvements verticaux des yeux. On ajoute souvent une seconde paire d'électrodes, placées à gauche de l'œil gauche et à droite de l'œil droit, pour enregistrer les mouvements horizontaux.

Au mouvement des globes oculaires s'ajoutent l'activité électrique des muscles qui permettent ce mouvement. La paupière, lorsqu'elle s'ouvre et se ferme, agit aussi par l'action d'un muscle, bien que l'essentiel de l'artefact sur l'EEG que cause un clignement soit dû à un mouvement réflexe du globe

oculaire vers le haut (phénomène de Bell). Les électrodes placées sur la peau ont une grande sensibilité à l'activité électrique des muscles, du fait de la proximité de ces derniers à ces premières.

c) L'activité électrique des muscles

Les muscles, bien que nous n'ayons pas cherché à spécifiquement enregistrer leur activité, contribuent à tous les autres signaux électrophysiologiques. Ce sont des muscles dits squelettiques, à commande consciente et constitués de fibres musculaires parallèles. Chaque fibre musculaire a une activité électrique au niveau d'une partie appelée plaque motrice ou jonction neuromusculaire. On parle de potentiel de plaque motrice, à l'origine des variations de potentiel électrique enregistrées lors d'une contraction. Ce potentiel prend naissance principalement sur la membrane postsynaptique, c'est-à-dire la membrane de la fibre musculaire, et est dû à l'ouverture de canaux récepteurs de l'acétylcholine qui entraîne un flux entrant d'ions positifs, principalement une entrée d'ions sodium et une sortie d'ions potassium.

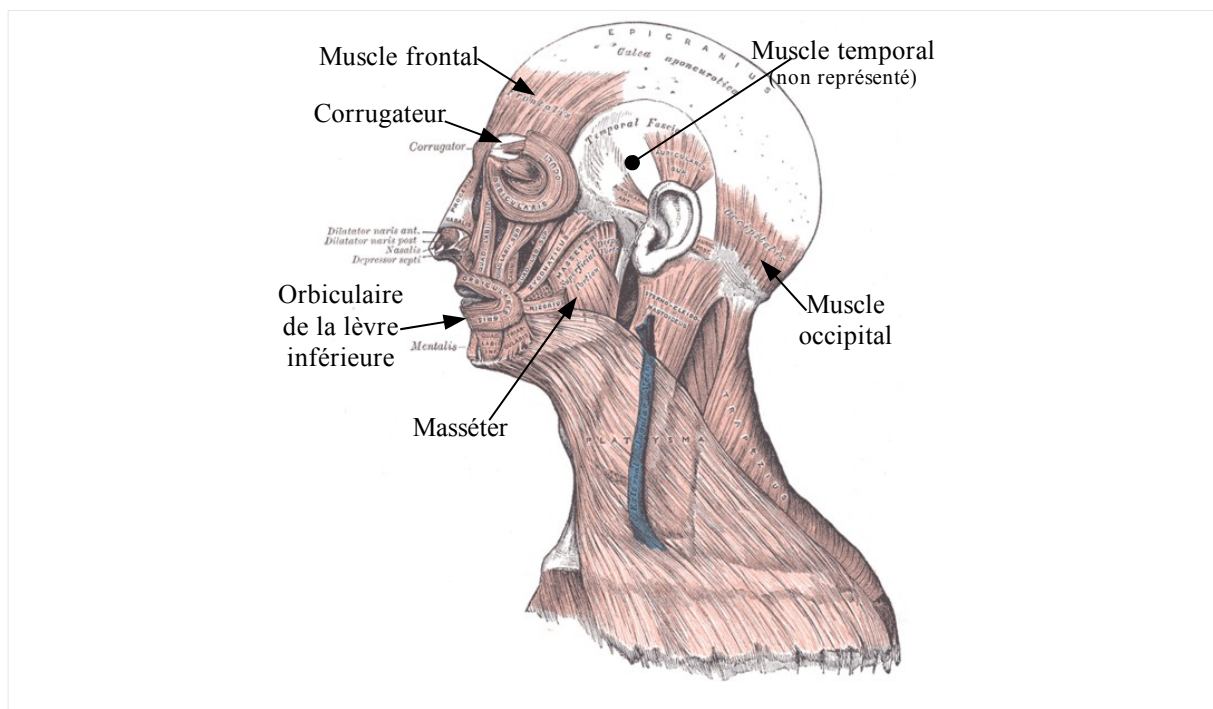


Fig. 2 – Principaux muscles de la tête, dont l'activité électrique peut perturber l'EEG. Certains muscles sont marqués en double, en version française, car ils sont cités à nouveau dans la suite de ce document.

(illustration de la 20^{ème} édition américaine du *Gray's Anatomy of Human Body*)

d) L'électrocardiographie

Le cœur est un muscle strié particulier car à commande involontaire. Pour enregistrer son activité, on place deux électrodes diamétralement opposées par rapport au cœur, ces deux électrodes formant une dérivation, par exemple une au bas des côtes côté gauche, et une au dessus d'une des deux clavicules ou omoplates (gauche ou droite). Sur l'ECG apparaissent des ondes de grande amplitude qui localisent dans le temps les contractions et dilatations lors d'un battement cardiaque.

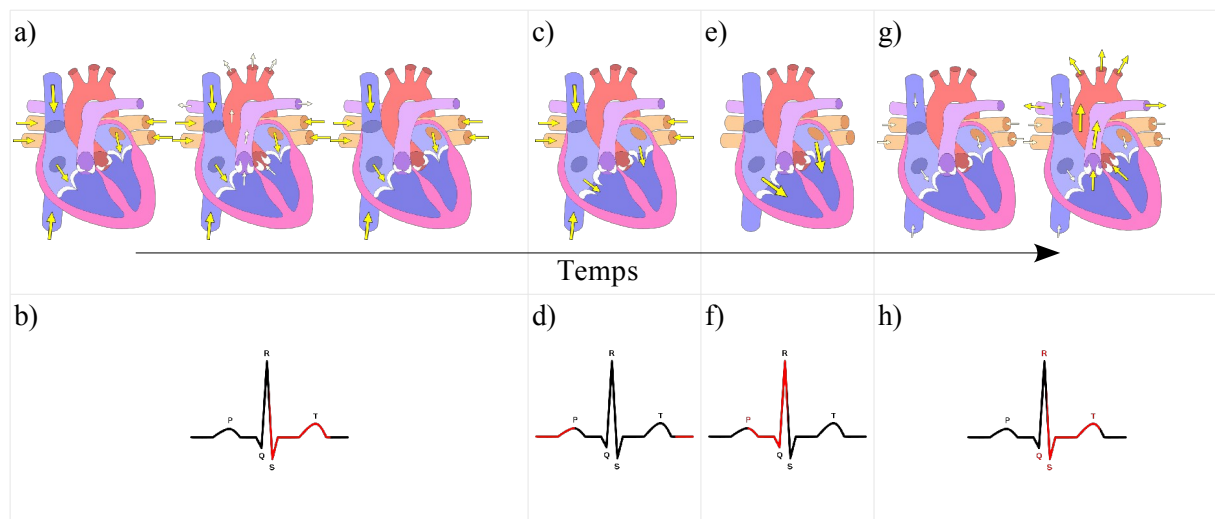


Fig. 3 – Séquence des contractions et dilatations au cours du cycle cardiaque (Silbernagl & Despopoulos, 1997). a,b) Pendant la systole ventriculaire et le début de la diastole, les oreillettes droite et gauche se dilatent pour permettre l'entrée du flux sanguin, des veines caves pour l'oreillette droite, et de la veine pulmonaire pour l'oreillette gauche. c,d) Les ventricules se relâchent et les valvules entre oreillettes et ventricules s'ouvrent, permettant ainsi le remplissage passif des ventricules. e,f) Toujours au cours de la diastole, et suite à l'onde P, la systole auriculaire, c'est-à-dire la contraction des oreillettes, éjecte le sang vers les ventricules ; on parle de remplissage actif des ventricules ; pendant cette phase, les oreillettes n'admettent plus de sang. g,h) Une nouvelle systole ventriculaire débute, amorcé par le complexe QRS, encore appelé systole électrique ; les ventricules se contractent jusqu'à provoquer l'ouverture des sigmoïdes et l'éjection du sang, du ventricule droit vers les poumons via l'artère pulmonaire, du ventricule gauche vers le reste de l'organisme via l'aorte ; l'onde T marque la fin de l'onde électrique et le point où les ventricules atteignent leur volume minimal, c'est-à-dire la fin de leur contraction.

L'événement électrique correspondant à une contraction précède cette dernière.

(illustration originale du cœur empruntée à Eric Pierce² et modifiée).

La recherche d'anomalie dans l'activité du cœur nécessite souvent plusieurs dérivations placées dans un même plan. Dans notre étude, nous nous sommes contentés d'analyser le rythme des battements cardiaques, ce qui est suffisant pour observer l'activité du système nerveux autonome via celle du cœur. Ce système n'est pas sous le contrôle direct de l'action consciente, il est piloté en majeure partie par l'hypothalamus et module l'activité de plusieurs organes à la fois. Deux composantes de modulation de l'activité de ces organes s'opposent continuellement, la composante (ortho-)sympathique et la composante parasympathique. La première, lorsqu'elle augmente son activité, est connue pour accélérer le cœur, entre autres effets, alors que la seconde, de même lorsque son activité augmente, ralentit le cœur. Une augmentation du rythme cardiaque peut être due à une augmentation de l'activité sympathique comme à une diminution de l'activité parasympathique, d'où la notion d'équilibre (ou de balance) ortho-para-sympathique.

3.3. Fatigue mentale et système nerveux central

Le système nerveux central étant notre organe vedette, que peut-on dire des mécanismes de la fatigue à son niveau ?

a) L'hypovigilance

Le passage de l'éveil au sommeil (dont l'hypovigilance est une étape) a été abondamment étudié, et l'on sait à présent que c'est un mécanisme très particulier qui ne saurait rendre compte de tous les aspects de la fatigue mentale. Le sommeil semble être un phénomène qui modifie l'activité d'à peu près tous les organes et le nombre de publications est tel que ce document ne peut prétendre inventorier ne serait-ce que les principales observations faites liées à la transition de l'état de veille à celui de sommeil. On note par exemple qu'un contingent d'études a porté sur le rôle de nombreux noyaux sous corticaux, entre autres des noyaux hypothalamiques qui auraient le pouvoir d'initier le basculement de l'éveil vers le sommeil, et certains auteurs défendent la thèse d'une transition rapide (Saper *et al.*, 2005). D'autres études ont exploité l'EEG, inventorié différents stades de la transition de la veille au sommeil et proposent à l'opposé un mécanisme progressif pour cette transition (Ogilvie, 2001). Avec la dérive vers le sommeil, l'activité corticale se synchronise sur des fréquences précises, quoique changeantes au cours de la transition, pour donner des ondes de plus grandes amplitudes, ce qui a pour effet d'inhiber les activités porteuses d'information, plus désorganisées, moins étendues et moins amples. C'est autour de ce genre d'hypothèses qu'un autre contingent d'études électrophysiologiques apportent des éléments sur les interactions entre cortex cérébral et thalamus (Steriade *et al.*, 1993).

2 Sous les termes de la licence Creative Commons dite By-Sa décrite à l'adresse suivante : <http://creativecommons.org/licenses/by-sa/2.0/>

b) La « fatigue des réseaux cérébraux »

Les neurones ayant des ressources limitées, et peut être aussi des mécanismes locaux de protection contre l'excès d'activité, on peut imaginer qu'une fatigue des circuits neuronaux (ou plus exactement neuronogliaux), semblable en certains aspects à la fatigue musculaire, soit possible, en particulier chez un individu qui réalise une tâche à forte charge de travail, prolongée au delà de ce à quoi il ou elle est habitué(e). L'hypothèse d'une limitation énergétique à l'activité cérébrale est supportée par les chiffres élevés des demandes du cerveau : alors que ce dernier ne représente que 2 % du poids corporel total, il représente à lui seul 15 % de l'irrigation sanguine totale, 20 % de la consommation d'oxygène et 25 % de la consommation de glucose³.

Cependant, à la différence des cellules musculaires, les neurones n'accumulent pas le lactate lors d'une activité intense ; au contraire, le lactate est une source d'énergie pour les neurones plutôt qu'un résidu du catabolisme. De plus, il est fourni par les astrocytes voisins plutôt que produit directement par les neurones. Cependant, les astrocytes pourraient ne pas pouvoir en fournir assez en situation d'activité soutenue et prolongée. Ils ont en outre la capacité d'influer sur l'activité neuronale (Volterra & Steinhäuser, 2004). Par exemple, le glutamate, principal neurotransmetteur du cortex cérébral, doit être éliminé de la fente synaptique une fois libéré, pour que les récepteurs sur le neurone postsynaptique soient libres en proportion suffisante. Des cytokines pro-inflammatoires ont exhibé un rôle dans la recapture et le transfert du glutamate par les astrocytes (Rönnbäck & Hansson, 2004). Ces cytokines sont présentes en concentrations anormales dans la plupart des maladies accompagnées de sensations de fatigue.

On découvre tout un registre d'études sur le métabolisme de la fatigue sous le terme « fatigue centrale », c'est-à-dire une fatigue en lien avec l'activité physique mais qui siège dans le système nerveux central. Par exemple, des acides aminés dits ramifiés (BCAA pour *branched chain amino acids*, en anglais) seraient consommés lors de l'effort physique (Newsholme *et al.*, 1991 ; Castell *et al.*, 1999) ce qui entraînerait une augmentation du taux de sérotonine, cette dernière ayant un rôle dans l'induction du sommeil. En revanche, il n'a pas été montré qu'une tâche mentale puisse avoir un même effet, quoique une étude récente a montré une diminution de la valine (un BCAA) dans le plasma après une série de tâches cognitives de 4 h (Nozaki *et al.*, 2009). Toujours au sujet du « métabolisme » de la fatigue, on trouve aussi des études sur les performances intellectuelles augmentées après une prise de produits que certains sportifs utilisent (par exemple, Watanabe *et al.*, 2002).

Une autre hypothèse concerne le rôle de l'adénosine (Porkka-Heiskanen *et al.*, 2002), molécule libérée tout comme le glutamate par l'exocytose des vésicules du bouton présynaptique, du fait de sa présence dans le milieu intracellulaire. L'adénosine se fixe sur la membrane présynaptique et inhibe l'ouverture

3 Pierre Magistretti, communication personnelle. Support de présentation disponible à l'adresse suivante : http://www.college-de-france.fr/media/pier_magi/UPL1943_Cours1_pierre.magistretti.pdf

des canaux calciques potentiel-dépendants. De plus, sa concentration intracellulaire augmente avec l'activité métabolique. Ce mécanisme pourrait avoir un rôle protecteur, en limitant la libération de glutamate après une forte activité, car le glutamate est neurotoxique à concentration élevée. Il apparaît probable que le mode d'action de l'adénosine se déclenche lors d'une activité liée à une tâche et qu'il entraîne alors des limitations de cette activité.

Il semblerait aussi que ce mécanisme ait une contribution dans la promotion du sommeil (Porkka-Heiskanen *et al.*, 2002). Par exemple, une augmentation de la concentration de l'adénosine extracellulaire, tout au long de la période d'éveil, a été observée dans une région cérébrale appelée télencéphale basal. De plus, il a été observé que l'adénosine, en concentration élevée, entraînait l'apparition de rythmes EEG semblables à ceux du sommeil (Benington *et al.*, 1995).

Enfin, les systèmes de l'adénosine et de l'interleukine 1 (une cytokine) seraient couplés chez le rat sain (Luk *et al.*, 1999).

Il n'y a qu'un pas pour émettre l'hypothèse que l'implication probable d'une même molécule, à la fois dans la fatigue liée à une activité mentale et dans l'hypovigilance, pourrait expliquer la possible transition de la première fatigue à la seconde.

c) Les fatigues subjective et cognitive

Enfin, si l'on adhère à l'hypothèse de Boksem et Tops, la sensation de fatigue, en particulier le mécanisme à l'origine du désengagement de la tâche, impliquerait un certain nombre de structures cérébrales telles que le noyau accumbens, le cortex orbito-frontal, l'amygdale baso-latérale, l'insula et le cortex cingulaire antérieur (Boksem & Tops, 2008). Le réseau dopaminergique, dont les nombreuses projections prennent racines pour l'essentiel dans le mésencéphale, est connu pour son implication dans les mécanismes de récompense et donc de prise de décisions.

Dans la fatigue centrale, liée donc à l'exercice physique, une perte de motivation est parfois observée. La sérotonine a été abondamment étudiée et son lien avec la dopamine attire à présent les intérêts (Meeusen *et al.*, 2006).

3.4. Quantifications

a) Variables électroencéphalographiques

En EEG, on enregistre autant de signaux que l'on a d'électrodes, hors référence et terre. La reproductibilité du placement des électrodes EEG est permise par l'utilisation d'un standard appelé « système 10-20 ». Cette reproductibilité s'entend à l'échelle des lobes cérébraux. Par conséquent les physiologistes s'intéressent aux observations en chaque électrode.

L'activité EEG, quelque soit l'électrode, présente bien souvent un contenu spectral avec des inhomogénéités intéressantes. Ainsi, on considère le partitionnement suivant du domaine fréquentiel (au Hertz près) : delta (0,1-3 Hz), thêta (3-7 Hz), alpha (7-13 Hz), bêta 1 (13-18 Hz), bêta 2 (18-30 Hz), gamma (30-100 Hz et plus). Ces intervalles sont appelées « bandes de fréquences ». Les bandes de plus hautes fréquences n'ont pas été considérées dans ce travail, en particulier parce qu'elles s'avèrent plus contaminées par l'activité des muscles voisins et que l'on cherche à mesurer surtout l'activité du cerveau. Autrement dit, la contribution des muscles à l'EEG est considérée comme du bruit.

L'interprétation de ces rythmes est liée au phénomène de synchronisation. Si les neurones d'une même région se font de plus en plus nombreux à décharger aux mêmes instants, phénomène rendu possible par les nombreuses connexions établies entre eux, le signal généré par l'ensemble de ces neurones présente alors un spectre fréquentiel avec un pic. Les mécanismes d'entraînement mutuel des activités des neurones qui permettent ces phénomènes de synchronisation s'opèrent sur une étendue de surface variable, parfois suffisamment large pour être observable au niveau d'une ou plusieurs électrodes.

On parle de synchronies locales car elles se limitent à une région, de même qu'il existe des synchronies distantes qui sont des synchronies entre les neurones de deux régions distantes (Rodriguez *et al.*, 1999 ; Varela *et al.*, 2001). Deux régions qui présentent une modulation de synchronie distante n'exhibent pas nécessairement une modulation correspondante de leurs synchronies locales respectives.

Ces deux types de synchronies sont étudiés par l'intermédiaire de quantifications que l'on peut découper naturellement en deux groupes correspondants : les mesures de synchronie locale et les mesures de synchronie distante. De plus, du fait du profil spectral particulier de l'EEG, ces mesures sont évaluées pour un contenu fréquentiel donné.

Ainsi, pour quantifier un niveau de synchronie locale, on peut estimer l'énergie du signal dans les différentes bandes de fréquences d'intérêt, cela en s'aidant par exemple de la transformée de Fourier du signal. Pour les synchronies distantes, on peut exploiter la partie correspondante à la bande de fréquences du spectre croisé qui s'écrit aussi comme la transformée de Fourier du corrélogramme entre les deux signaux.

D'autres quantifications s'aident d'une « réécriture » du signal qui permet la distinction entre les concepts d'amplitude et de phase. On estime pour cela le signal analytique d'un signal réel x (obtenu ici par filtrage dans une des bandes de fréquences d'intérêt) qui est le signal complexe z tel que :

$$z(t) = x(t) + i \cdot H\{x\}(t) = a(t) \cdot e^{j\phi(t)}$$

ou H est la transformée de Hilbert, a est l'amplitude du signal et ϕ sa phase.

Les mesures de synchronie locale reposent principalement sur la notion d'amplitude, tandis que les mesures de synchronie distante exploitent aussi (ou seulement) les phases des signaux étudiés.

La quantification de liens dynamiques entre des sites distants préfigure la notion de plus en plus populaire dans le milieu scientifique des réseaux fonctionnels. En effet, on peut construire le réseau des interactions entre sites distants, et interpréter ensuite les caractéristiques de ces réseaux.

b) Variables électrooculographiques

L'analyse de l'EOG s'appuie sur les clignements palpébraux (des paupières). Il est bien entendu possible d'extraire et d'étudier les mouvements horizontaux et verticaux autres que les clignements, mais cela n'a pas non plus été couvert par cette thèse. Dans la plupart des protocoles expérimentaux dits de laboratoire, on demande aux sujets de ne pas bouger des yeux, pour limiter la contamination des activités électriques oculopalpébrales dans l'EEG.

Les clignements se présentent dans l'EOG comme des déflexions orientées par convention vers le bas. Les principales quantifications des clignements sont leur amplitude moyenne, leur durée moyenne et leur nombre sur une fenêtre de temps choisie, ou de manière équivalente, leur fréquence. Ces quantifications ont montré des variations avec la fatigue mentale, dans différents contextes (Stern *et al.*, 1994 ; Morris & Miller, 1996). On peut également s'intéresser à des quantifications moins utilisées telles que la variabilité de l'amplitude ou de la durée, ou encore diverses mesures (moyenne et écart-type par exemple) des intervalles inter-clignements afin de quantifier la variabilité du rythme des clignements.

c) Variables électrocardiographiques

Comme écrit précédemment, l'analyse de l'ECG normal porte essentiellement sur les battements cardiaques et en particulier les ondes R plutôt que sur les fluctuations du signal par exemple. La première mesure du rythme de ces battements est la fréquence cardiaque, en général estimée comme la fréquence instantanée en chaque onde R ; sur une fenêtre de temps, on prendrait la moyenne des fréquences instantanées des ondes R présentes. La variabilité du rythme cardiaque a également fait l'objet de nombreuses études et quantifications (lire Task Force, 1996, pour une introduction). Parmi ces quantifications, on peut citer l'écart-type des intervalles dits R-R, c'est-à-dire des intervalles inter-battements, d'une onde R à l'onde R suivante. On dénombre aussi diverses mesures spectrales du signal R-R, obtenu par interpolation des intervalles R-R. La figure 4 donne une illustration des étapes de la construction de ce signal dérivé.

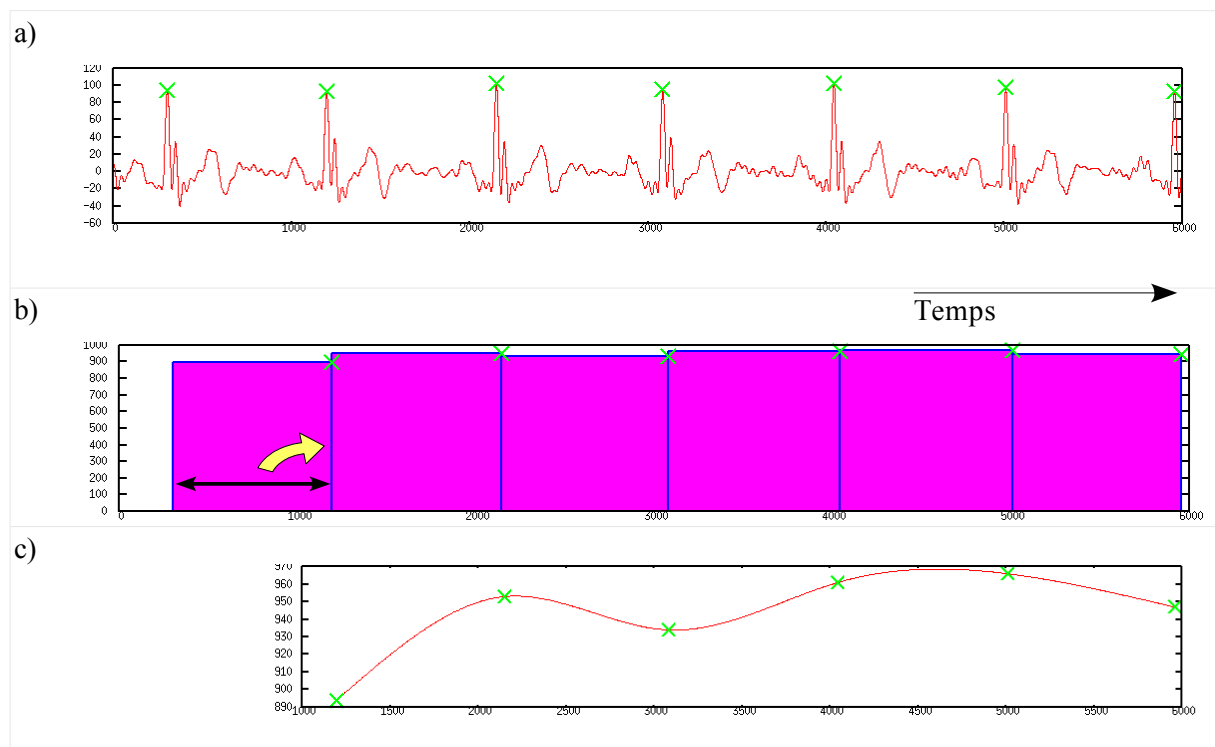


Fig. 4 – Construction du signal R-R interpolé. a) ECG filtré avec ses imperfections (dans cet exemple, le complexe QRS apparaît comme coupé en deux pics) ; les ondes R sont marquées par des croix vertes. b) Signal R-R construit comme la série des intervalles de temps entre ondes R successives ; chaque onde R est remplacée par l'intervalle correspondant. c) Signal R-R interpolé par splines cubiques.

Par exemple, la puissance dans la bande dite hautes fréquences (HF pour *high frequency*, soit 0,15-0,4 Hz) refléterait le niveau d'activation parasympathique, tandis que la puissance dans la bande basses fréquences (LF pour *low frequency*, soit 0,04-0,15 Hz) serait une fonction des activations sympathique et parasympathique. Citons également la puissance autour de la fréquence de 0,1 Hz, dont le lien avec la charge de travail a été beaucoup discuté (Nickel & Nachreiner, 2003). À l'exception de la fréquence cardiaque, ces mesures traitent toutes de la variabilité du rythme cardiaque, ou *heart rate variability* (HRV) en anglais.

4 – Analyse de données

4.1. Vue générale de la chaîne des traitements

a) Le détecteur, ses entrées, ses sorties

L'emploi d'un détecteur permet d'obtenir une information que l'on ne saurait obtenir autrement, en utilisant par ailleurs des informations plus aisément accessibles. Ces informations sont relatives à un état, celui d'un individu, et donc à des variables (des choses susceptibles de varier).

La procédure générale consiste à concevoir une expérience dans laquelle la tâche qu'auront à réaliser les sujets sera aussi proche que possible du contexte d'application réel, tout en permettant la lecture des informations que l'on souhaiterait obtenir par l'intermédiaire du détecteur.

Ainsi, la première étape (toute la procédure en fait) repose sur le choix, d'un côté, des informations que l'on souhaite obtenir par le biais du détecteur, c'est-à-dire les informations qui seront disponibles dans l'environnement contrôlé de l'expérience et qui ne le seront plus dans l'environnement réel, et de l'autre côté, des informations à enregistrer qui constitueront l'entrée du détecteur, la matière première à partir de laquelle la fatigue pourra (ou non) être détectée.

Pour alléger le style d'écriture, désignons ces informations respectivement par informations de sortie et informations d'entrée.

Dans le cas des études réalisées au cours de cette thèse, les informations d'entrée étaient invariablement les enregistrements de l'activité physiologique. On se dirigeait donc vers des applications dans lesquelles les opérateurs (les pilotes ?) seraient équipés d'électrodes EEG, entre autres.

Les informations de sortie étaient quant à elles toujours des mesures comportementales liées à une consigne (par exemple, le sujet répondait à une « question » pour laquelle l'expérimentateur connaissait la réponse), plus, pour certaines études, des mesures de fatigue subjective (des réponses répétées à un même questionnaire).

Le point crucial ici est que les deux types d'information doivent être bien définis et isolés l'un de l'autre pour ne pas introduire de biais dans la méthode. En effet, la procédure, du moins ce qui en est présenté, nécessite toujours la connaissance de la (vraie) « sortie » de manière à pouvoir tester la validité des réponses du détecteur en simulant en quelque sorte la situation où le détecteur doit prédire la sortie sur la seule base des informations d'entrée. Il est donc possible pour l'expérimentateur de fauter en enrichissant les informations d'entrée avec des informations de sortie, pas nécessairement la réponse que doit rendre le détecteur, mais n'importe quelle information qui ne sera en réalité pas disponible dans le contexte réel d'application visé.

b) Étapes de la méthode

Le cœur de la chaîne des traitements implique les étapes suivantes :

- La tâche, une fois réalisée, est tout d'abord découpée en fenêtres de temps ; la durée des fenêtres est à choisir, mais un point important est que la définition de ces fenêtres ne doit s'appuyer que sur les informations d'entrée ; dans les études menées au cours de cette thèse, les fenêtres étaient glissantes et de durée fixe, donc non calées aux stimulations ou événements liés à la tâche ; ce type de paradigmes d'analyses en temps-réel est qualifié d'asynchrone.
- Parmi les fenêtres de temps segmentées, sur la base des informations de sortie, certaines sont assignées à une condition 1 (un premier état mental, par exemple l'état « non fatigué »), d'autres sont choisies pour représenter une condition 2 (par exemple l'état « fatigué ») ; une construction en plus de deux conditions (ou classes) est possible.
- Les signaux des fenêtres de temps assignées à l'une ou l'autre des classes sont quantifiés en un nombre constant de variables (soit n ce nombre), si bien que l'on représente à présent chaque fenêtre de temps (ou individu statistique) comme un point dans un espace à n dimensions.
- Les signaux quantifiés sont fournis, comme données d'entrée, à un algorithme (une « boîte noire ») dont la donnée de sortie est la classe ; cet algorithme construit puis éprouve le détecteur sur la base des différences qu'il relève, dans les signaux quantifiés, entre les classes.

Cette section présente les détails du dernier traitement énoncé. La construction du détecteur est permise par les outils de classification, tandis que la mise à l'épreuve du détecteur est l'objet des procédures de validation.

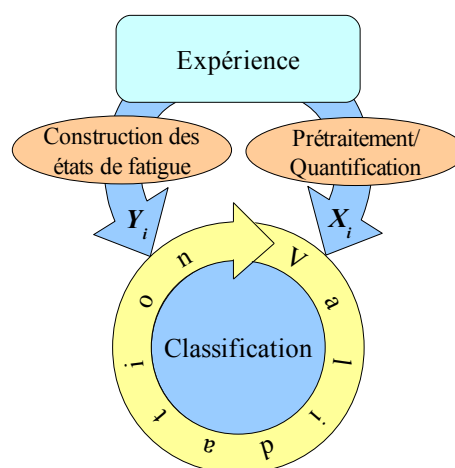


Fig. 5 – Schéma général de la chaîne des traitements.

4.2. Classes et classification

a) Principes et enjeux

La définition de classes est une première procédure des plus déterminantes pour l'étude. Pour les études réalisées au cours de cette thèse, les classes sont invariablement l'état « non fatigué » d'un côté, et l'état « fatigué » de l'autre.

Ensuite, la classification est l'affaire d'un algorithme dont l'objet est la construction du détecteur, que l'on appellera également classifieur. Cette étape constitue le cœur de la procédure d'analyse, car c'est elle qui, par la recherche de corrélats, offre une observation dans l'espace des données d'entrée liée à la mise en opposition des classes les unes aux autres.

Comme décrit précédemment, un tel algorithme prend en entrée les fenêtres de temps sous la forme de points dans un espace à n dimensions, et en sortie les mêmes fenêtres de temps sous la forme de leur appartenance à telle ou telle classe (une seule variable qualitative, binaire dans les cas présentés dans cette thèse). Les données d'entrée se représentent donc comme autant de groupes de points, dans cet espace à n dimensions, qu'il y a de classes. Un détecteur doit autant que possible pouvoir prédire la classe à laquelle appartient n'importe quel point de cet espace, y compris des points autres que ceux utilisés pour la construction du détecteur. En effet, on souhaite que le détecteur puisse reconnaître des situations semblables, sans être parfaitement identiques à celles auxquelles il a été précédemment confronté. Le détecteur correspond donc à un partitionnement de l'espace d'entrée (Fig. 6).

Ensuite, on part du postulat que les données ne sont pas parfaites, qu'elles sont en quelque sorte entachées de quelques erreurs. Cela est dû au fait que les quantifications offrent une fenêtre d'observation partielle sur le phénomène à étudier. La figure 6 illustre un cas simulé mais représentatif de beaucoup de problèmes réels ; dans cet exemple, il semble inapproprié de vouloir séparer les deux nuages de points « à tout prix ». Certaines heuristiques peuvent s'avérer des plus utiles. Dans la mesure où la problématique de départ a été bien posée (que l'on étudie par exemple un seul phénomène par classe) et où les mesures et quantifications sont adaptées aux phénomènes à étudier, on peut espérer que les régions représentatives des classes soient de formes simples, par exemple connexes voire, avec un peu de chance, convexes.

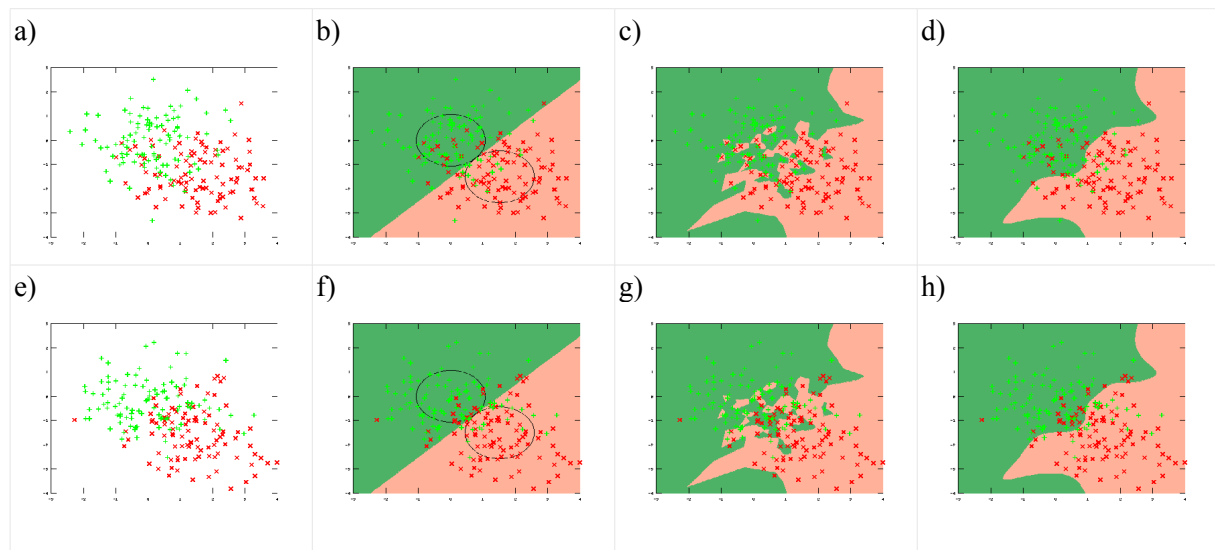


Fig. 6 – Exemple simulé de deux classes (verte et rouge) décrites en deux dimensions (axes). a) Ensemble des données disponibles pour la construction du classifieur (données dites d'apprentissage) : chaque point représente par exemple une fenêtre de temps ; une classe est l'ensemble des points d'une même couleur. b) Régions théoriques à retrouver ; les classes sont ici issues de distributions gaussiennes dont une isocourbe est donnée en noir. c) Régions permettant une séparation parfaite des classes au sens des seules données disponibles. d) Régions calculées par un algorithme qui cherche un compromis entre l'adéquation aux données et la simplicité de la forme des régions. e,f) Nouvel ensemble de données issu de la même distribution théorique (données dites de test ou de validation). g,h) Solutions des deux partitionnements précédents ; cette fois-ci la seconde solution se comporte mieux que la première, car moins de points verts tombent dans la zone rouge et de points rouges dans la zone verte.

b) Deux approches pour la classification

Il existe deux grands modes opératoires pour la construction d'un classifieur. Le premier consiste à travailler classe par classe (nuage de points par nuage de points) pour modéliser avec plus ou moins d'approximation la distribution des points de chaque nuage. C'est l'approche modélisante, où chaque classe est décrite dans l'espace des mesures, sous une forme alternative au nuage de points, une forme plus adaptée en particulier à l'interprétation. Les algorithmes qui implémentent cette approche ont en général l'intérêt de produire des classifieurs que l'on peut mettre à jour, par exemple suite à l'ajout de points supplémentaires ou même d'une classe supplémentaire. Ces algorithmes offrent bien souvent un moyen de contrôle sur la forme des régions.

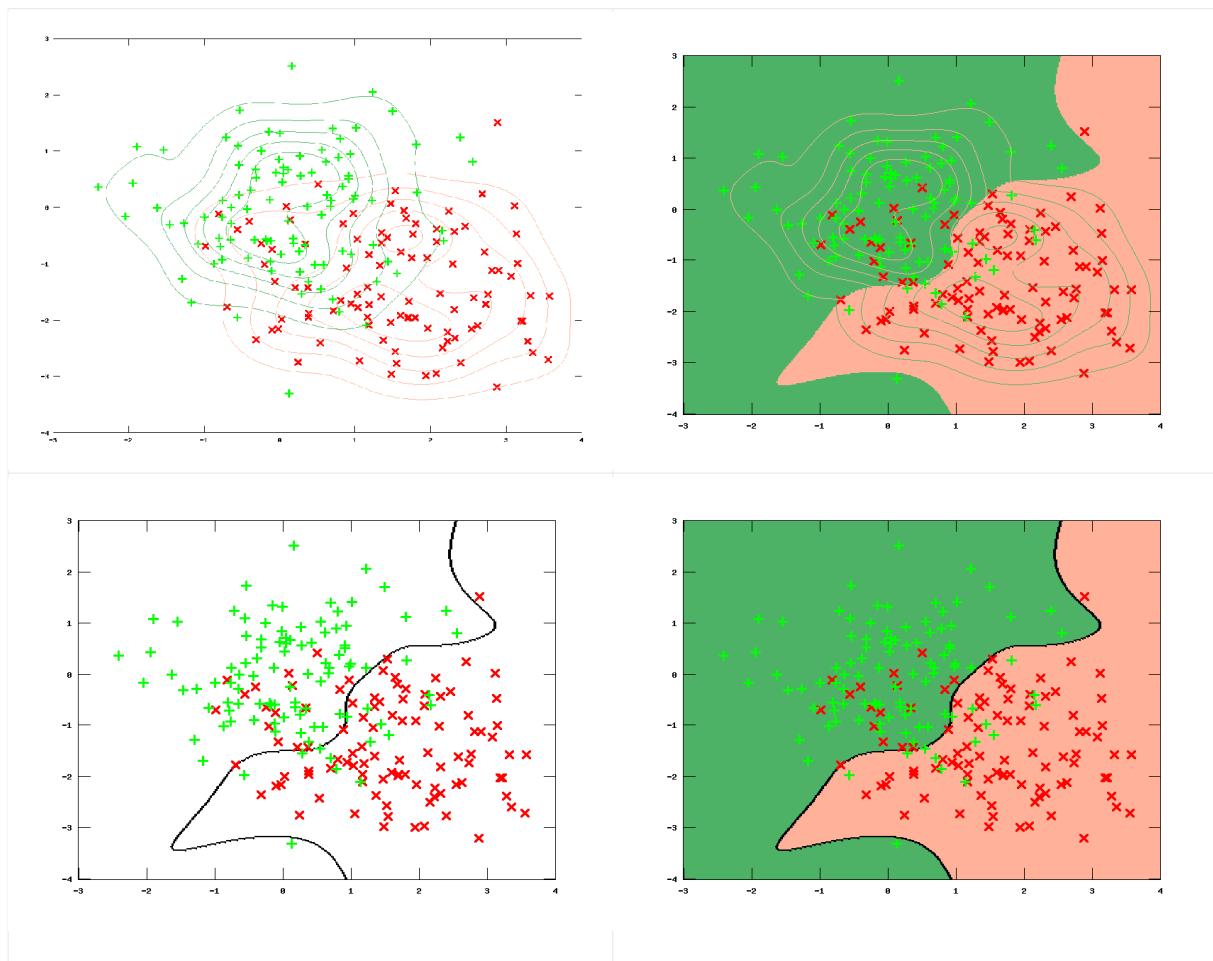


Fig. 7 – Exemples de classifieurs au comportement identique, pour deux classes (verte et rouge) et un espace de quantification à deux dimensions ($n = 2$). La classification est le processus de partitionnement de l'espace d'entrée en autant de régions qu'il y a de classes. Haut : ce partitionnement peut se faire par l'emploi d'une représentation de chaque nuage de points qui permet l'estimation en tout point de l'espace de la probabilité d'appartenir à la classe (les lignes de niveau) ; le partitionnement est alors aisé, la classe sélectionnée pour un point donné étant la plus vraisemblable. Bas : le partitionnement peut se faire aussi par la construction de la frontière de séparation entre les différentes régions.

L'autre approche, que l'on décrit plus facilement dans le cas de deux classes, avec l'image de deux nuages de points en recouvrement partiel, consiste à construire une surface de séparation entre les classes. C'est l'approche discriminante. La surface de séparation est également porteuse d'information sur les classes, en particulier sur ce qui les différencie, mais dans une moindre mesure que des distributions de classe comme dans l'approche précédente. En revanche, le calcul d'une séparatrice est bien souvent plus rapide, si bien que les algorithmes de classification basés sur cette approche sont en

général favorisés lorsque l'interprétation de chacune des classes comparées aux autres n'est pas le souci premier.

Un autre avantage de cette méthode, plus délicat à comprendre, est la possibilité de contraindre la forme de cette séparatrice. C'est un avantage car, lorsque l'on a plusieurs solutions qui permettent une même précision de séparation des nuages de point, un choix raisonnable est de sélectionner la solution la plus simple. Ce principe est connu sous le nom de principe du rasoir d'Occam⁴. La simplicité d'une séparatrice est celle de sa représentation. Une séparatrice simple sera par exemple un hyperplan (une droite dans un espace à deux dimensions, un plan dans un espace à trois dimensions, un sous-espace de dimension $n - 1$ dans un espace à n dimensions), et une séparatrice plus complexe exhibera des courbures comme dans les exemples de la figure 7.

Opter pour une solution simple est une stratégie pour faire peser plus la contribution des nuages de points pris comme des ensembles, que celle de points pris isolément, ceci afin par exemple de minimiser la contribution des quelques valeurs aberrantes (des points anormalement éloignés du reste du nuage, encore appelés *outliers* en anglais), s'il y en a.

En outre, en concentrant l'effort de classification sur les frontières entre régions, on peut obtenir un partitionnement plus précis là où il est important d'être précis, puisqu'une erreur sur la frontière fait très vite tomber de nombreux points du mauvais côté. En effet, les données de problématiques réelles forment souvent des nuages de points fort entremêlés, difficile à séparer.

Une technique a rencontré un net succès dans la communauté scientifique et au delà : les séparateurs à vastes marges, ou *support vector machines* en anglais, SVM en abrégé. Le terme original anglais et sa traduction française ne semblent pas équivalents. Ces deux termes reflètent pourtant deux aspects d'une même caractéristique clef.

c) Le séparateur à vastes marges (SVM)

Si on considère un problème de classification binaire (à deux classes), et que l'on souhaite estimer une séparatrice linéaire, on introduit les notations suivantes :

- soit \mathbf{x}_i le vecteur représentatif de la fenêtre de temps i dans l'espace H à n dimensions (le point i), et $y_i \in \{-1, 1\}$ l'étiquette de classe (ces valeurs représentatives ne sont pas choisies au hasard) ;
- soit \mathbf{w} le vecteur orthogonal à l'hyperplan séparateur et b le biais tels que : $f_i = \langle \mathbf{w}, \mathbf{x}_i \rangle + b$ et $\hat{y}_i = \text{signe}(f_i)$ avec par conséquent $\hat{y}_i \in \{-1, 1\}$ l'étiquette de classe estimée par le classifieur linéaire (\mathbf{w}, b) ;
- soit le problème d'optimisation suivant, dans lequel intervient un paramètre c positif :

4 « *Entia non sunt multiplicanda praeter necessitatem* » ou « *pluralitas non est ponenda sine necessitate* » (Guillaume d'Ockham, 14^e siècle) : les multiples ne doivent pas être utilisés sans nécessité.

$$\begin{aligned} \min_{\mathbf{w}, b} c \cdot \sum_i \epsilon_i + \frac{\|\mathbf{w}\|^2}{2} \\ (\langle \mathbf{w} | \mathbf{x}_i \rangle + b) \cdot y_i \geq 1 - \epsilon_i, \forall i \quad (1) \\ \epsilon_i \geq 0, \forall i \end{aligned}$$

La résolution de (1) permet d'obtenir la séparatrice linéaire. Le problème (1) peut se réécrire sous une forme utile à l'implémentation et à une interprétation plus poussée, mais contentons nous d'une description aussi simple que possible. Le premier terme de l'expression à minimiser, couplé aux contraintes qui suivent l'expression, permet l'adéquation de la solution aux données, c'est-à-dire ce qui est le plus important. En effet, les ϵ_i sont les erreurs que s'autorise l'algorithme, au sens où cet algorithme s'efforce de rendre les valeurs prédites f_i égales aux y_i . En pratique, $\epsilon_i = |f_i - y_i|$ si f_i n'est pas du bon côté de y_i , c'est-à-dire inférieure à 1 si $y_i = 1$ ou supérieure à -1 si $y_i = -1$. Ces valeurs -1 et 1 définissent en quelque sorte une marge autour de la droite d'équation $f = 0$ qui est en fait la séparatrice, car en dehors de cette marge, l'erreur n'est pas comptabilisée pour peu que le point \mathbf{x}_i tombe tout de même du bon côté de la séparatrice.

Ces points qui sont soit du mauvais côté de la séparatrice, soit dans la marge, sont appelés vecteurs supports, d'où le terme anglais original *support vector machine*. Étant donné que seuls ces vecteurs supports contribuent à l'erreur cumulée (le premier terme dans l'expression à minimiser dans (1)), l'algorithme opère en minimisant le nombre de vecteurs supports et donc en minimisant la largeur de la marge, ce qui a pour conséquence de rendre \mathbf{w} très grand. Ainsi, les points qui sont du bon côté de la séparatrice sont aussi à l'extérieur de la marge (ou juste à la limite pour les points les plus proches de la séparatrice).

Pour modérer cet effet de minimisation de la marge, un second terme intervient dans l'expression à minimiser dans (1). Ce terme est dit de régularisation, car il s'oppose à l'effet du premier terme de l'expression, si bien d'ailleurs qu'il est à l'origine de la traduction française de SVM. Grâce à ce terme, l'algorithme peut produire une solution unique s'il s'en présentait plusieurs qui offraient un même minimum de l'erreur d'adéquation aux données. De plus, dans le cas d'une séparatrice non linéaire (de forme plus complexe), la complexité de cette séparatrice est minimisée dans le même temps. Ce dernier point est ici moins important à comprendre car il n'est pas considéré dans cette section, ce dès lors que les notations avaient été posées.

Pour simplifier à l'extrême le fonctionnement du SVM au regard de ce qui a été décrit, dans un domaine de fonctionnement raisonnable, les contraintes combinées au premier terme permettent un ajustement « grossier » de la séparatrice par minimisation de l'erreur de séparation (en considérant tous les points), puis le second terme permet un ajustement plus fin en ne considérant cette fois-ci que les points les plus problématiques (les vecteurs supports). Ce second terme est un terme de régularisation, c'est-à-dire qu'il s'oppose au premier terme, ce grâce à l'introduction de la notion de marge. L'heuristique de la régularisation a fait ses preuves sur le plan expérimental.

Enfin, l'existence de deux termes dans une expression à optimiser nécessite de pondérer un terme par rapport à l'autre. C'est le rôle du paramètre c dans le problème (1). Ce paramètre est à optimiser, et pour cela plusieurs techniques existent, mais impliquent l'usage d'une procédure de validation, ce qui est l'objet de la section suivante.

Au sujet de ce paramètre, un cas particulier amène le processus de classification à un certain niveau de trivialité, lorsqu'il est possible de séparer parfaitement les deux classes. Dans ce cas, les ε_i sont tous nuls, et le terme d'adéquation aux données n'a plus d'intérêt, car les contraintes du problème (1) sont toutes satisfaites. Le réglage du paramètre c n'est alors plus nécessaire, ce dernier n'ayant plus de raison d'être.

Dans ce qui suit, seul le cas d'un classifieur discriminant est considéré, par simplicité, étant donné qu'un classifieur modélisant peut se réduire à un classifieur discriminant, la détermination d'une séparatrice étant triviale une fois que les régions ont été définies.

4.3. Validation et mesures de pouvoir discriminant

a) Évaluation de la classification

Si un point tombe du bon côté de la séparatrice, c'est-à-dire dans la région de la classe à laquelle il appartient réellement, on dit qu'il est bien classé ; et s'il tombe du mauvais côté, on dit qu'il est mal classé. Ainsi, si l'on considère un groupe de points, on peut comptabiliser le nombre de points bien classés. Le rapport de ce nombre sur le nombre total de points est appelé taux de bonne(s) classification(s), et est l'indicateur le plus fréquemment utilisé pour quantifier la précision d'une discrimination. Plus le taux de bonnes classifications est élevé, meilleure a été la classification.

Pour donner du sens à un taux de bonnes classifications, on peut le comparer à d'autres taux de bonnes classifications obtenus dans des conditions de référence, ou le comparer à un niveau dit de la chance. En effet, on peut imaginer un classifieur qui déciderait de la classe d'un individu statistique de manière arbitraire, sans connaissance aucune des informations que l'on a sur cet individu (classer un point sans tenir compte de sa position). Si l'on a deux classes avec autant d'individu dans l'une que dans l'autre, un tel classifieur atteindrait tout de même 50 % de bonnes classifications. Ainsi, un taux de bonnes classifications doit être significativement supérieur à cette valeur que le hasard saurait obtenir pour prétendre que le détecteur puisse être efficace.

Une autre valeur de référence parfois plus sévère que la précédente est la proportion que représente la classe majoritaire. Si les classes sont équilibrées (même nombre d'individus par classe), cette valeur est égale à la précédente. Elle est supérieure en revanche dans le cas contraire. Un détecteur peut obtenir ce taux de bonnes classifications s'il donne toujours la même réponse en faveur de la classe

majoritaire. La seule information exploitée par un tel détecteur est la connaissance de la classe majoritaire ; il est donc agnostique de l'état courant représenté par le point \mathbf{x}_i .

Enfin, dans une perspective de détection temps-réel, le détecteur permet de suivre l'état de fatigue à intervalle de temps régulier. Cet intervalle de temps est défini par la durée de la fenêtre de temps évoqué précédemment, sur laquelle le signal est quantifié. Si l'on a la possibilité d'évaluer le détecteur pour différentes tailles de fenêtres, on pourrait vouloir intégrer ce paramètre de latence à l'évaluation. En effet, la détection, quelque soit l'état, a lieu à la fin de chaque fenêtre de temps et parfois des contraintes temporelles limitent le délai avec lequel on peut se permettre d'attendre une réponse du détecteur. De plus, si le détecteur retourne des réponses à un rythme plus élevé que nécessaire, il est possible de cumuler ces réponses pour n'en obtenir qu'une seule qui sera plus fiable.

Il existe ainsi une dualité entre le taux de bonnes classifications et l'unité de temps nécessaire à chaque prise de décision du détecteur. La notion de taux de transfert d'information a été introduite (ITR, de l'anglais *information transfert rate*) et plusieurs formules ont été proposées. Celle qui serait la plus adéquate serait la suivante (Nykopp, 2001 ; Kronegg *et al.*, 2005) :

$$ITR[bits.s^{-1}] = \frac{H(\hat{Y}) - H(\hat{Y}|Y)[bits.décision^{-1}]}{T[s.décision^{-1}]}$$

$$H(\hat{Y}) = -\sum_{i=1}^2 p(\hat{y}_i) \cdot \log_2(p(\hat{y}_i))$$

$$H(\hat{Y}|Y) = -\sum_{i=1}^2 \sum_{j=1}^2 p(y_i) \cdot p(\hat{y}_j|y_i) \cdot \log_2(p(\hat{y}_j|y_i))$$

Où \hat{Y} représente les étiquettes de classes estimées (prédites par le détecteur), Y représente les vraies étiquettes que le détecteur avait à retrouver, les $p(\cdot)$ sont des probabilités que l'on peut estimer à partir des matrices de confusions et T est la durée de la fenêtre de temps utilisée.

Cette notion d'ITR est d'usage dans la communauté des interfaces cerveau-machine.

b) Principe de la validation

L'objectif de la classification est d'obtenir un taux de bonnes classifications aussi élevé que possible. Le fait est que l'on construit un classifieur avec certaines données, disponibles avant sa « mise en production » en tant que détecteur, et que la performance d'un détecteur s'évalue plutôt à la longue, c'est-à-dire sur de nouvelles données acquises ultérieurement.

Parallèlement à cela, les données utilisées pour la construction du classifieur ne fournissent qu'une information partielle sur les régions théoriques représentatives des classes. Par exemple, si l'on imagine des régions convexes, on peut espérer ne pas commettre beaucoup d'erreurs au cœur de ces régions, mais la prise de décision devient plus incertaine aux frontières. Or, la construction du classifieur implique l'ajustement de paramètres, souvent en grand nombre, ce qui offre assez de

souplesse à l'algorithme de construction du classifieur pour s'adapter aux formes de ces régions ; assez de souplesse... Ou trop de souplesse. Lorsque le classifieur modélise des motifs trop fins, propres aux données utilisées pour sa construction, mais pas à d'autres données semblables sans être identiques, on parle de sur-apprentissage (une illustration en est donné figure 6).

Certains classifieurs intègrent des stratégies pour lutter contre ce phénomène, comme décrit précédemment pour le SVM, mais elles ont une efficacité limitée et n'offrent pas d'estimation de l'impact de ce phénomène.

Introduisons le vocabulaire usité (à nouveau pour certains termes). Les données utilisées pour la construction du classifieur sont appelées données d'apprentissage, du fait que le processus de construction est appelé apprentissage. Les données supplémentaires sont appelées données de test ou données de validation. Les procédures qui exploitent l'idée d'une estimation de la performance du classifieur sur des données autres que celles qui ont servi à l'apprentissage sont appelées procédures de validation.

c) La validation croisée

Il existe plusieurs de ces procédures. L'idée centrale est invariablement de découper un ensemble de données en deux sous-ensembles, un premier qui servira à l'apprentissage, et un second pour tester ou « valider » ce qui a été appris. Ces procédures ne sont pas spécifiques à la classification. Elles peuvent (doivent ?) être utilisées dès lors que des paramètres sont estimés afin d'aboutir à un modèle qui est censé représenter plus que les simples données utilisées pour l'estimation.

La procédure de validation employée dans les études menées au cours de cette thèse est appelée validation croisée. Elle consiste en la « découpe » de l'ensemble des données disponibles en un nombre donné (soit N ce nombre) de sous-ensembles approximativement de mêmes effectifs. On réalise N itérations en prenant tour-à-tour l'un des N sous-ensembles comme ensemble de test ou de validation, et les $N - 1$ autres sous-ensembles recombinaés pour former l'ensemble d'apprentissage. Lors de chaque itération, un classifieur est appris sur les données d'apprentissage, puis un taux de bonnes classifications est calculé sur les données de test (ou de validation), à l'aide du classifieur appris. Enfin, on moyenne les N taux de bonnes classifications obtenus en test (ou en validation) pour obtenir l'estimation de la performance en classification que l'on peut espérer obtenir avec l'ensemble de la chaîne de traitements utilisée, incluant les prétraitements sur les informations d'entrée, les quantifications extraites, et l'algorithme de classification.

4.4. Support d'interprétation

L'interprétation des taux de bonnes classifications a déjà été décrite. Ils ont pour utilité de donner une indication sur le pouvoir en détection qu'offre les signaux et quantifications utilisés en combinaison de l'algorithme de classification.

Nos esprits scientifiques seraient bien déçus s'ils devaient se limiter à ces simples taux de bonnes classifications. Un classifieur, une fois « appris », permet de décrire chacune des classes dans l'espace de mesures (espace des quantifications) ou ce qui différencie les classes les unes des autres, toujours dans cet espace de mesures, d'où à un champ ouvert à l'interprétation. Un SVM linéaire est par exemple un hyperplan de séparation entre deux classes. Pour rappel, au cours de la thèse les deux classes sont toujours censées représenter l'une l'état « pas fatigué » et l'autre l'état « fatigué ». Le vecteur orthogonal à l'hyperplan s'inscrit dans l'espace de mesures, et on peut considérer sa projection sur chacune des dimensions de l'espace comme la contribution relative des quantifications correspondantes. En effet, comme illustré par la figure 8 avec un exemple de deux classes et un espace de description à deux dimensions, plus cette projection est grande, plus la quantification correspondante est déterminante dans la classification. Le choix de la lettre w provient du mot anglais *weight(s)*, poids.

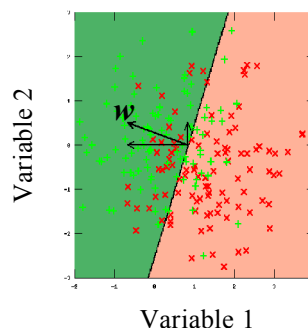


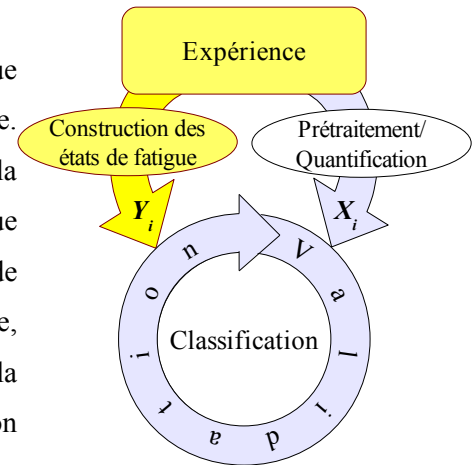
Fig. 8 – Hyperplan séparateur et contribution des variables individuelles au vecteur orthogonal à l'hyperplan. On constate que le projeté de w sur l'axe des abscisses est plus grand que le projeté sur l'axe des ordonnées. L'axe des abscisses est donc plus discriminant, c'est-à-dire que l'on séparerait mieux les deux classes avec une séparatrice verticale dont on choisirait l'abscisse, qu'avec une séparatrice horizontale.

Cependant, l'usage de cette propriété pour l'interprétation du rôle des variables est peu fréquent dans la littérature. À notre connaissance, seuls Mourão-Miranda et collaborateurs ont proposé cette interprétation (Mourão-Miranda *et al.*, 2005).

5 – Axes de recherche

5.1. L'approche générale

Dans ce travail, le but applicatif est lié à la détection de la fatigue objective, c'est-à-dire à la prévention des chutes de performance. Cependant, pour bien des tâches, on ne sait pas évaluer la performance autrement que par la fréquence des erreurs, et lorsque la première erreur peut être fatale, l'anticipation des défaillances de l'opérateur nécessite un autre critère. C'est là que la fatigue mentale, telle que conçue populairement, paraît intéressante. Comme la sensation subjective de fatigue semble être un signal d'alarme, on peut penser que l'opérateur aurait eu l'occasion de sentir une fatigue



avant de commettre une erreur. La détection de la fatigue subjective devient alors le nouvel enjeu. La fatigue subjective, de part sa nature, se mesure en interrogeant l'opérateur sur son niveau de fatigue. Deux problèmes se posent alors. D'une part, il faudrait que cette interrogation se fasse à une fréquence suffisamment élevée et qu'en même temps, elle ne cause aucune gêne à l'opérateur dans la réalisation de la tâche. D'autre part, l'éducation et l'environnement social des uns et des autres amènent certains d'entre nous à vouloir parfois tricher sur les rapports de fatigue, par exemple en minimisant cette dernière, volontairement ou involontairement, ou à l'inverse en l'invoquant avec exagération pour justifier d'une activité insuffisante. Un individu adapte son comportement en fonction des conséquences que ce dernier peut avoir. Si par exemple un pilote s'imagine que ses chances de promotion seraient menacées s'il se montrait trop souvent fatigué, alors il tolérera un niveau plus élevé de fatigue, donc de risque. De plus, on ignore la précision des auto-évaluations dynamiques de fatigue. Finalement, une mesure objective est nécessaire, toujours afin d'anticiper les défaillances de l'opérateur. Le comportement de ce dernier pourrait faire un bon prédicteur mais il n'est pas toujours aisé d'observer le comportement de manière automatique, ni de le décrire sous la forme de signaux continus qui sont mieux adaptés aux outils d'analyse actuels. Une autre solution, celle choisie ici, est d'utiliser des mesures de l'activité de l'organisme, en espérant y trouver des corrélations avec la fatigue.

Le principal inconvénient de l'approche physiologique est le manque de connaissances dont on connaîtrait suffisamment bien le domaine de reproductibilité, pour guider les recherches. En effet, l'essentiel des connaissances dans ce domaine a été obtenu par inférence. Des conditions expérimentales données ont permis une observation donnée ; la reproduction de celle-ci ne signifie pas que les conditions expérimentales initiales aient été réunies. Il faut normalement beaucoup de temps à

la communauté scientifique pour arriver à la conclusion que telle observation constitue un marqueur fiable de fatigue mentale plutôt que d'un phénomène parent, au fil des études qui reproduisent cette observation en modulant telle ou telle condition expérimentale.

Il faut donc induire l'état de fatigue dans un contexte expérimental, d'une part suffisamment contrôlé pour savoir quand il y a fatigue et quand il n'y a pas fatigue, d'autre part suffisamment semblable au contexte d'application visé pour espérer une reproductibilité à ce second contexte des corrélats de la fatigue mentale extraits du premier contexte.

Nous avons eu l'opportunité de mettre en place un tel protocole, impliquant des élèves pilotes au cours d'une simulation de vol. Nous avons également analysé les données de deux expériences de laboratoire sans lien véritable avec la tâche de pilotage d'aéronefs, dans l'espoir d'apporter des observations supplémentaires au corps des études du même genre.

Les grandes lignes de la démarche exploitée dans cette thèse sont les suivantes. D'un côté on s'appuiera sur les mesures subjectives et comportementales pour définir tant bien que mal le niveau de fatigue mentale sur des périodes de temps prédéfinies dans un contexte expérimental qui facilite cette définition. De l'autre côté, on cherchera les motifs de l'activité physiologique corrélés avec la fatigue mentale telle que construite.

5.2. Objectifs des études

a) Première étude : classification multi-modalités

Dans la recherche de marqueurs physiologiques qui permettent la détection de la fatigue mentale, nous nous sommes principalement appuyés sur l'EEG. Cependant, les activités oculo-palpébrale et cardiaque ont fait l'objet de plus d'études, en particulier chez les conducteurs de la route pour la première et chez les pilotes d'aéronefs pour la seconde. Ces techniques d'enregistrement sont plus aisées d'installation et offrent une fenêtre d'observation sur l'activité d'organes dont les variations au cours des tâches citées sont couramment rapportées par les opérateurs eux-mêmes. L'interprétation des observations faites est souvent plus immédiate que pour l'EEG.

Il s'agissait donc pour nous de comparer les trois techniques (ou modalités) d'enregistrement évoquées, à savoir l'EEG, l'EOG et l'ECG, pour évaluer l'intérêt de chacune prise indépendamment, mais aussi leur complémentarité dans le cas d'une combinaison de leurs quantifications respectives.

b) Deuxième étude : reconstruction de sources

Nous avons mené une seconde étude dans laquelle nous nous sommes intéressés seulement à l'EEG et en particulier aux synchronies à distance. De la littérature, il ressortait que les manifestations de la fatigue étaient assez variables d'une tâche à une autre. Dans l'idée de développer une procédure

applicable à plus d'un contexte d'étude ou de détection de la fatigue mentale, il nous est apparu pertinent de tenter d'accroître la résolution spatiale de l'EEG en espérant suivre ainsi l'activité de réseaux fonctionnels dont la finesse de la structure spatiale serait à la limite de ce que l'on peut normalement appréhender avec l'EEG. Pour cela, nous avons employé des techniques de reconstruction des sources corticales à l'origine des signaux électriques mesurés sur le scalp.

Nous souhaitions tester la faisabilité d'une analyse qui combinait reconstruction des sources corticales, mesures de synchronies à distance et classification. Pour cela, une simulation paraissait adéquate. De plus, nous avons exploité les données d'une expérience EEG constituée d'une tâche de compatibilité spatiale (test de Simon) pour laquelle nous avons construit des états de « fatigue » basés sur la performance instantanée, plus précisément le temps de réponse moyen, et non plus sur le temps passé à la tâche, la fatigue subjective (non mesurée) ou la performance moyenne sur la durée d'un bloc d'essais.

c) Troisième étude : simulation de vol

La troisième étude présentée dans ce document avait pour objectif l'appréhension d'un contexte d'application réelle, ceci initialement afin de reproduire des résultats précédemment obtenus en laboratoire, entre autres buts. Le contexte choisi était celui d'une simulation de vol à l'intérieur d'un cockpit d'Alphajet converti en simulateur, avec le concours de pilotes en tant que sujets.

La liberté qu'apportait cette expérience, du fait de la durée de la simulation et de la relative complexité de la tâche de pilotage, a été l'occasion de chercher une manière différente de poser la problématique de la fatigue mentale, une manière de guider l'analyse par une proposition initiale d'application d'un détecteur de fatigue mentale. Dans ce qui est présenté, les questions concernant l'électrophysiologie, telles celles posées dans les deux études précédentes, ont été finalement comme reléguées à plus tard, et l'enjeu était ici plutôt de présenter des résultats en classification qui puissent revêtir un sens plus accessible à tout un chacun.

La principale question posée était : « peut-on prédire l'évolution générale du niveau de fatigue mentale vers la fin de la simulation à partir d'informations disponibles dans une partie antérieure de la simulation ? »

d) Structure

La suite de ce chapitre donne des éléments d'information supplémentaires et inventorie en particulier les points de méthode communs aux différentes études. Ensuite, les trois dernières sections de ce chapitre donnent chacune un résumé des trois études.

5.3. Matériel, prétraitements et quantifications

a) Matériel et sujets

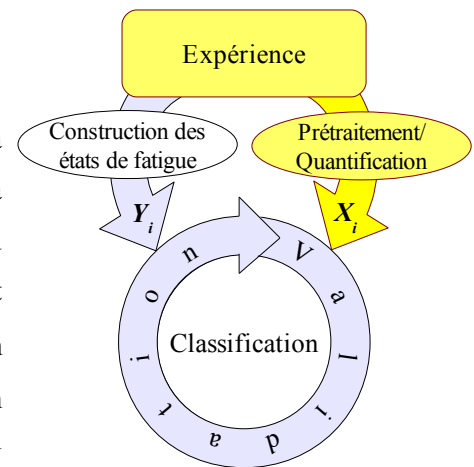
Quelle que fut l'expérience, chaque sujet était équipé d'un casque à 32 ou 60 électrodes EEG, placées suivant le système 10-20. La totalité du matériel utilisé pour l'acquisition des signaux a été conçu par la société Brain Products. L'impédance des électrodes était rendue inférieure ou égale à 10 k Ω . Les électrodes EEG, en comprenant la référence et la terre, étaient reliées à un amplificateur Brain Amp, tandis que les électrodes non-EEG étaient connectées à un amplificateur différentiel Brain ExG MR, sauf indication contraire. Ces autres électrodes étaient en général au nombre de 9, dont 2 électrodes d'électromyographie (EMG) placées sur la partie inférieure du masséter, 2 électrodes d'EOG horizontal placées à gauche de l'œil gauche et à droite de l'œil droit, 2 électrodes d'EOG vertical placées au dessus et en dessous de l'œil dominant, et enfin 3 électrodes d'ECG, dont une commune aux deux dérivations, placée au bas des côtes sur le côté gauche, une au dessus de l'omoplate gauche et une au dessus de la clavicule droite.

L'enregistrement de l'EMG avait pour seule fonction de permettre à l'expérimentateur de dépister au cours de l'expérience un comportement inapproprié du sujet, marqué par des contractions excessives ou trop régulières du muscle sous-jacent aux électrodes, et le cas échéant, pour que l'expérimentateur puisse demander au sujet de se détendre, à l'occasion des pauses. Le signal enregistré n'a pas été analysé.

Les signaux étaient échantillonnés soit à 500 Hz, soit à 1 kHz, et filtrés lors de l'acquisition entre 0,1 Hz et la fréquence maximale (la moitié de la fréquence d'échantillonnage).

b) Fenêtrage

Le signal, quelle que soit la modalité électrophysiologique, était découpé en fenêtres de temps de durée fixe. Ces fenêtres étaient glissantes, avec un recouvrement entre fenêtres successives de 50 %. Dans les expériences de laboratoire pour lesquelles la tâche était une succession d'essais, les fenêtres n'étaient donc pas calées d'une quelconque manière aux événements propres à la tâche (stimulus ou réponse), à l'exception de la première fenêtre de chaque bloc qui démarrait à la première stimulation, marqueur du début de la tâche. Cette première fenêtre était ensuite rejetée pour ne pas prendre en compte l'activité physiologique peu stable du sujet lorsque celui-ci est en train de se réappropriier le rythme de la tâche.



c) Quantification de l'EOG et de l'ECG

L'EOG vertical était filtré entre 1 et 5 Hz puis une procédure de détection des clignements a été appliquée afin d'aider au marquage visuel des clignements dans l'EOG par une même personne. Une fois les clignements localisés, leur position, leur amplitude et leur durée étaient estimées pour pouvoir calculer sur chaque fenêtre de temps les quantifications inventoriées ci-après avec entre parenthèses les abréviations utilisées :

- le nombre de clignements (BN),
- l'amplitude moyenne des clignements (BA(m)), ou 0 s'il n'y avait aucun clignement,
- la durée moyenne des clignements (BD(m)), ou 0 s'il n'y avait aucun clignement,
- l'écart-type de l'amplitude des clignements (BA(sd)), ou 0 s'il y avait moins de 1 clignement,
- l'écart-type de la durée des clignements (BD(sd)), ou 0 s'il y avait moins de 1 clignement,
- l'écart-type des intervalles entre clignements successifs (BI(sd)), ou 0 s'il y avait moins de 2 clignements.

Une des deux voies d'ECG, celle présentant des ondes R les plus amples, était filtrée entre 5 et 15 Hz ou, pour l'étude de simulation de vol, entre 5 et 25 Hz. Cette différence provient d'une appréciation subjective de l'apparence du signal filtré, en particulier de l'amplitude de l'onde R par rapport à celles des autres ondes. Une procédure de détection des ondes R a été ensuite appliquée afin d'aider au marquage visuel des ondes R dans l'ECG par une même personne. La position des ondes R a permis la construction du signal R-R en ré-échantillonnant à 4 Hz après interpolation par splines cubiques, puis le calcul des quantifications suivantes pour chaque fenêtre de temps :

- la fréquence cardiaque (HR),
- l'écart-type des intervalles R-R (RR(sd)),
- l'énergie dans la bande 0,04-0,15 Hz du signal R-R (LF),
- l'énergie dans la bande 0,15-0,4 Hz du signal R-R (HF),
- le rapport des énergies LF sur HF (LF/HF),
- l'énergie du signal R-R dans la bande 0,09-0,11 Hz (0.1Hz).

d) Quantification de l'EEG

L'EEG a tout d'abord fait l'objet d'une correction des artefacts dus aux clignements oculo-palpébraux. Pour cela, les clignements ont été détectés puis un segment d'EEG de 2 s centré sur le pic du clignement a été décomposé en composantes singulières. Les composantes corrélées à plus de 0,7 avec l'EOG vertical ont été rejetées puis l'EEG a été reconstruit à partir des composantes restantes (Wallstrom *et al.*, 2004).

Les segments d'EEG contaminés par des activités vraisemblablement d'origine musculaire ont été rejetés. Ce marquage, essentiellement manuel, a été facilité par l'usage d'une procédure de détection de transition de l'état « EEG propre » à l'état « EEG contaminé » (et inversement) de type « somme cumulée » (du logarithme du rapport de vraisemblance), *cusum* en anglais, les états étant représentés par des portions d'EEG choisies.

Le signal a ensuite été filtré à l'aide d'un filtre à réponse impulsionnelle finie ajusté pour laisser passer les fréquences entre 0,1 et 100 Hz, puis le signal a été décimé à 250 Hz.

À ce stade, la référence a été modifiée pour une référence moyenne. Le calcul de cette référence moyenne impliquait les signaux des différentes électrodes EEG à l'exception des électrodes composant la couronne la plus périphérique, hors électrodes occipitales. La figure 9 marque les électrodes exclues du calcul de la référence moyenne.

Cette sélection a pour but de limiter l'étendue spatiale de la contamination de l'EEG par l'activité des muscles péricrâniens particulièrement visible sur les électrodes périphériques. En effet, le passage en référence moyenne permet de gommer le potentiel de la référence matérielle, potentiel normalement visible en copies conformes sur toutes les électrodes. L'estimation de ce potentiel de référence ne nécessite pas forcément toutes les électrodes, d'autant plus si un sous-groupe d'électrodes est le siège d'une activité fortement corrélée et à grande amplitude, susceptible de contribuer significativement à la moyenne sans présenter la répartition uniforme du signal de la référence dans l'EEG.

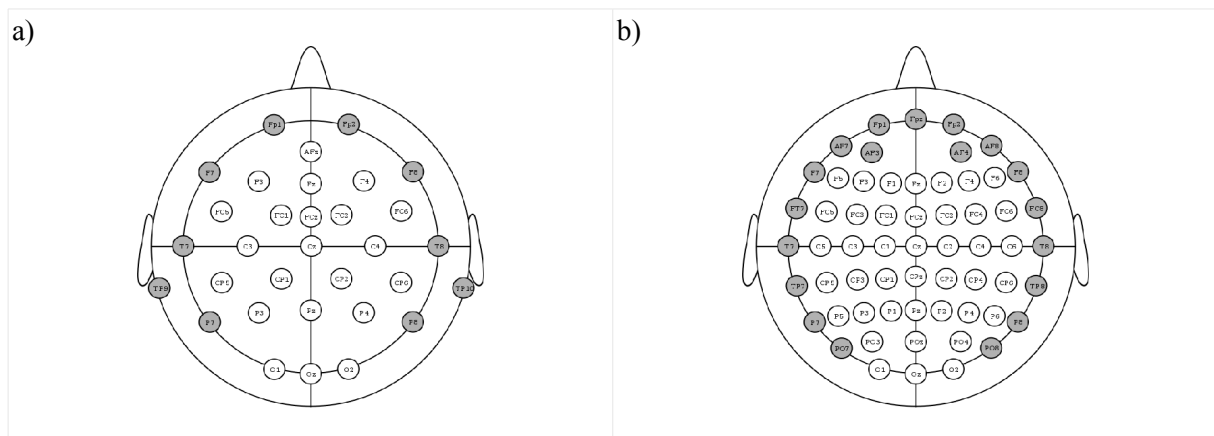


Fig. 9 – Électrodes exclues du calcul de la référence moyenne (en grisé). a) Casque à 32 électrodes utilisé pour les expériences **classification multi-modalités** et **simulation de vol**. b) Casque à 60 électrodes utilisé pour l'expérience **reconstruction de sources** (illustrations empruntées à Marius 't Hart⁵ et modifiées).

5 Sous les termes de la licence Creative Commons dite By-Sa décrite à l'adresse suivante : http://creativecommons.org/licenses/by-sa/3.0/nl/deed.en_GB (lien fourni par l'auteur)

Les quantifications employées ont été, selon l'étude, l'amplitude moyenne dans les bandes de fréquences 3-7 Hz, 7-13 Hz et 13-18 Hz, la cohérence dans les mêmes bandes de fréquences et la « valeur de verrouillage de phase » ou *phase locking value* (PLV), également dans ces trois bandes de fréquences.

L'amplitude moyenne est la moyenne des amplitudes instantanées estimées par transformation de Hilbert, au cours de la période d'intégration (de la fenêtre de temps).

La cohérence se calcule quant à elle à partir des spectres croisés notés S (Franaszczuk *et al.*, 1985) :

$$Coh_{xy}^f = \frac{|S_{xy}^f|}{\sqrt{S_{xx}^f \cdot S_{yy}^f}}$$

$$S_{xy}^f = \sum_n TF^f\{\mathbf{x}_n\}^t \cdot TF^f\{\mathbf{y}_n\}$$

TF est ici la transformation de Fourier, restreinte à un domaine fréquentiel f . Le spectre croisé entre les signaux x et y a été estimé par périodogramme, c'est-à-dire en découpant les fenêtres de temps en sous-fenêtres représentées par des vecteurs \mathbf{x}_n et \mathbf{y}_n . Les spectres croisés ont été calculés d'abord à une résolution fréquentielle plus fine que nécessaire, puis les composantes spectrales croisées de fréquence comprise dans la bande considérée ont été sommées.

La valeur de verrouillage de phase (nous utiliserons l'abréviation PLV par la suite) se calcule à l'aide des phases et donc ici par l'intermédiaire des signaux analytiques des deux signaux (Lachaux *et al.*, 1999) :

$$PLV_{xy}^f = \frac{1}{T} \left| \sum_t e^{\phi_x(t) - \phi_y(t)} \right|$$

Pour une période d'intégration T (durée de la fenêtre de temps), en tout échantillon d'instant t .

L'amplitude moyenne est une mesure de synchronie locale, tandis que la cohérence et la PLV sont des mesures de synchronie distante.

Un commentaire important sur ces mesures porte sur la contribution des amplitudes et celle des phases. La cohérence exploite à la fois information d'amplitude et de phase, même si il y a normalisation du spectre croisé par les énergies individuelles des signaux. La PLV exploite quant à elle uniquement l'information de phase et est donc susceptible d'apparaître à l'usage plus complémentaire des mesures de synchronie locale.

e) Reconstruction de sources

La reconstruction de sources en EEG (pour revue, lire Baillet, 2001) s'appuie sur la distribution de la matière grise dans le cerveau. Cette matière grise contient l'essentiel des corps cellulaires des neurones ainsi que leurs dendrites. Pour la localiser, on emploie l'imagerie par résonance magnétique (IRM), plus précisément une image dite T1, et une série d'algorithmes comme ceux fournis par le logiciel

Brainvisa⁶. Nous n'avons retenu que le cortex cérébral, car les neurones des structures profondes sont plus souvent étoilés, avec des arbres dendritiques étendus dans toutes les directions, si bien que les champs électromagnétiques émis par leurs courants s'annulent.

La reconstruction de sources s'opère en deux étapes. La première est la modélisation de la contribution des courants électriques du cortex cérébral aux potentiels électriques mesurés sur le scalp. Pour cela, on réduit le cortex cérébral à une surface, par simplification, puis on segmente cette surface en un nombre éventuellement grand de petites régions dont l'activité électrique résultante est modélisée par un dipôle électrique d'orientation connue (définie par la forme locale de la surface corticale). Dans la pratique, la surface corticale est représentée par un maillage, en général celui de l'interface entre matière blanche et matière grise, en assimilant chaque sommet de ce maillage au centre d'une de ces régions, avec donc un dipôle en chaque sommet. La propagation du signal porté par ces dipôles est ensuite modélisée en considérant les différents milieux séparant ces dipôles des électrodes. Il y a les méninges avec principalement le liquide céphalo-rachidien (LCR) qu'elles contiennent, le crâne, et la peau (le scalp). Le LCR et le scalp ont une conductivité élevée, à la différence du crâne qui agit comme une barrière. Par approximation, on considère tous ces milieux comme isotropes, ce qui n'est pas le cas du crâne en pratique. Par conséquent, la conductivité de chacun des milieux est le principal paramètre qui régit la propagation du signal, outre bien entendu l'étendue spatiale des milieux.

Cette première étape de modélisation a été réalisée ici à l'aide d'un algorithme appelé méthode des sphères en recouvrement (*overlapping spheres* en anglais), développé par Huang *et al.* (1999), et disponible dans la boîte à outils BrainStorm⁷. Comme toute méthode, elle ajoute son lot d'approximations au modèle, avec principalement la considération de seulement deux milieux : celui à l'intérieur d'une sphère approximant la surface intérieure du crâne, et celui à l'extérieur. L'usage d'une sphère simplifie considérablement les calculs par rapport aux méthodes qui traitent les maillages tels quels en échappant en plus à certains problèmes numériques que rencontrent ces dernières méthodes. En revanche, c'est une approximation spatialement médiocre en certaines régions du crâne. Étant donné que la modélisation de la propagation du signal peut être résolue indépendamment pour chaque électrode, on réalise la procédure d'approximation électrode par électrode ; on adapte la sphère de manière à ce que la surface interne du crâne soit représentée de façon réaliste au voisinage de l'électrode et donc aussi des sources qui contribuent pour l'essentiel du signal reçu par cette électrode. La simplification qui consiste à ne considérer que la surface interne du crâne est appelée « approche du crâne isolant » (*insulated skull approach* en anglais, soit ISA en abrégé ; Hämäläinen & Sarvas, 1989 ; Mosher *et al.*, 1999). L'ISA se justifie par le fait que les courants secondaires ne peuvent pas circuler dans le crâne, et qu'ils s'arrêtent donc à sa surface interne, pour ceux originaires du milieu

6 Disponible sur le site <http://brainvisa.info>

7 Disponible sur le site <http://neuroimage.usc.edu/brainstorm>

intra-cranien. Le potentiel en revanche se propage au delà, et de nouveaux courants « secondaires » peuvent naître dans le scalp du fait des différences de potentiel et de la conductance moins basse de la peau, bien qu'on les néglige également.

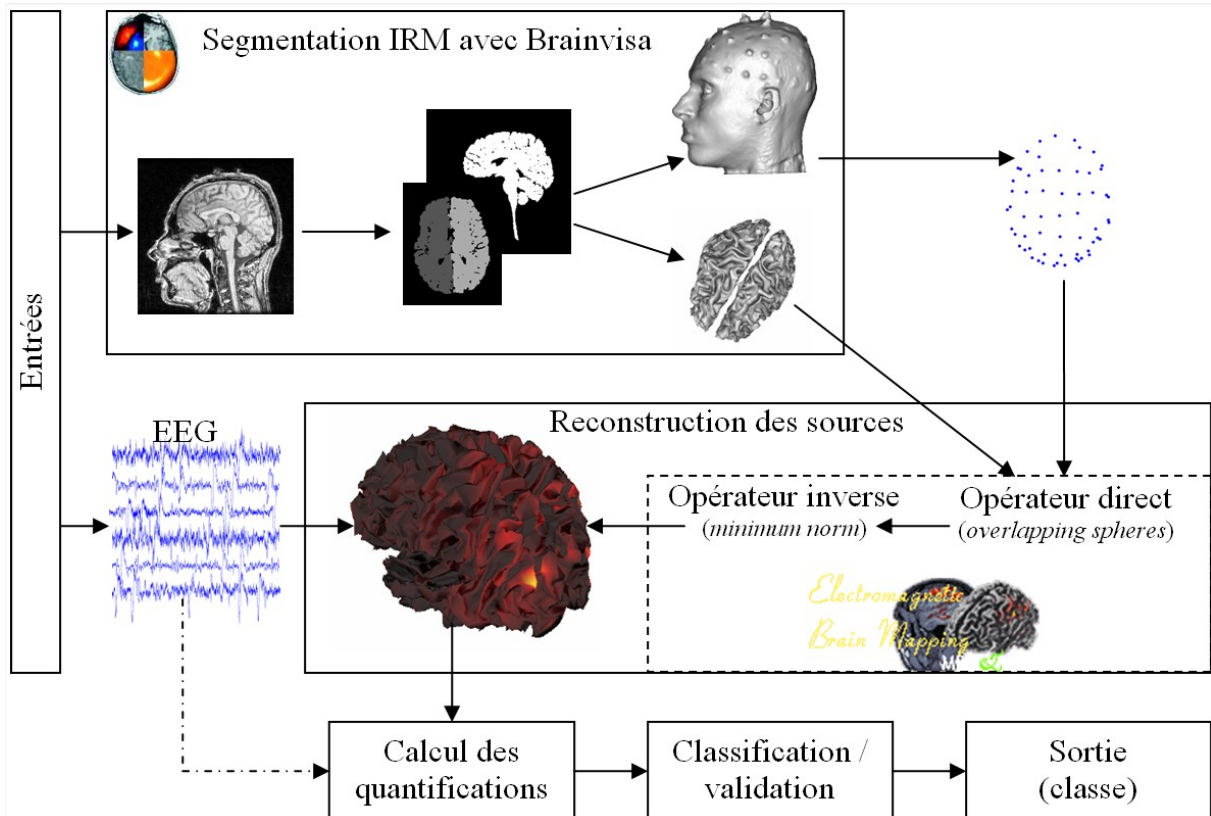


Fig. 10 – Principe de la reconstruction et intégration dans la chaîne de traitements. En plus des signaux EEG prétraités à « projeter » sur la surface corticale, l'IRM du sujet est segmenté pour en extraire un modèle anatomique de la distribution des sources. Dans cette expérience, la position des électrodes a été également extraite de l'IRM, grâce à un marquage du gel électrolytique appliqué entre le scalp et chaque électrode tel que les patches de gel apparaissent sur l'IRM. La segmentation des maillages utiles à la détermination de la position et de l'orientation des dipôles électriques de courant (nos sources) a été réalisé à l'aide du logiciel Brainvisa. Le calcul des opérateurs direct et inverse a été mené à l'aide de la boîte à outils Matlab BrainStorm. Les quantifications usuelles peuvent ensuite être calculées sur les sources reconstruites.

Une fois la contribution des sources (les dipôles) au signal des électrodes évaluée sous la forme d'une matrice avec autant de lignes que d'électrodes et de colonnes que de sources, il faut estimer la transformation inverse. C'est la seconde étape de la reconstruction de sources. Nous avons employé

une des méthodes les plus simples qui soient, la solution inverse de norme minimale, ou *minimum norm inverse solution* (MN) en anglais (Hämäläinen & Sarvas, 1989). Pour une matrice « directe » A , la matrice inverse est :

$$G = A'(AA' + \alpha I)^{-1}$$

α est un coefficient de régularisation. Ce coefficient est fixé traditionnellement à 10 % de la valeur singulière maximale de A . Cette proportion sera notée λ et c'est elle que l'on fera varier pour ajuster la régularisation. I est la matrice identité dans l'espace des électrodes. Le terme αI peut être interprété comme une estimation du bruit de mesure (Dale & Sereno, 1993). C'est donc à la fois un paramètre de lissage, car l'ensemble des sources qui contribuent à l'activité d'une électrode augmente à même niveau de contribution seuil lorsqu'on augmente α , ainsi que comme un seuil sur le rapport signal sur bruit.

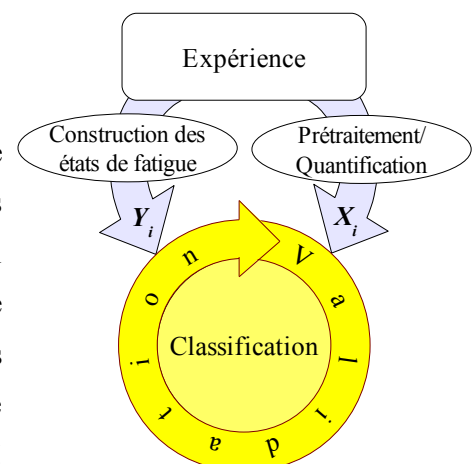
Au final, G est une matrice avec autant de lignes que de sources et de colonnes que d'électrodes, et elle est indépendante de l'activité, donc du temps ; chaque échantillon relevé sur l'ensemble des électrodes peut être projeté sur les n dipôles choisis sans la connaissance des échantillons précédents. n est choisi égal à environ 10000, l'objectif étant de gagner en résolution spatiale (on parle de sources massivement distribuées), mais ce nombre peut être réduit pour limiter le temps de calcul, en particulier lorsqu'on estime des synchronies à distance, ce qui nécessite de considérer tous les couples possibles de sources.

Les prétraitements ont tous été réalisés avant la « projection » de l'activité de scalp sur le cortex cérébral. De plus, une optimisation possible pour le calcul des quantifications au niveau des sources consistait à calculer les signaux analytiques et les spectres croisés sur les signaux EEG avant projection, puis à « projeter » ces grandeurs complexes sur le cortex cérébral. Ceci est possible du fait de la nature à la fois linéaire et constante dans le temps de l'opérateur de reconstruction des sources qui est une simple matrice. Toutes les opérations de filtrage dans le temps ou de manière équivalente dans l'espace de Fourier pouvaient s'appliquer aussi bien avant que après le produit par cette matrice.

5.4. Classification, validation et interprétation

a) Classification

Lors de l'apprentissage du SVM, le paramètre c évoqué précédemment était à régler. Bien souvent, dans les deux premières études, nous avons plus de variables que de fenêtres de temps, si bien que les classes pouvaient être séparées sans erreur et que ce paramètre pouvait être ignoré. Dans les autres circonstances (lorsque le nombre de variables n'excédait pas de le nombre de fenêtres de temps), nous avons toléré la procédure simplifiée qui



consistait à utiliser une même valeur pour c , ce pour tous les apprentissages dont les résultats étaient ensuite fusionnés ou comparés l'un à l'autre. Cette valeur est donnée pour chaque expérience dans la partie correspondante. Elle a été fixée à 1 ou à 0,1, selon les expériences.

Dans la troisième étude, le c a été réglé et la procédure est décrite dans la partie correspondante.

b) Validation

À l'exception de la troisième étude, la validation a été opérée par validation croisée en 10 sous-ensembles de fenêtres de temps. Ces 10 sous-ensembles ont été formés en affectant aléatoirement chaque fenêtre de temps à l'un d'entre eux, de manière à ce que ces sous-ensembles soient d'effectifs équilibrés.

Pour rappel, ce partitionnement en 10 sous-ensembles implique que nous ayons obtenu pour chaque validation croisée 10 taux de bonnes classifications et 10 SVM, donc 10 vecteurs orthogonaux aux hyperplans séparateurs.

Les fenêtres de temps successives avaient un recouvrement entre elles. Nous avons supprimé tout recouvrement de fenêtres entre celles qui servaient à l'apprentissage et celles qui servaient au test en ignorant les fenêtres de l'ensemble d'apprentissage qui avaient un recouvrement avec au moins l'une des fenêtres de l'ensemble de test. Cette procédure réduisait parfois considérablement les taux de bonnes classifications, ce qui justifie pleinement son usage. En effet, l'augmentation observée des taux de bonnes classifications lorsque le recouvrement des fenêtres entre les ensembles d'apprentissage et de test n'était pas traité était très probablement due à une grande ressemblance entre les fenêtres qui partageaient un segment en commun. Or, les données de test doivent être différentes des données d'apprentissage.

c) Interprétation

Le vecteur w du SVM présente un intérêt clair pour la comparaison des quantifications à la suite d'un apprentissage. Une première précaution est à prendre : les variables doivent au préalable avoir été normalisées de manière à ce que le nuage global des points (les deux classes mélangées) ne se dispersent pas plus le long d'une dimension que d'une autre. La normalisation consiste en le centrage et la réduction des données. On opère en divisant chaque variable par son écart-type, car l'écart-type est un estimateur de la dispersion d'une distribution.

Les choses se compliquent lorsqu'on souhaite prendre une moyenne de plusieurs vecteurs, par exemple lorsque l'on a un vecteur par sujet. Dans la pratique, on obtient autant de vecteurs que l'on a réalisé d'apprentissage, soit autant que de sujets multiplié par le nombre d'itération de la validation croisée. Bien que la norme de ces vecteurs soit ajustée sur la base d'un critère propre au SVM, donc *a priori* un même critère pour tous les vecteurs, la dépendance aux données n'est pas résolue à cause de l'existence

d'un autre paramètre. Ce paramètre, désigné par b , est le biais de l'hyperplan, c'est-à-dire, pour une orientation w donnée, la position de l'hyperplan. w et b ont été estimés de manière inter-dépendante, ce qui retourne cette complication contre b , si l'idée venait à l'expérimentateur d'exploiter ce second paramètre. Notre stratégie pour rendre plusieurs w homogènes, pourvu que le nombre de variables soient grands, est de normaliser ces vecteurs. Ainsi, on moyenne les contributions relatives de chaque variable. Dans la pratique, nous avons observé que cette transformation n'entraînait pas de grands changements dans les résultats.

Enfin, nous avons opéré un second raffinement dans ce moyennage de w . Lorsque l'on n'a pas l'opportunité de tester un classifieur sur des données suffisamment différentes pour être assuré de la généralisabilité⁸ des résultats obtenus, par exemple des données acquises sur les mêmes sujets lors d'une seconde session d'enregistrement réalisée en un jour différent, il est intéressant de pouvoir observer ce qui est commun à une majorité de sujets, et ce qui diffère et pourrait donc peut être induire le détecteur en erreur. Ainsi, on observe parfois que certaines variables sont relativement discriminantes pour la plupart des sujets, mais qu'elles ne varient pas dans le même sens d'un sujet à un autre. Dans ce cas, on peut douter de l'intérêt de ces variables. Cette variabilité inter-individuelle n'est pas systématiquement gommée par l'opération de la moyenne, car un déséquilibre d'amplitude relativement grande peut être vite atteint par l'effet du hasard. Nous avons donc décidé de diviser chaque moyenne par l'écart-type correspondant, pénalisant ainsi fortement les variables à forte variabilité inter-individuelle. Cette pénalisation n'a été appliquée qu'à travers les sujets, et pas lors du moyennage des vecteurs d'un même sujet.

d) Statistiques

Tous les tests statistiques ont été menés en prononçant la significativité de l'effet testé lorsque le risque de première espèce était inférieur ou égal à $P=5\%$.

Un test fréquemment utilisé pour comparer deux populations d'échantillons appariés est le test de Wilcoxon (Saporta, 2006). Lorsque de nombreuses comparaisons ont été réalisées (20 ou plus), une correction a été appliquée, soit celle de Bonferroni, soit par contrôle du taux de faux positifs (FDR, de l'anglais *false discovery rate*) telle que implémentée par Benjamini et Hochberg (1995). Idéalement, il aurait fallu réaliser cette correction systématiquement. Ce nombre minimal de 20 comparaisons a été choisi du fait qu'à $P=5\%$, on a en théorie au moins 1 faux positif (contre « moins de 1 » à moins de 20 comparaisons) et que bien souvent, dans les situations où peu de comparaisons sont à faire, le test employé offre une « résolution » en probabilité insuffisante pour qu'un effet puisse rester significatif après correction, quelle que soit la taille de cet effet.

8 Ce mot n'existe pas en soi mais est déjà très utilisé. Il désigne la capacité de généralisation.

5.5. Résumé : classification multi-modalités

Une tâche à choix binaire avec alternance de deux sous-tâches suivant une séquence prédictible (protocole dit de *task switching*) a été réalisée pendant 6 blocs consécutifs d'environ 20 min chacun. Une première analyse du comportement de 13 sujets montrait une augmentation du taux d'erreurs au cours des 4 premiers blocs, suivie d'une relative stagnation pour les 2 derniers blocs. Les scores de fatigue subjective ont également varié dans le sens d'une augmentation de la fatigue.

Pour chaque sujet, le bloc avec le minimum du taux d'erreurs (classe « pas fatigué » ou des « bonnes performances ») a été opposé au bloc avec le maximum du taux d'erreurs (classe « fatigué » ou des « mauvaises performances »). À l'échelle de la population des sujets, le bloc des bonnes performances arrivait avant le bloc des mauvaises performances. La fatigue subjective croissait également pour tous les sujets. Ainsi, trois aspects de la fatigue étaient réunis : la diminution de performance, le temps passé à la tâche et l'augmentation de la sensation de fatigue.

L'analyse des signaux électrophysiologiques a été réalisée par classification sujet-par-sujet des fenêtres de temps du bloc « pas fatigué » *versus* celles du bloc « fatigué ». Le signal a donc été découpé en fenêtres de temps de durée allant de 4 s à 30 s puis quantifié de la manière décrite sections 5.3.c) et 5.3.d) en 6 variables pour l'ECG, 6 pour l'EOG et 96+1488 d'EEG (32 électrodes, 3 bandes de fréquences). Plusieurs procédures de classification-validation ont été réalisées : une avec l'ECG seul, une avec l'EOG seul, une avec les amplitudes moyennes de l'EEG (96 premières variables), une avec les PLV de l'EEG (1488 variables suivantes) et une en groupant toutes les variables.

Le principal résultat en termes de comparaison est la supériorité du pouvoir de détection qu'offrait l'EEG sur ceux que permettent l'ECG et l'EOG. Les PLV de l'EEG étaient plus discriminantes que les amplitudes moyennes, en particulier pour de petites tailles de fenêtres. Les deux modalités extracérébrales, lorsqu'elles étaient ajoutées à l'EEG, permettaient une augmentation, relativement uniforme à travers les différentes tailles de fenêtres, des taux de bonnes classifications de quelques points. Pour donner quelques chiffres, à 4 s, les taux moyens de l'EOG et de l'ECG étaient au niveau de la chance (environ 50 %), celui des amplitudes moyennes de l'EEG était à 71 %, celui des PLV à 89 % et celui de la combinaison des modalités à 93 %. À 30 s, le taux moyen de l'EOG était à 76 %, celui de l'ECG à 85 %, celui des amplitudes moyennes et les PLV à 95 % et celui de la combinaison de modalités à 97 %.

L'étude des contributions individuelles des différentes variables dans la classification, sur la base de toutes les modalités combinées et en segmentant les signaux en fenêtres de 30 s, a apporté des observations pour une part en accord avec certaines de nos attentes au vu de la littérature. Par exemple, les clignements se sont faits plus nombreux dans l'état « fatigué » que dans l'état « pas fatigué » et en même temps moins amples et plus longs. Les amplitudes moyennes de l'EEG ont globalement augmenté en 7-13 Hz et diminué en 13-18 Hz. En revanche, la fréquence cardiaque a

diminué et la variabilité du rythme cardiaque a augmenté. Les quantifications les plus discriminantes en poids relatif étaient des variables d'ECG (fréquence cardiaque, écart-type des intervalles R-R et puissance dans la bande LF), mais de manière générale les variables d'ECG présentaient aussi les plus fortes variabilités inter-individuelles.

La classification des états de fatigue mentale tels que construits a donc été possible, avec des taux de bonnes classifications dépassant les 90 % et des variations physiologiques, exploitées par les classifieurs individuels, que l'on retrouve chez une majorité de sujets.

5.6. Résumé : reconstruction de sources

a) Analyse de données simulées

Deux simulations ont été réalisées à l'aide de l'IRM d'un sujet et du choix de deux régions corticales distantes. La première simulation consistait en une modulation d'amplitude, simultanée mais sans corrélation ajoutée entre les signaux dominants des deux régions. Le niveau de modulation était différent d'une classe à l'autre, représentant ainsi deux états mentaux. La seconde simulation consistait en une modulation de la durée de synchronisation totale entre les signaux dominants des deux régions, sans modifier d'une quelconque manière l'amplitude des signaux.

Les données simulées ont été générées au niveau du cortex cérébral pour être ensuite projetées sur le scalp. Le signal a été segmenté en fenêtres de temps puis quantifié tour-à-tour à l'aide de l'amplitude moyenne, de la cohérence et de la PLV. Une procédure de classification-validation a été réalisée pour chacune des deux simulations et pour les différentes quantifications considérées individuellement au niveau du scalp (60 électrodes), au niveau de 60 sources reconstruites, au niveau de 504 sources reconstruites et au niveau des seuls deux signaux corticaux utiles originaux.

La reconstruction de sources apportait un gain systématique en termes de taux de bonnes classifications, quelle que fut la quantification, quoique non (seulement presque) significatif pour la première simulation. Conformément à nos attentes, les amplitudes moyennes offraient de meilleurs taux de bonnes classifications dans la première simulation et les cohérences et PLV offraient quant à elles de meilleurs taux de bonnes classifications dans la seconde simulation. Les PLV en particulier ont permis de retrouver fidèlement les deux régions d'intérêt, ce qui a nécessité l'usage d'une procédure particulière du fait de la nature distante des couplages quantifiés par ce type de mesures. Le mélange de signaux causé par les projections cortex-scalp puis scalp-cortex a en revanche induit des effets de transfert d'information entre amplitude et phase, si bien que les amplitudes moyennes permettaient une classification au dessus du niveau de la chance dans la seconde simulation, à tous les niveaux sauf avec la paire de signaux originaux. De même, les mesures de couplages permettaient une discrimination d'égale efficacité dans la première simulation, là encore à l'exception des signaux

originaux. Enfin, les taux de bonnes classifications augmentaient avec le nombre de sources et la valeur du coefficient de régularisation pour le calcul de la solution inverse de norme minimale.

b) Analyse de données EEG réelles

Une tâche de compatibilité spatiale, dans laquelle le sujet devait répondre à la direction (gauche ou droite) d'une flèche (le stimulus) sans tenir compte de la position (également gauche ou droite) de cette même flèche, a été réalisée sur une durée totale de 4x10 min (4 blocs). Une première analyse du comportement des 12 sujets a montré l'effet Simon attendu (temps de réponse plus long pour les essais avec incompatibilité entre la caractéristique non pertinente du stimulus et la réponse à donner elle-même fonction de la caractéristique pertinente du stimulus), ainsi que de multiples effets d'interaction entre un essai et l'essai suivant sur le temps de réponse, pour différents paramètres (principalement compatibilité stimulus-réponse et position du stimulus).

Pour chaque sujet, le temps de réponse a été corrigé pour les effets observés et le signal a été segmenté en fenêtres de 20 s. Un temps de réponse moyen a été calculé sur chaque fenêtre. Deux classes ont ensuite été construites l'une comme l'ensemble des fenêtres associées à un temps de réponse moyen bas, l'autre comme l'ensemble des fenêtres associées à un temps de réponse moyen élevé. L'EEG constitué de 60 électrodes a ensuite été analysé par classification sujet-par-sujet des signaux originaux, des sources corticales reconstruites au nombre de 60 et des sources corticales reconstruites au nombre de 500 environ. Amplitudes moyennes, cohérences et PLV ont été utilisées indépendamment pour quantifier les activités aux trois échelles spatiales décrites.

La seule différence significative entre le niveau du scalp et le niveau des 60 sources était une diminution des taux de bonnes classifications pour les amplitudes moyennes. Les cohérences offraient les meilleurs taux de bonnes classifications au niveau du scalp en comparaison avec les amplitudes moyennes et les PLV et décroissait également avec la reconstruction de sources. Seules les PLV permettaient une augmentation (non significative) au niveau des 60 sources par rapport au niveau du scalp, sans toutefois dépasser les taux de bonnes classifications obtenus avec les amplitudes moyennes et les cohérences au niveau du scalp. Enfin, les taux de bonnes classifications au niveau du scalp, quelle que fut la quantification utilisée, étaient déjà relativement bas (environ 70 %).

Les observations faites en simulations ne se retrouvent pas sur les vraies données, ce qui laisse penser à une défaillance de la technique utilisée pour la reconstruction de sources. Une double hypothèse émise porte sur l'activité électrique d'origine musculaire, vraisemblablement pertinente dans la discrimination de niveaux de performances mais non prise en compte dans les modèles directs actuels pour la reconstruction de sources. De plus, les PLV pourraient être moins sensibles aux activités d'origine musculaire et donc mieux adaptées à des situations de contamination de l'EEG par de telles

activités lorsque l'on souhaite ignorer cette contamination. La technique de l'inverse de norme minimale est également discutée.

5.7. Résumé : simulation de vol

a) Méthodes

Une même simulation de vol (durée environ égale à 2 h 45) a été réalisée par 13 élèves-pilotes placés aux commandes d'un simulateur d'Alphajet français. Le plan de vol faisait intervenir une alternance de segments de vol marqués ou non chacun par une série de manœuvres, lesquelles portaient sur le cap, la vitesse ou l'altitude. Les segments consistant en le maintien de ces trois paramètres de vol à des valeurs, appelés *Straight and Level flying Tasks* (SLT), ont été retenus pour l'analyse de deux mesures de la variabilité dans la tenue des paramètres de vol, puis pour la classification sujet-par-sujet de fenêtres de temps de 10 s pour la prédiction de deux niveaux de performance. Les fenêtres de temps ont été d'une part traduites par les amplitudes moyennes et les PLV pour les 32 électrodes d'EEG en trois bandes de fréquences (3–7 Hz, 7–13 Hz, 13–18 Hz), et ont d'autre part été réparties en deux classes construites comme les parties hautes et basses de la distribution d'une mesure de variabilité dans la tenue des paramètres de vol, mesure calculée sur la durée des fenêtres de temps (10 s). La classification a été évaluée par une procédure de validation qui construisait l'ensemble d'apprentissage comme l'ensemble des fenêtres de temps de deux à quatre SLT contigus et l'ensemble de test comme l'ensemble des fenêtres de temps d'un seul SLT, ultérieur aux SLT d'apprentissage. Outre la tâche « classique » de classification de fenêtres de temps du SLT de test, un critère de décision à l'échelle de la totalité des données de test a été défini : le taux de fenêtres prédites comme associées à des bonnes performances était rapporté aux taux réels correspondant, observés sur les SLT d'apprentissage, selon qu'il était supérieur, inférieur ou dans l'intervalle ; cette mesure à trois modalités était comparée à sa valeur attendue, fonction du taux de fenêtres réellement associées à de bonnes performances ; le critère était atteint lorsque les mesures « prédites » et réelles concordaient.

b) Résultats

L'analyse des mesures comportementales de la variabilité dans la tenue des paramètres de vol montrait une légère diminution du premier au deuxième SLT (débutant respectivement à 17 et 44 min après le décollage) puis une augmentation soutenue jusqu'au dernier SLT. Selon les configurations de validation, le taux de bonnes classifications moyens était compris entre 68,1 et 74,6 %, tandis que le nombre de sujets, pour lesquels le critère à l'échelle du SLT de test était atteint, valait 12 ou 13, pour 13 sujets, à l'exception d'une configuration pour laquelle ce nombre valait 11. Pour une majorité de sujets, la performance tendait à diminuer sur le SLT de test par rapport aux SLT d'apprentissage, mais

de 3 à 5 sujets se différenciaient de cette majorité sans que la non-diminution (augmentation ou « stabilisation ») de leur performance n'ait été mal prédite.

c) Conclusion

La capacité à prédire l'évolution générale de la performance en fin de simulation à partir de données issues d'une partie antérieure de la simulation, comme illustrée par le nombre de pilotes pour lesquels cette évolution était correctement prédite, permet de conclure favorablement quant à la faisabilité d'une telle approche. De plus, la mesure comportementale utilisée pour la construction des classes présente une évolution conforme à celle de mesures largement utilisées dans la littérature.

6 – Étude 1 : classification multi-modalités

6.1. L'expérience

Lors de la tâche, le sujet était installé dans une pièce, sur un fauteuil, face à un écran d'ordinateur. La tâche se déroulait comme une succession d'essais, chaque essai étant un stimulus visuel affiché à l'écran. Le stimulus se présentait sous la forme d'une paire chiffre-lettre (chiffre de 1 à 8 ; lettre parmi A, E, O, U, G, K, M ou R). Le sujet devait répondre alternativement par rapport au chiffre (tâche A), selon qu'il était pair ou impair, ou à la lettre (tâche B), selon qu'elle était une voyelle ou une consonne. À chaque essai, la paire chiffre-lettre était présentée, mais une seule tâche correspondait. Ainsi, les deux sous-tâches s'enchaînaient suivant un motif prédictible AABB (Fig. 11). Le sujet connaissait la tâche courante de par la prédictibilité de cette dernière et l'indice visuel donné par la zone dans laquelle apparaissait le stimulus. En effet, 4 zones carrés étaient visibles à l'écran, en permanence, et la paire chiffre-lettre, d'un essai au suivant, changeait de carré dans le sens horaire.

Le sujet disposait d'un boîtier à deux boutons pour répondre avec l'index et le majeur de la main droite.

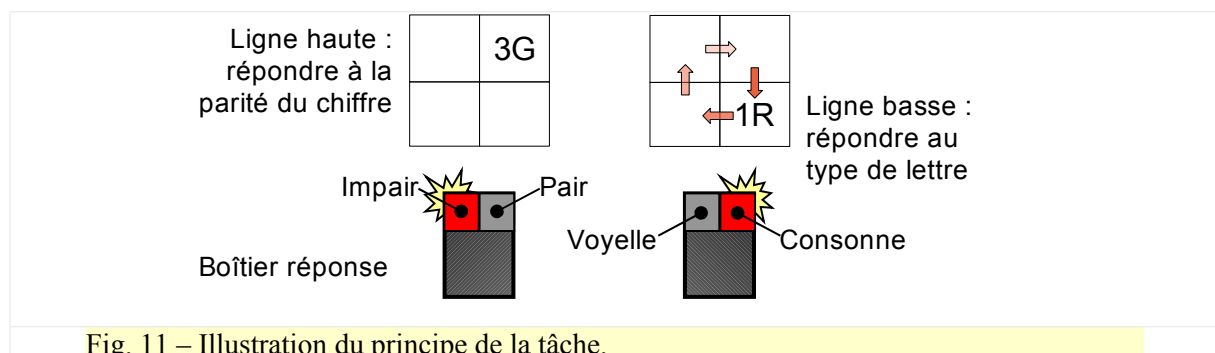


Fig. 11 – Illustration du principe de la tâche.

Les essais s'enchaînaient ainsi au nombre de 660 par bloc, avec un intervalle réponse-stimulus suivant (IRS) variable, aléatoirement choisi parmi les valeurs suivantes : 300, 600 ou 1500 ms. Le sujet devait répondre dans un intervalle de 2500 ms et à défaut le stimulus disparaissait et un nouvel essai débutait au terme d'un IRS.

Cette tâche est dite de *task switching* car formée de l'alternance de deux (sous-)tâches. Elle est proche de la tâche proposée par Rogers et Monsell (Rogers & Monsell, 1995). Bien que le sujet puisse connaître la tâche à réaliser sur la paire chiffre-lettre courante, l'alternance des tâches et la durée de la série des essais peuvent l'amener à commettre des erreurs.

Ainsi, le taux d'erreurs, mais aussi le temps de réponse et la variabilité du temps de réponse représentaient des variables comportementales intéressantes, car suspectées chacune de refléter une notion de performance.

6.2. Déroulement

Treize sujets masculins droitiers (à l'exception d'un ambidextre), âgés entre 23 et 31 ans, ont participé à l'expérience décrite, toujours pendant l'après-midi après le déjeuner (heure minimale de début : 14 h ; heure maximale de fin : 17 h 45). Une session était constituée d'un premier bloc d'entraînement avec retour sonore en cas d'erreur, puis d'un second demi-bloc d'entraînement sans retour sonore. L'installation du sujet avait ensuite lieu, avec la pose des électrodes. Un fois installé à une distance d'environ 90 cm de l'écran, dans une pièce dont l'éclairage était réglé à 4,5 lux, le sujet réalisait 6 blocs de 660 essais chacun, avec une courte pause à chaque transition entre blocs. Les 6 blocs totalisaient environ 2 h de tâche.

À chaque pause entre blocs, le sujet remplissait un questionnaire constitué d'une échelle en 7 niveaux de fatigue avec une description pour chaque niveau, adaptation française d'une échelle conçue par Pearson et Byars (1956).

6.3. Données comportementales et subjectives

a) Étude préliminaire : introduction

Le corps de la session d'enregistrement étant divisé en 6 blocs, il est naturel d'étudier en premier lieu les variations de certains paramètres d'un bloc à l'autre. La figure 12 présente les variations des principaux indicateurs étudiés. Outre l'indice de fatigue subjective qui dérive directement des valeurs mesurées à l'aide du questionnaire soumis après chaque bloc, il s'agit de variables comportementales : taux d'erreur, temps de réponse moyen (en comptabilisant une non-réponse comme le temps maximal autorisé au sujet pour répondre, à savoir 2500 ms), variabilité du temps de réponse (quantifiée par l'écart-type, là encore en incluant tous les essais) et coût du changement de tâche en termes de temps de réponse et de taux d'erreurs et non-réponses. Ce coût du changement de tâche, appelé *switch cost* en anglais, est la différence de performance entre les essais qui donnent lieu à un changement de tâche et ceux qui ne donnent pas lieu à un changement de tâche. On peut observer ce coût aussi bien sur le temps de réponse moyen que sur les taux d'erreurs et non-réponses. Il est possible d'estimer chacune de ces variables sur une sous-population d'essais comme celle partageant une même durée d'intervalle réponse-stimulus (IRS).

b) Étude préliminaire : résultats

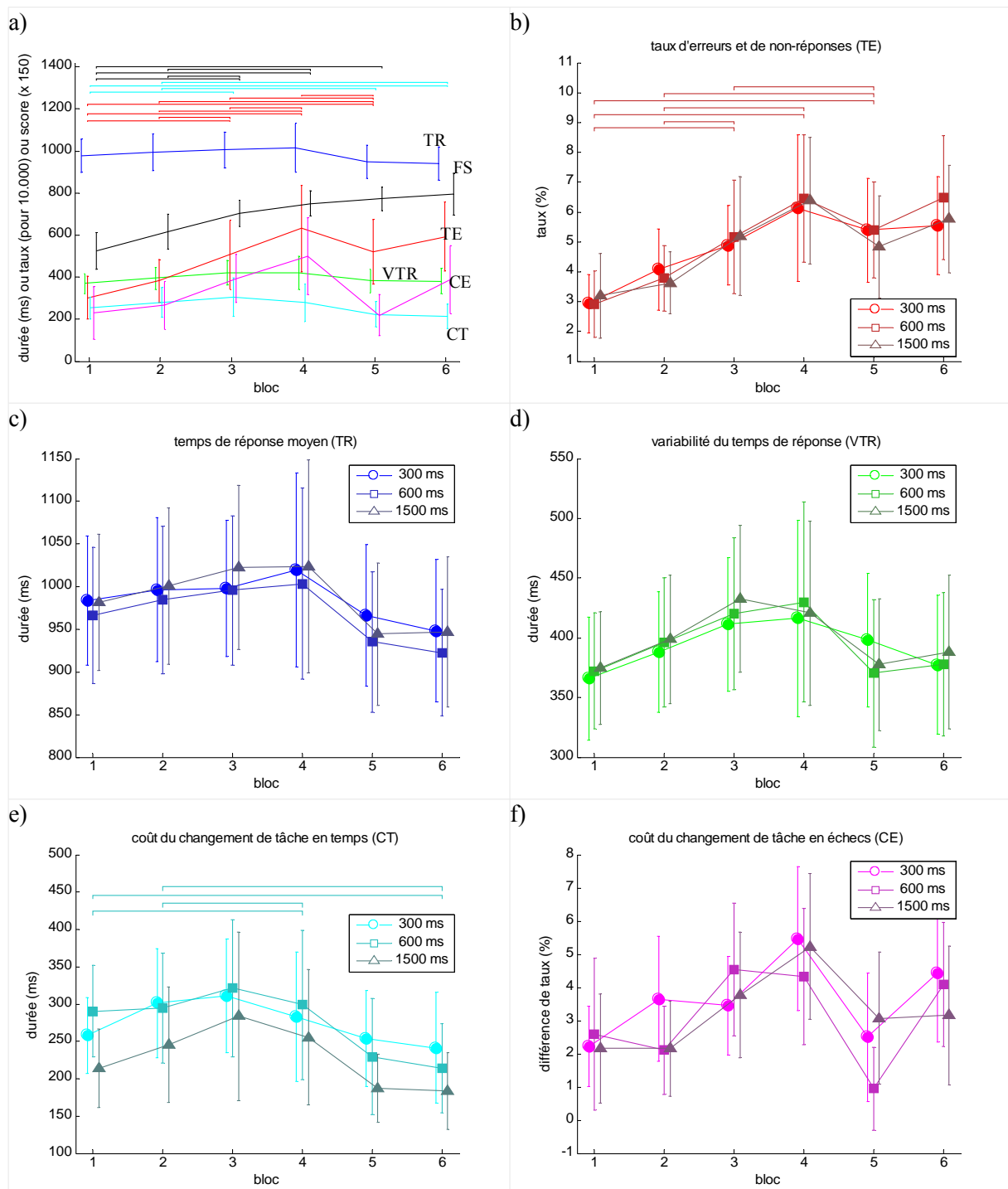


Fig. 12 – Évolution des variables comportementales et subjective sur les 6 blocs. a) Temps de réponse moyen (TR, bleu), variabilité du temps de réponse (VTR, vert), taux d'erreurs et de non-réponses (TE, rouge), coût du changement de tâche en temps de réponse (CT, cyan) et coût du changement de tâche en erreurs et non-réponses (CE, magenta) ; à ces variables comportementales est ajoutée la fatigue subjective (courbe noire). b,c,d,e,f) Chacune des 5 variables comportementales pour les 3 intervalles réponse stimulus (ISR ; de la plus claire à la plus sombre : 300, 600 et 1500 ms).

La figure 12, en particulier la partie a), présente le profil des variables comportementales au cours des 6 blocs qu'a duré l'expérience. Les valeurs moyennes pour les 13 sujets sont accompagnées des intervalles de confiance (à 95 %) sur ces moyennes. Ces intervalles de confiance reflètent pour beaucoup la variabilité inter-individuelle et les tests pour échantillon appariés (Wilcoxon avec contrôle du FDR), qui s'affranchissent de cette variabilité, permettent la mise au jour de différences significatives.

C'est le cas du taux d'erreurs et de non-réponses (TE) entre la plupart des couples possibles de blocs parmi les 5 premiers blocs, avec un accroissement maximal entre le premier et le quatrième blocs de plus de 3 points (un doublement du TE). De même, des différences apparaissent pour le coût de changement de tâche en termes de temps de réponse entre les premiers et derniers blocs, avec principalement une augmentation du bloc 1 au bloc 3 et globalement une diminution du bloc 1 au bloc 6 par exemple.

Outre ces critères comportementaux, la fatigue subjective pour chacun des deux premiers blocs est significativement inférieure à celle de chacun des trois blocs suivants (blocs 3 à 5).

Les autres parties (b,c,d,e,f) de la figure 12 portent sur chacun des critères comportementaux factorisés par la variable d'intervalle réponse-stimulus (IRS). Des différences significatives apparaissent pour les mêmes types de variables que précédemment.

Le comportement moyen des sujets présente une dynamique commune à la plupart des variables étudiées, avec une dégradation progressive au cours des trois voire quatre premiers blocs (soit de 1 h à 1 h 30 de tâche), puis une stagnation voire inversion de la tendance sur les deux blocs suivants. Les variables de temps de réponse ont diminué vers la fin de l'expérience, à la différence des variables d'erreur.

c) Discussion et construction des classes

Les résultats de la figure 12 ont été la base de réflexion pour la construction des états mentaux « fatigué » et « pas fatigué ». Nous voulions voir combinées les contributions de l'expression subjective des sujets, du temps passé à la tâche, et de la performance. Nous avons en réalité privilégié le temps passé à la tâche puisque cette stratégie a déjà été employée dans des études exploitant un protocole expérimental similaire (par exemple, Lorist *et al.*, 2000). Cela a eu pour conséquence que nous avons assimilé l'état « pas fatigué » à un bloc précoce parmi les 6, et l'état « fatigué » à un autre bloc plus tardif. En revanche, à la différence des études évoquées, nous ne nous sommes pas contentés de considérer le premier bloc et le dernier bloc de chaque sujet. Pour sélectionner précisément les deux blocs, nous avons impliqué la notion de performance, quantifiée par le taux d'erreurs : le bloc sur lequel le sujet réalisait son nombre minimum d'erreurs a été assimilé à son bloc « pas fatigué », tandis que celui sur lequel le sujet commettait le plus d'erreurs devint son bloc « fatigué ». Le taux d'erreurs a

été préféré au temps de réponse moyen à cause de l'absence de différence significative pour le second et de la décroissance sur les deux derniers blocs qui semble révéler un changement de stratégie plutôt qu'un effet direct de la fatigue.

d) Étude des blocs sélectionnés

Il se trouve que cette dernière procédure était compatible avec le critère de précédence du bloc « pas fatigué » sur le bloc « fatigué ». De plus, les scores de fatigue subjective augmentaient tous entre les blocs « pas fatigué » et les blocs « fatigué » respectifs, autrement dit l'effet de l'augmentation de la fatigue subjective était on ne peut plus significatif. Cela dit, les tâches de *task switching* sont connues pour favoriser la corrélation entre fatigue subjective, performance et temps passé à la tâche (Lorist *et al.*, 2000).

Outre ces variables, des effets significatifs ont été relevés entre blocs sélectionnés pour la variabilité du temps de réponse (tous essais : $+69,7 \pm 47,2$ ms, et essais d'IRS égaux à 300 ms : $+77,6 \pm 39,5$ ms) et pour le taux d'erreurs et de non-réponses (tous essais : $+4,7 \pm 1,6$ %, et tous IRS : respectivement $+4,6 \pm 1,5$ %, $+4,9 \pm 2,2$ % et $+4,7 \pm 1,8$ %).

Les tests de Wilcoxon pour échantillons appariés ont été corrigés pour comparaisons multiples (ici par la méthode de Bonferroni) à travers les variables sans toutefois inclure les différents IRS ensemble, car la corrélation d'une même variable sur différents IRS est très élevée, comme l'illustre la figure 12.

6.4. Données physiologiques – Classification

a) Introduction

Les deux classes étant définies, nous pouvions chercher les différences dans les signaux physiologiques entre les deux situations que représentent les classes. Nous avons pour cela appliqué la chaîne de traitements détaillées chapitres 5.3. sans faire de reconstruction de sources (ce n'est pas l'objet de cette étude) et 5.4.. Il a fallu en outre découper chaque bloc en fenêtres de temps dont nous avons fait varier la durée (de 4 à 30 s), mais toujours avec un recouvrement de 50 % entre fenêtres successives.

L'ECG à 4 et 6 s a été quantifié seulement par la fréquence cardiaque (HR) et l'écart-type des intervalles R-R (RR(sd)) car la procédure de filtrage pour les autres quantifications nécessitait plus d'échantillons de signal que disponibles.

b) Résultats

La figure 13 présente les taux de bonnes classifications moyens à travers les sujets pour les modalités prises seules (en séparant pour l'EEG amplitudes moyennes et PLV) et toutes ensemble.

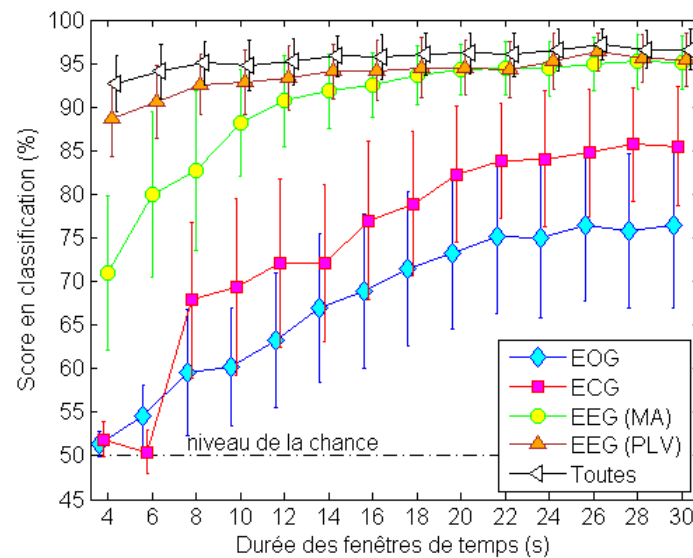


Fig. 13 – Taux de bonnes classifications pour les différentes modalités prises seules ou combinées par concaténation de leurs variables associées (courbe « Toutes »). Les abscisses représentent la taille des fenêtres de temps (en s), les ordonnées présentent les taux de bonnes classifications et les différentes courbes correspondent aux modalités.

Comme représenté par les intervalles de confiance, seuls les taux de bonnes classifications pour l'EOG et l'ECG à 4 et 6 s ne se démarquent pas significativement du niveau de la chance.

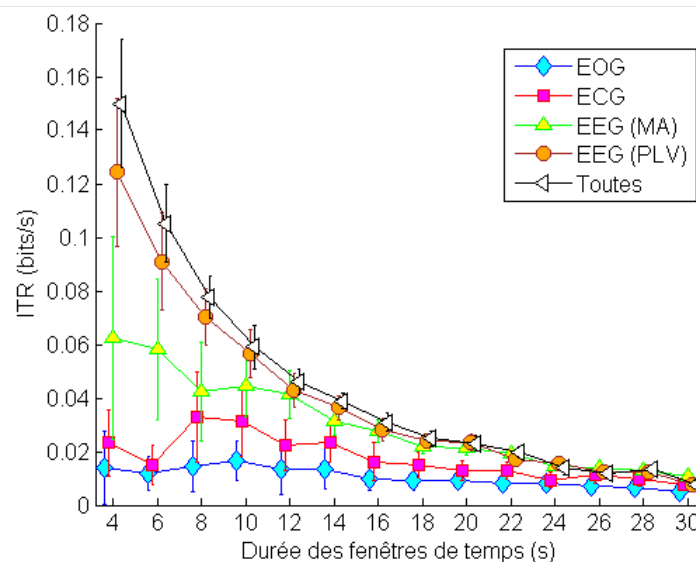


Fig. 14 – Taux de transfert d'information correspondant aux taux de bonnes classifications présentés précédemment.

Si l'on considère les modalités deux à deux, aux différentes durées de fenêtre, on trouve des effets significatifs (Wilcoxon pour échantillons appariés, sans correction) partout sauf entre l'EOG et l'ECG entre 4 et 18 s et à 26 s, entre les amplitudes moyennes et les PLV à 14 s, entre 18 et 24 s et à 28 et 30 s, et enfin entre les amplitudes moyennes (ou de même, les PLV) et la combinaison des modalités à 28 s.

Les ITR correspondant aux taux de bonnes classifications figure 13, comme décrits section 4.3.a) , sont présentés figure 14. Le maximum est obtenu à 4 s.

c) Discussion

La supériorité de l'EEG par les seules amplitudes moyennes est illustrée dans les figures précédentes. On remarque aussi l'apport faible mais significatif des quantifications de l'EOG et de l'ECG. En fait, cet apport était plus important lorsque le taux de bonnes classifications n'était pas déjà très élevé, c'est ce que nous avons observé en ajoutant les quantifications de l'EOG et de l'ECG aux seules amplitudes moyennes de l'EEG, à de courtes durées de fenêtres de temps.

C'est un point non négligeable, car dans un contexte de détection, une période d'intégration plus courte permettra une détection plus précoce. Et dans le cas où l'exactitude de la détection est prioritaire sur sa latence, il est possible, plutôt que de prendre une décision au terme d'une durée longue, de découper cette longue période d'intégration en fenêtres de temps plus courtes et de cumuler les réponses du détecteur pour ces fenêtres. Le taux de bonnes classifications n'est pas nécessairement le même. Cet aspect est modélisé par la notion de taux de transfert d'information. Les résultats obtenus plaident en faveur de fenêtres de temps les plus courtes possibles.

À ce sujet, les PLV montrent une nette supériorité sur les amplitudes moyennes, en particulier à de courtes tailles de fenêtres. Peut-être l'emploi d'une fenêtre de Hanning pénalise-t'il l'estimation des amplitudes plus que celles des phases.

6.5. Contribution des variables à la classification

a) Introduction

Pour dépendre partiellement ce qui est commun aux 13 sujets, l'examen des poids SVM est plus approprié. Une durée de fenêtres de temps de 30 s a été retenue pour obtenir des quantifications ECG aussi proprement estimées que possible (Salahuddin *et al.*, 2007). La procédure est décrite dans la section 5.4.c) . À la différence de la procédure de classification en elle-même appliquée sujet-par-sujet, on peut se pencher ici sur la problématique de la variabilité inter-individuelle. Pour cela, nous présentons à la fois le poids moyen de chaque variable après réduction, c'est-à-dire en divisant sa moyenne par son écart-type à travers les sujets, et les mêmes poids sans réduction (Fig. 15).

b) Résultats

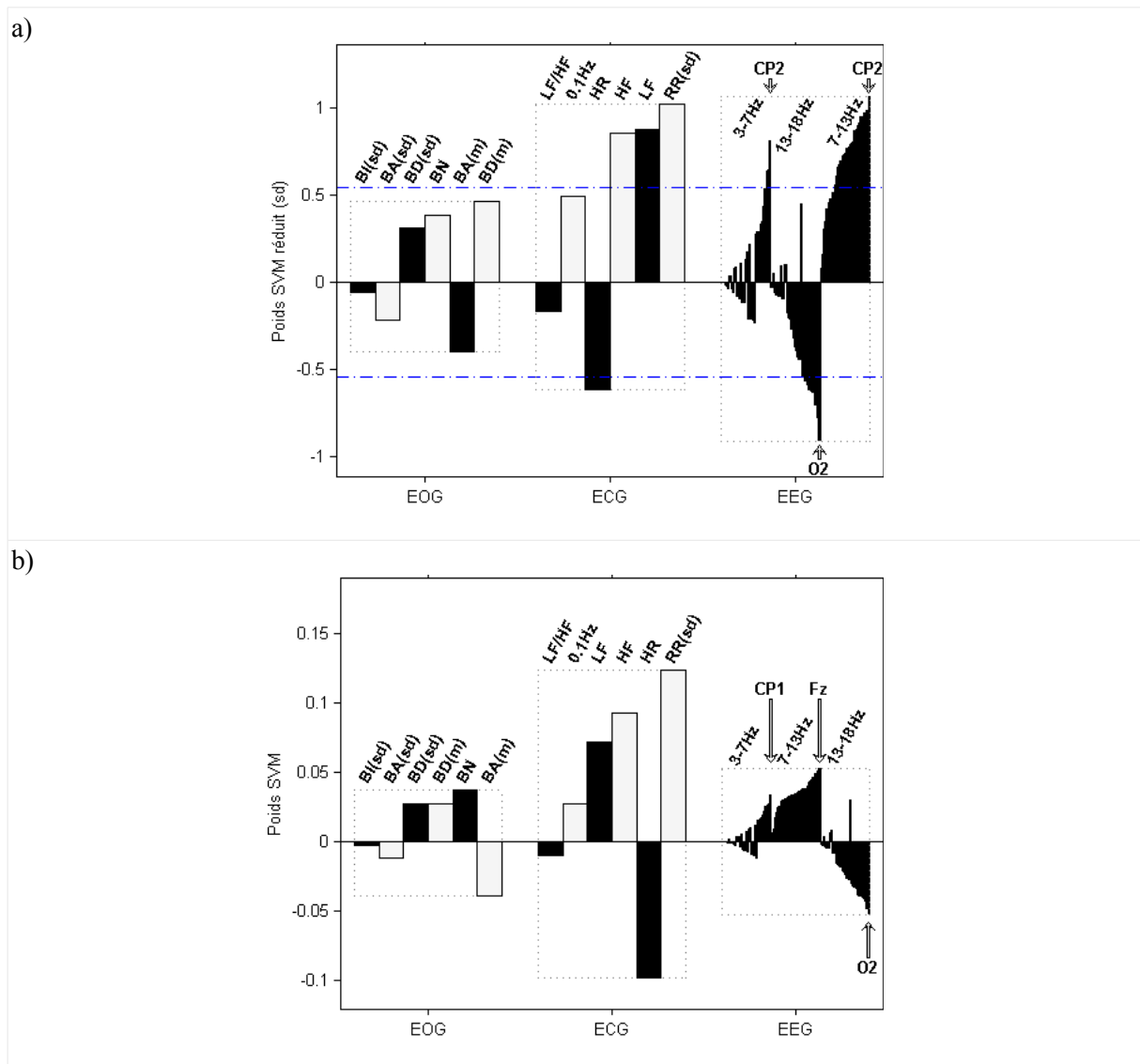


Fig. 15 – Poids SVM moyennés avec (a) et sans (b) réduction par l'écart-type inter-sujet, pour des fenêtres de temps de 30 s et l'ensemble des quantifications EOG, ECG et EEG (PLV comprises, bien qu'absentes ici). Les quantifications sont regroupées par modalité, et ordonnées à l'intérieur de chaque modalité (et bande de fréquences pour les amplitudes moyennes de l'EEG) par poids absolu croissant. Le signe des poids est représentatif de leur sens de variation d'une classe à l'autre (positif = croissant avec la fatigue). a) Les lignes horizontales bleues marquent l'intervalle de confiance à 95 % autour de 0.

Les poids des différentes modalités présentent des variabilités inter-individuelles différentes (moyenne des écart-types des poids \pm intervalle de confiance) : $0,075 \pm 0,017$ pour l'EOG, $0,098 \pm 0,031$ pour l'ECG, $0,057 \pm 0,003$ pour les amplitudes moyennes de l'EEG et $0,023 \pm 0,000$ pour les PLV de l'EEG.

Tous les poids des amplitudes moyennes en 7-13 Hz sont positifs, tandis que les poids des amplitudes moyennes pour la bande 13-18 Hz sont significativement négatifs (test de Wilcoxon contre zéro des poids moyens correspondant) tant avec que sans réduction.

La partie a) de la figure 15 présente l'intervalle de confiance commun à toutes les variables, placé autour de zéro. Ainsi, les variables qui sortent de l'intervalle atteignent un certain niveau de significativité, dans la limite de l'absence de correction qui serait ici nécessaire.

En particulier, aucun poids des variables d'EOG n'exhibe une tendance significativement commune aux sujets.

La figure 16 donne une représentation lissée des poids SVM moyens réduits relatifs à l'EEG sur le scalp pour les différentes bandes de fréquences, toujours dans le cas de fenêtres de temps de 30 s.

Les cartes moyennes de PLV ont été obtenues ici par la procédure suivante : les poids SVM des PLV ont tout d'abord été ordonnées en valeur absolue croissante, toutes fréquences confondues ; pour chaque couple électrode-fréquence, les rangs de tous les poids relatifs aux PLV impliquant l'électrode à la fréquence donnée ont été sommés ; cette somme a été ensuite multipliée par le signe de la somme des mêmes rangs mais eux-mêmes chacun multiplié par le signe du poids correspondant ; enfin, la même procédure de moyennage et réduction (à travers les sujets) appliquée aux amplitudes moyennes a été appliquée aux cartes de sommes signées de rangs.

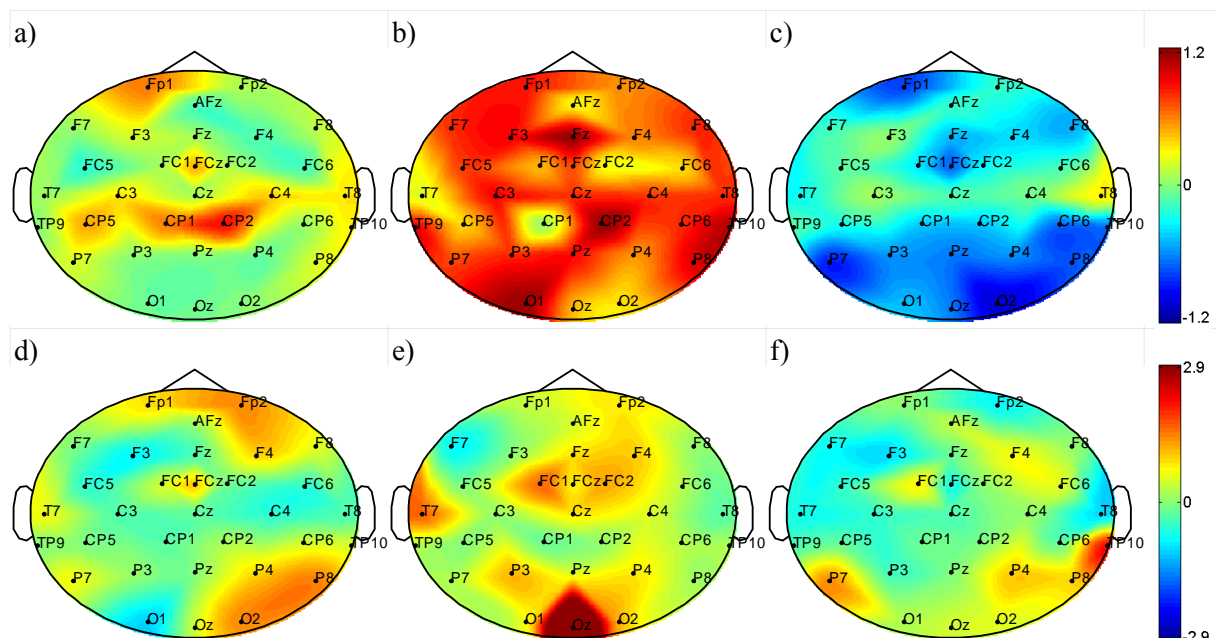


Fig. 16 – Topographies de scalp à partir des poids SVM réduits appris en combinaison de modalités sur des fenêtres de 30 s. a,b,c) Amplitudes moyennes respectivement en 3-7 Hz, 7-13 Hz et 13-18 Hz ; le rouge correspond à une augmentation avec la fatigue, et le bleu à une diminution. d,e,f) PLV pour les mêmes bandes de fréquences.

c) Discussion

Les amplitudes moyennes de l'EEG tendent à augmenter en 3-7 Hz, principalement en CP1-CP2, augmentent de manière généralisée en 7-13 Hz, et diminuent sur la quasi-totalité du scalp pour la bande 13-18 Hz.

Les cartes d'amplitudes moyennes et de PLV sont très différentes. L'activité en FCz présente un niveau de discrimination qui démarque cette électrode de ses voisines, tant en amplitudes moyennes qu'en PLV. En 3-7 Hz, on observe une augmentation de synchronie locale, mais aussi une plus forte implication de l'électrode dans des synchronies à distance (pour un pouvoir discriminant relativement élevé), tandis que l'on observe la situation opposée pour la bande 13-18 Hz, à savoir une diminution de la synchronie locale pour un pouvoir discriminant comparable et un faible pouvoir discriminant pour la PLV avec une tendance baissière des synchronies distantes.

Ensuite, l'ensemble des PLV impliquant l'électrode Oz en 7-13 Hz présente le plus fort pouvoir discriminant toutes variables prises en compte. L'augmentation des rythmes alpha semble donc pilotée par cette électrode alors même que la synchronie locale correspondante ne présente pas le plus fort pouvoir discriminant parmi ceux des amplitudes moyennes de la même bande de fréquences.

Pour en revenir aux quantifications des modalités extracérébrales, on remarque que toutes les variables d'ECG sont pénalisées par une variabilité inter-sujet élevée. Cela dit, cette pénalisation peut être due à des variations importantes chez peu de sujets, en sens opposé à la majorité. C'est le cas pour la fréquence cardiaque pour laquelle on trouve en réalité seulement 1 sujet présentant une (forte) augmentation avec la fatigue (poids supérieur à 0,3) et 8 sujets avec une diminution modérée à forte (poids entre $-0,3$ et $-0,1$). Les variables d'ECG n'en restent pas moins utiles à la classification.

Ensuite, le nombre de clignements et leur durée augmentent avec la fatigue alors que leur amplitude diminue, quoique non significativement. La variabilité du rythme cardiaque augmente quelque soit la quantification choisie.

6.6. Discussion

a) Différences entre états « pas fatigué » et « fatigué »

Les variables de temps de réponse ont diminué vers la fin de l'expérience, à la différence des variables d'erreur. Les sujets auraient donc pour une part accéléré la cadence pour finir l'expérience plus vite, quitte à la « saboter » en faisant plus d'erreurs.

L'interprétation de la contribution de chaque variable physiologique amène à des parallèles avec de nombreuses autres études. Pour commencer, les changements dans l'activité oculopalpébrale correspondent à ceux rapportés dans des situations de fatigue mentale, avec ou sans privation de sommeil : les clignements se font plus nombreux, mais moins amples et plus longs (Morris & Miller,

1996 ; Stern, 1994). Il semblerait qu'à charge de travail suffisamment élevée, le réflexe de clignement soit normalement inhibé puis que, avec l'installation de la fatigue mentale, cette inhibition s'amenuise. Il est reconnu cependant que ces résultats attendus sont très dépendants de la tâche visuelle. De plus ici, les poids rapportés sont faibles, et leurs écarts à 0 ne sont pas significatifs, ce qui implique une variabilité inter-individuelle élevée comme rapporté par d'autres auteurs (Ingre *et al.*, 2006).

La diminution de la fréquence cardiaque est également en accord avec une étude qui avait montré un effet modérateur du temps passé à la tâche (Floru *et al.*, 1985). En revanche, on ne retrouve pas la marque d'un accroissement de la charge de travail subjective (Roscoe, 1992), manifestée par une augmentation de la fréquence cardiaque et une diminution de la variabilité du rythme cardiaque. Ici, les variables d'ECG autres que la fréquence cardiaque et le rapport LF/HF quantifie divers aspects de cette variabilité cardiaque. L'augmentation commune à ces quantifications pourrait être redondante, malgré le fait qu'elles reflètent normalement des composantes différentes de modulation de l'activité cardiaque. En effet, les quantifications spectrales de l'ECG sont susceptibles d'être improprement estimées, car les durées des fenêtres de temps sont pour la plupart en théorie trop courtes par rapport aux fréquences considérées (Salahuddin *et al.*, 2007). Il se pourrait donc que la forte augmentation de la variabilité cardiaque telle que reflétée par l'écart-type des intervalles R-R se manifeste par une augmentation en puissance à toutes les fréquences, ce que l'on observerait là plutôt que des manifestations des composantes sympathique et parasympathique du système nerveux autonome. De plus, les poids moyens relativement élevés des quantifications d'ECG ne sont pas soutenus par les taux de bonnes classifications. Lorsque l'on considère les taux de bonnes classifications obtenus à l'aide de ces seules quantifications, il semblerait que l'ECG soit fort utile à la discrimination chez une moitié des sujets, et très peu chez les autres. Cette variabilité est reflétée par les larges intervalles de confiance des taux de bonnes classifications à toutes les tailles de fenêtres et des poids SVM moyens.

En ce qui concerne les amplitudes moyennes de l'EEG, la diminution en 13-18 Hz est en contradiction avec l'observation, faite par d'autres auteurs pour d'autres tâches, d'une activité croissante de certains muscles péricrâniens. En effet, il a été montré que sur une période de 20 min, les sujets, pour maintenir leur niveau de performance, présentaient une augmentation de l'activité de muscles tels que le frontal, les corrugateurs et l'orbiculaire de la lèvre inférieure, augmentation absente chez les sujets qui laissaient librement leur performance décroître (van Boxtel & Jessurun, 1993 ; Waterink & van Boxtel, 1994). Cette activité musculaire a été attribuée à l'effort mental déployé par les sujets qui souhaitaient compenser la baisse de performance de laquelle la fatigue mentale pourrait bien avoir été à l'origine. De manière générale, l'activité motrice augmente avec la fatigue comme un mécanisme d'auto-stimulation (Bonnefond *et al.*, 2006). Dans notre étude, la fatigue mentale a amené un effet contraire, probablement en raison d'un temps passé à la tâche plus long, ce qui suggérerait que la dynamique de la fatigue est relativement complexe, au fur et à mesure qu'elle s'installe chez l'individu.

En outre, la bande 13-18 Hz reflète l'activité liée à divers processus cognitifs et la moindre réduction en amplitudes moyennes en C3 et C4 pourrait bien correspondre à une augmentation d'activité liée aux erreurs comme décrit dans une étude récente (Koelewijn *et al.*, 2008). On peut ajouter que le passage en référence moyenne est susceptible d'avoir étalé la diminution dominante d'activité dans cette bande de fréquence et donc d'avoir gommé partiellement les diverses augmentations comme celle que l'on imagine ici en C3-C4.

Ensuite, l'activité en 7-13 Hz augmente de manière généralisée à l'ensemble du scalp, avec l'électrode Oz en « chef d'orchestre ». Cette observation laisse penser aux résultats obtenus par les études sur l'hypovigilance. Le phénomène pourrait être d'autant plus fort qu'une diminution des rythmes de cette bande de fréquences a été observée entre 12 et 24 h, en situation d'activité modérée (Aeschbach *et al.*, 1997 ; Aeschbach *et al.*, 1999). Les sessions ayant toutes eu lieu l'après-midi, les rythmes circadiens ont dû s'opposer à cette augmentation.

Enfin, en ce qui concerne la bande 3-7 Hz, un motif attire l'attention ; il s'agit d'une augmentation d'activité en FCz, visible tant dans les amplitudes moyennes que dans les PLV. Cette augmentation laisse penser à la négativité d'erreur (*Error-Related Negativity*, ERN) dont on sait qu'elle apparaît tout particulièrement dans cette bande de fréquences (Luu *et al.*, 2004).

b) Terrain d'amélioration de la méthode

Les résultats en classification appellent une question concernant une corrélation troublante : qu'elle est l'influence du nombre de variables ? Plus celui-ci est élevé, plus le taux de bonnes classifications correspondant est également élevé.

Oublions dans un premier temps l'existence de la validation et réfléchissons à ce qui se passe lors de l'apprentissage. Un nombre de variables plus grand permet à l'algorithme d'apprentissage de compenser la limitation que pose le choix d'une séparatrice linéaire. En effet, une séparatrice linéaire n'offre qu'un seul axe possible de discrimination. Pour qu'une telle séparatrice permette bien de discriminer, il faut que les distributions des points de chacune des classes soient simples, qu'elles ressemblent d'autant que possible à la représentation qu'on se fait de nuages homogènes de points. Or, chaque variable ajoutée offre un degré supplémentaire de liberté dans l'ajustement de la séparatrice, son propre pouvoir discriminant aussi petit soit-il, si bien qu'en général, une augmentation du nombre de descripteurs creuse l'espace entre les deux classes et amène les groupes de points à occuper des volumes plus facilement identifiables.

Un raisonnement plus rigoureux au sujet de ce phénomène procède en considérant en premier lieu l'influence du nombre de dimensions (de variables) sur le volume d'une boule. Comme illustré figure 17, le volume de la boule inscrite dans l'hypercube unité décroît avec le nombre de dimensions.

En effet, si l'on compare ce volume, dans l'ordre croissant du nombre de dimensions (1, 2 et 3), on a : $1 > \pi/4 > \pi/6$.

On peut ensuite traduire cette situation de la boule à travers des espaces de dimensionnalité croissante comme celle de boules de volumes décroissants dans un même espace. Ainsi, si l'on considère une technique de classification qui exploite des éléments de nature sphérique pour définir les régions représentatives des classes, éléments que l'on appelle fonctions à base radiale, comme par exemple des gaussiennes dont les isocourbes sont des hypersphères, on se place directement dans un cas d'application de ce phénomène de perte de volume avec le nombre de dimensions.

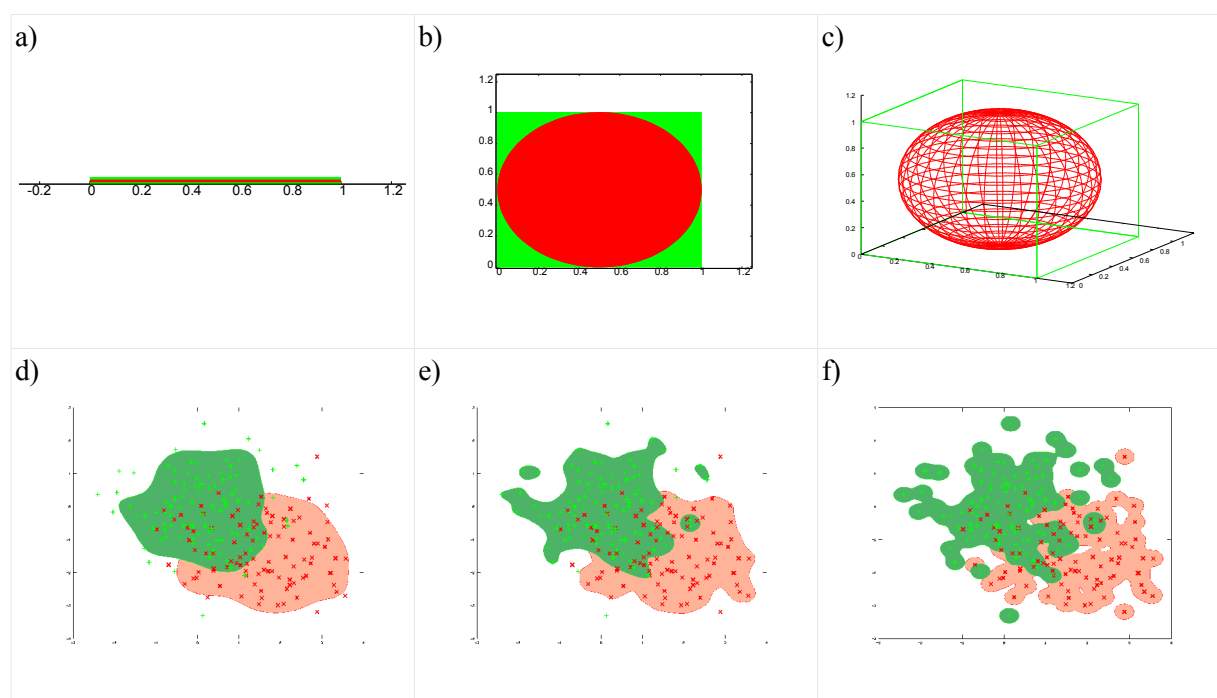


Fig. 17 – Effet du nombre de dimensions. a,b,c) Comparaison du « cube » de côté 1 et de la « boule » de diamètre 1 respectivement à 1, 2 et 3 dimensions ; si le volume du cube est pris comme unité de volume, le volume de la boule est respectivement 1, $\pi/4$ et $\pi/6$. d,e,f) Comparaison, à même nombre de dimensions, de la section intérieure d'une surface construite comme une somme de gaussiennes, à raison d'une gaussienne centrée en chaque point ; la dispersion des gaussiennes est constante pour chaque figure et décroît d'une figure à l'autre, de la gauche à la droite ; les sections rouges, en partie masquée par les sections vertes correspondantes, ont la même surface d'une figure à l'autre, tandis que chaque section verte est construite à l'aide du même seuil que celui de la section rouge correspondante.

Toujours en lien avec la figure 17, un mélange de fonctions radiales, c'est-à-dire une somme de fonctions identiques placées une en chaque point d'apprentissage, est un exemple de classifieur qui

opère ici en attribuant un point quelconque de l'espace à la classe dont la valeur du mélange correspondant serait la plus élevée. Si, en restant dans le même espace, on diminuait le rayon des fonctions à base radiale, on constaterait que le volume occupé par chaque classe pour une valeur donnée du mélange (le volume délimité par une isocourbe donnée) diminuerait. De plus, comme illustré figure 17, même en choisissant une isocourbe adaptée de manière à obtenir un volume intérieur de taille constante quel que soit la grandeur des rayons, la forme de ce volume deviendrait plus irrégulière.

Le SVM linéaire n'opère pas exactement de la même décrite, mais il n'échappe pas pour autant à ce phénomène. La séparabilité des données d'apprentissage est donc meilleure à nombre de variables plus grand et une conséquence directe est un risque accru de sur-apprentissage, s'il est possible de sur-apprendre. Ainsi, les classifications à l'aide des amplitudes moyennes de l'EEG, celles utilisant les PLV encore plus, pourraient avoir bénéficié d'un avantage supplémentaire « injuste » sur celles exploitant seuls l'EOG et l'ECG.

Ce problème de sur-apprentissage éventuel concerne avant tout la validation, même si on tente de le traiter dès l'apprentissage du classifieur par l'emploi d'heuristiques. Une procédure de validation est pourtant employée ici, mais est-elle suffisante ? Autrement dit, peut-on sur-apprendre sans que cela entraîne une diminution du taux de bonnes classifications sur les données de test ? La réponse (à la seconde question) pourrait bien être oui, car pour supprimer le sur-apprentissage, il faut que les données de test et d'apprentissage aient été acquises dans des conditions qui, par leurs différences, garantissent la généralisabilité des résultats obtenus aux situations que l'on souhaite couvrir. Ayant appliqué la procédure complète sujet-par-sujet, il faudrait par exemple vérifier que les résultats obtenus soient stables d'un jour à l'autre pour un même sujet, donc tester le classifieur sur les données d'une autre session. Ne pas avoir procédé ainsi rend probable la contribution de fluctuations sans rapport avec la fatigue, donc plus ou moins dues au hasard, mais de dynamique semblable à celle de la fatigue sur la durée de l'enregistrement étudié.

Les effets qui se superposent à la fatigue au cours d'une session seraient probablement désynchronisés de la fatigue sur une autre session, si par exemple ces effets sont périodiques. Malheureusement, il est assez difficile de convaincre les sujets de participer plusieurs fois à une expérience consistant en une tâche longue et pénible. En bref, il y a un possible effet de chance qui n'est pas du tout contrôlé ni quantifié dans les résultats en classification. Ces derniers sont donc contestables et à considérer comme préliminaires.

La solution employée par la majeure partie des études scientifiques consiste à étudier des motifs communs à tous les sujets, en faisant ainsi d'une pierre deux coups. En effet, un effet stable à travers les sujets est très probablement stable pour un même sujet à travers le temps, dans la limite de ce qui a été fixé dans l'expérience comme commun à tous les sujets (la tranche d'âge des sujets, ou le moment

de la journée pendant lequel se sont déroulées les sessions, par exemple). C'est par ce moyen que l'analyse précédente de la contribution de chaque quantification à la classification conserve son intérêt, mais au risque de me répéter, dans une approche de détection, on devrait idéalement ne pas avoir à résoudre la variabilité inter-individuelle, du fait que l'on peut adapter le détecteur à l'individu.

Cependant, en l'absence d'un grand nombre d'enregistrements par sujet, soit il faut se résoudre à développer des quantifications qui prennent en compte cette variabilité, puis à appliquer la procédure de classification-validation en classant les fenêtres de tous les sujets à la fois, soit on peut tenter de développer des techniques de validation, ou construire des classes telles qu'il y ait moins d'effets superposés à craindre. La première solution a l'inconvénient soit d'être hasardeuse, en particulier si l'on n'a aucun *a priori* et que l'on doit opérer par essai-erreur, soit de nécessiter une connaissance fine des différences entre individus. La deuxième solution ne résout pas directement le problème, mais permet une correction au moins partielle. Il faut par contre avoir des hypothèses sur la dynamique des effets susceptibles de se superposer à ceux de la fatigue mentale pour pouvoir réduire leur contribution.

Ces phénomènes, pour ceux qui seraient périodiques, seraient principalement des rythmes lents, tels que la période serait supérieure à la durée d'un bloc (environ 20 min). Ces rythmes pourraient donc être des rythmes ultradiens ou circadiens. Outre des rythmes, on peut envisager que l'apprentissage de la tâche par le sujet ait pu avoir un effet, la dynamique de cet effet étant une variation monotone (un rythme de période infinie). Le concept « un bloc = une classe » est susceptible de souffrir de telles superpositions d'effets. Une construction différente des classes est donc à envisager.

7 – Étude 2 : reconstruction de sources

7.1. Introduction

L'EEG ayant paru pertinente pour la détection de la fatigue mentale dans l'étude précédente, nous nous sommes concentrés à partir de ce point sur cette seule modalité, avec les mesures de synchronie à distance en principales candidates.

L'intérêt levé pour ces quantifications de lien dynamique entre des sites distants remonte aux observations et hypothèses liées à certains phénomènes cognitifs tels que la mémoire, la perception, l'attention et la conscience (Rodriguez *et al.*, 1999 ; Varela *et al.*, 2001). Le concept d'« assemblée résonante de neurones » propose que les processus mentaux qui nécessitent une mise en commun simultanée de nombreuses informations de natures différentes, et donc encodées en des sites cérébraux éventuellement éloignés les uns des autres, soient permis par la synchronisation des populations de neurones responsables chacune de la représentation d'une de ces informations. Ces populations déchargeraient de manière synchronisée pour que les représentations nécessaires à l'émergence de la fonction cognitive soient en quelque sorte émises toutes en même temps. Pour cela, un mécanisme d'entraînement mutuel des activités pourrait agir, l'une ajustant ses rythmes à ceux des autres. Un réseau fonctionnel se mettrait ainsi en place, de manière vraisemblablement éphémère, et sur une étendue spatiale qui pourrait rendre possible l'observation de ce réseau à l'aide de l'EEG.

Notre objectif dans cette étude est d'estimer l'intérêt de la reconstruction de sources, avec l'espoir d'obtenir des marqueurs EEG avec une plus fine résolution spatiale et une meilleure précision en localisation. Cependant, les techniques de reconstruction de sources n'ont eu pour l'instant que quelques usages relativement limités. On ne trouve en particulier aucune étude traitant à la fois de reconstruction de sources, de mesures de couplage et de classification. Par exemple, certaines ICM exploitent la forte régionalisation corticale d'une fonction, contexte dans lequel la reconstruction de sources permet d'améliorer la détection sans l'usage de techniques de classification (Qin *et al.*, 2004 ; Congedo *et al.*, 2006 ; Kamousi *et al.*, 2007). D'autres en revanche emploient un algorithme de classification en quantifiant des activités reconstruites, mais via des mesures de synchronie locale telles que la puissance (Lotte *et al.*, 2007 ; Noirhomme *et al.*, 2008). On recense également des études qui décrivent des synchronies à distance entre sources reconstruites, mais en dehors du contexte de détection ou d'ICM (Cosmelli *et al.*, 2004 ; David *et al.*, 2006 ; Astolfi *et al.*, 2007 ; ten Caat *et al.*, 2008). D'autres abordent même la recherche d'une information de structure dans les réseaux fonctionnels ainsi estimés, mais toujours sans faire de classification (De Vico Fallani *et al.*, 2007).

Nous avons donc d'abord réalisé une simulation pour vérifier les propriétés attendues de la reconstruction de sources dans un contexte de classification, en utilisant et comparant plusieurs mesures de synchronies, aussi bien locales que distantes.

7.2. La simulation

L'IRM T1 d'un des sujets de l'expérience « réelle » présentée plus loin a été segmentée pour en extraire l'interface entre la matière blanche (sous corticale) et la matière grise du cortex cérébral, sous la forme d'un maillage à 10014 sommets représenté figure 18. Un dipôle de courant a été placé en chaque sommet, sa direction étant défini comme la moyenne des normales des faces adjacentes du maillage. Deux dipôles ont été sélectionnés de manière plus ou moins arbitraire, en faisant en sorte qu'ils soient éloignés l'un de l'autre. Une région autour de chacun de ces dipôles a été définie en retenant les sommets du maillage situés à une distance « géodésique » inférieure ou égale à 7 arêtes (ou 6 sommets intermédiaires).

Deux signaux ont été générés, chacun par filtrage d'un bruit gaussien par le filtre de fonction de transfert suivante :

$$H(z) = \frac{1 - z^{-1}}{1 - 2r \cos(2\pi f / F_s) z^{-1} + r^2 z^{-2}}$$

avec $f=10$ Hz et F_s la fréquence d'échantillonnage, soit, en notant x le bruit et y le signal obtenu, $y(t) = x(t) - x(t-1) - 2r \cos(2\pi f / F_s) y(t-1) + r^2 y(t-2)$. Ces deux signaux étaient ensuite modifiés de manière à ce qu'ils partagent une propriété particulière que l'on doit pouvoir détecter avec une mesure de synchronie soit locale soit distante. Cette propriété variait quantitativement de manière à obtenir deux modes d'activité chacun représentatif d'un état mental virtuel (Fig. 19). Ces signaux ont été diffusés sur les régions définies avec un maximum d'amplitude au centre, puis un bruit blanc a été ajouté en chaque source (un bruit différent par source), ceci afin de simuler la part d'activité cérébrale non corrélée aux états mentaux d'intérêt. Enfin, cette activité corticale a été projetée sur le scalp (60 électrodes) de manière à obtenir les signaux EEG simulés utilisés ensuite pour les analyses (Fig. 18).

Les mesures de synchronie ont été évaluées en jouant sur la nature des deux signaux utiles générés au niveau cortical. Pour tester les mesures de synchronies à distance, les deux signaux ont été construits de manière à entrer en synchronie totale pour une durée de 100 ms dans l'état 1 (les fenêtres de signal ainsi générées forment la classe 1) et de 500 ms dans l'état 2 (classe 2). Ce motif était répété avec une période de 1 s. Les données produites dans cette configuration constituaient la base de la « simulation 2 » (Fig. 19).

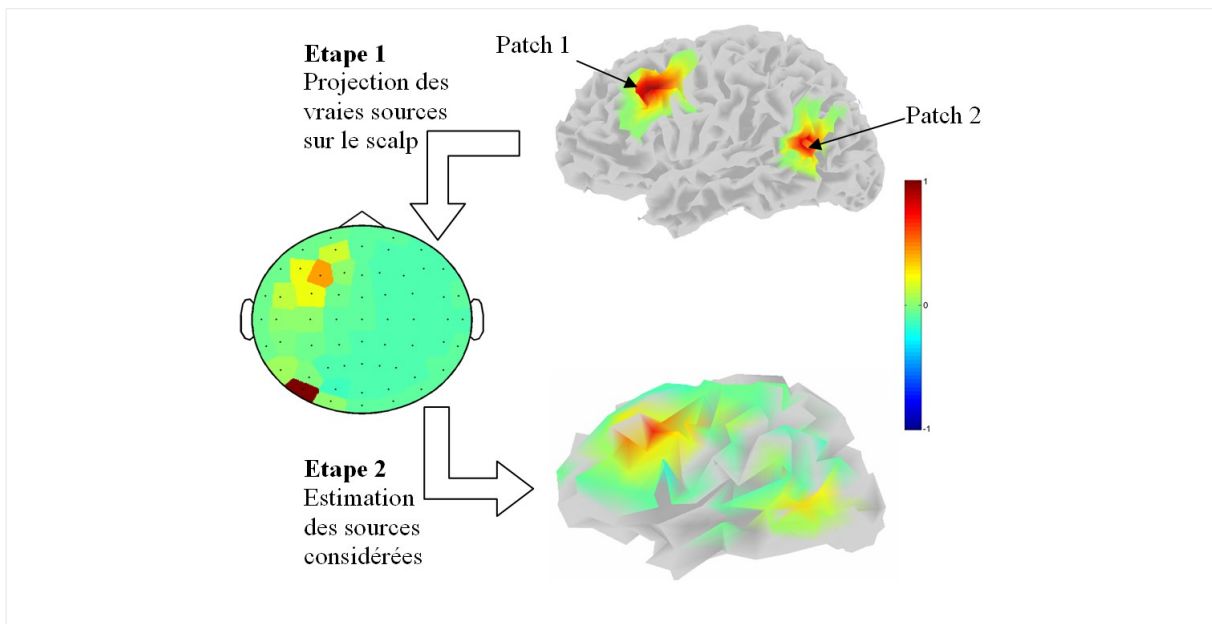


Fig. 18 – Principe de la simulation. Deux signaux utiles sont générés, répliqués chacun sur un groupe de dipôles avec une amplitude décroissante le long de la distance au foyer de ce groupe. Du bruit est ajouté en tout dipôle, puis le signal des dipôles (les sources) est projeté sur les électrodes situées sur le scalp à l'aide de la matrice directe. C'est la première étape (1), celle de construction des données simulées. Puis en appliquant la chaîne de traitements normale, vient l'étape de la reconstruction de sources (2) sur un maillage de dipôles différent de celui initialement utilisé pour la génération des données, en particulier moins précis (ici, environ 500 dipôles contre 10000 dans le maillage du haut).

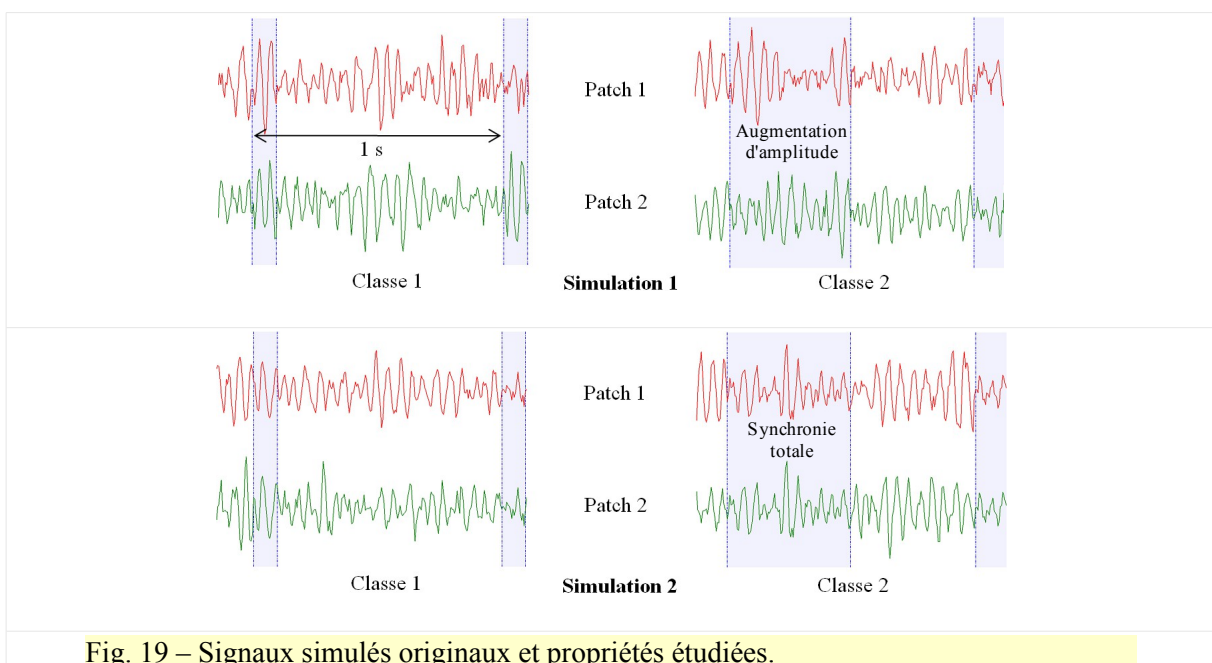


Fig. 19 – Signaux simulés originaux et propriétés étudiées.

La « simulation 1 » quant à elle créait un lien moins explicite avec dans l'état 1 une augmentation simultanée de l'amplitude des signaux par un facteur 1,2 ce pour une durée de 100 ms toutes les 1 s, et dans l'état 2 une même augmentation pour une durée de 500 ms toutes les 1 s.

Le signal de scalp était segmenté en fenêtres de 20 s, si bien que dans la classe 2, les signaux étaient en moyenne, à travers les 20 s de chaque fenêtre, plus amples dans la simulation 1 ou plus couplés dans la simulation 2, comparés à ceux de la classe 1.

Dans la simulation 1, une mesure de synchronie locale devrait permettre la détection d'une augmentation de l'amplitude moyenne pour les fenêtres de temps de la classe 2, en mettant à jour les deux sites d'activité utile sans donner d'information sur le caractère simultané du phénomène. Étant donné que seule l'amplitude des signaux a été modifiée, une mesure sensible uniquement à la phase comme la PLV ne devrait pas permettre la distinction du signal de l'état 1 de celui de l'état 2.

À l'opposé, dans la simulation 2, l'amplitude des signaux n'était pas modifiée. Par conséquent une mesure de synchronie locale devrait se montrer aveugle au changement d'état, alors qu'une mesure de couplage appliquée à la paire de signaux devrait quantifier avec succès la différence entre les deux états.

La différence entre les deux états d'une même simulation était seulement quantitative. Ce choix, en réalité bien souvent lapidaire pour la classification, était censé créer une similitude avec les vraies données auxquelles la procédure a ensuite été appliquée. En effet, les classes ont été construites comme correspondant aux valeurs respectivement hautes et basses d'une variable comportementale. À la différence des situations créées par les ICM, on n'avait pas nécessairement deux états bien différents qui alternaient, mais plus probablement un unique « état » qui était modulé.

Pour chaque simulation, les données ont été générées 20 fois de manière à obtenir des estimées robustes.

Trois quantifications différentes ont été étudiées : l'amplitude moyenne (locale), la cohérence (distante), et la PLV (distante). Comme expliqué dans la section 3.4.a) , l'amplitude moyenne exploite l'information d'amplitude du signal, la PLV utilise la phase, et la cohérence prend en compte les deux types d'information. Seule la bande de fréquences 7-13 Hz a été étudiée.

À la différence de l'étude précédente, la taille des fenêtres de temps n'a pas fait l'objet d'une étude, et a été fixé à 20 s. Dans la simulation, le choix de ce paramètre ne devrait pas avoir d'importance. Les différentes quantifications ont été utilisées séparément pour comparaison sur la base du taux de bonnes classifications. Enfin, les quantifications ont été calculées sur les sources reconstruites, mais également sur les signaux de scalp, pour comparer cette fois-ci les échelles d'observation, toujours par le biais de la classification.

Deux méthodes ont été employées pour la reconstruction de sources. Celle illustrée figure 18 a nécessité le calcul d'un maillage avec un nombre modéré de sommets, ce pour deux raisons. Premièrement, une limitation technique est posée par les mesures de couplage qui entraînent le calcul d'un nombre de variables proportionnels au nombre de couples ; par exemple 10014 sources donnent 50135091 couples de sources. Deuxièmement, l'utilisation pour la reconstruction des sources du même modèle de sources que celui utilisé pour générer les données aurait constitué un biais, car en situation réelle, aussi précise puisse être la segmentation de l'IRM, on ne modélise jamais parfaitement la distribution des sources du signal mesuré sur le scalp (Kaipio & Somersalo, 2005). En particulier, la résolution spatiale réelle de la colonne corticale ne peut pas être atteinte. Ainsi, le maillage original à 10014 sommets a été décimé spatialement avec la fonction *reducepatch* de Matlab pour obtenir un maillage à 504 sommets. Ce dernier maillage étant trop déformé par rapport à ce que l'on aurait pu espérer obtenir d'une nouvelle segmentation de l'IRM, chacun de ces sommets a été translaté sur la position du sommet le plus proche dans le maillage original. Pour la résolution du problème inverse par la méthode d'inversion avec minimisation de la norme, plusieurs valeurs du coefficient de régularisation α (voir section 5.3.e) pour rappel) ont été considérées, respectivement $\lambda=1, 5, 10, 15$ et 20 % de la plus grande valeur singulière de la matrice directe A . Pour indication, le choix arbitraire le plus courant dans la littérature est $\lambda=10\%$.

La seconde approche consiste simplement à ne retenir que 60 sources, autant que d'électrodes EEG, cela afin de ne pas induire un nombre de variables pour la classification différent entre le niveau scalp et le niveau cortical. Le choix de ces 60 sources s'est fait sur la base de la considération suivante : plusieurs sources contribuent au signal mesuré en une électrode, mais ces sources étant plus ou moins éloignées de l'électrode, leur contribution respective est variable. Ces sources ont donc été sélectionnées comme celles ayant chacune le plus fort coefficient pour l'électrode correspondante, dans la matrice directe A . Cette stratégie consistant à estimer dans un premier temps la source qui contribue le plus au signal de l'électrode revient à conserver l'essentiel du signal de l'électrode en réduisant la contribution des autres sources. Une telle technique fait penser à un « laplacien amélioré ». Nous avons fait l'hypothèse que le coefficient de régularisation aurait moins d'impact dans une telle configuration, et a été choisi égal à $\lambda=10\%$ de la plus grande valeur singulière de A .

Enfin, pour évaluer visuellement le pouvoir de localisation de chaque quantification, les poids SVM issus des apprentissages effectués lors de la procédure complète de validation ont été utilisés de la même manière que dans l'étude précédente, par normalisation puis moyennage à travers les itérations de validation croisée. Pour aller plus loin dans la validation de la procédure, la statistique de Student a été calculée sur les mêmes données, pour chaque variable, sans la segmentation de la validation croisée. Ces estimations sont comparées aux poids SVM, la principale différence conceptuelle étant

que les premières sont estimées de manière univariée alors que les seconds sont construits par une procédure multivariée. L'estimateur univarié est :

$$T_{score,i} = \frac{T_i - \overline{T_{p,i}}}{\sigma(T_{p,i})} \quad \text{avec} \quad T_i = \frac{\overline{X_i^{(1)}} - \overline{X_i^{(2)}}}{\sqrt{\frac{(n_1-1) \cdot \sigma^2(X_i^{(1)}) + (n_2-1) \cdot \sigma^2(X_i^{(2)})}{n_1 + n_2 - 2}}}$$

L'indice p dénote des permutations, réalisées au nombre de 100, des étiquettes de classe (marquée ici par les exposants (1) et (2)), créant ainsi une distribution de T à partir de fausses classes. Le véritable T pour la variable i est T_i . Celui-ci est donc centré-réduit sur la base d'une distribution de référence indépendante des classes.

Pour comparer plus spécifiquement la cohérence et la PLV des sources reconstruites sur le plan de la localisation des activités couplées, nous avons développé une technique descriptive, et une technique quantitative.

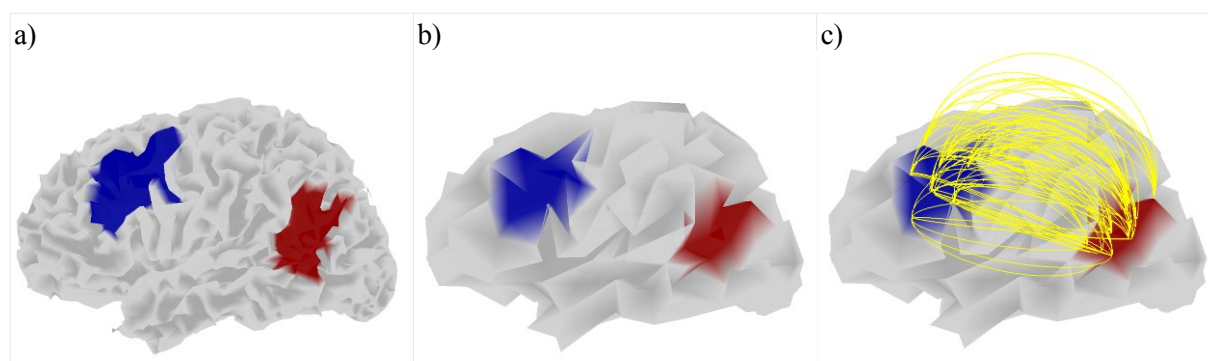


Fig. 20 – Base pour l'évaluation quantitative du pouvoir de localisation offert par les mesures de synchronie à distance au niveau cortical. Les régions à retrouver ont été définies sur le maillage complet (a, zones bleue et rouge) ; la première étape consiste donc à définir une adaptation de ces régions au maillage utilisé pour la reconstruction des sources (b). Pour ce faire, nous avons appliqué une simple règle du plus proche sommet voisin. Les régions étant définies, les couples de sources que l'on espère voir ressortir de l'ensemble des couples possibles par l'analyse sont ceux qui relient une source d'une région à une source de l'autre région (c, liens jaunes).

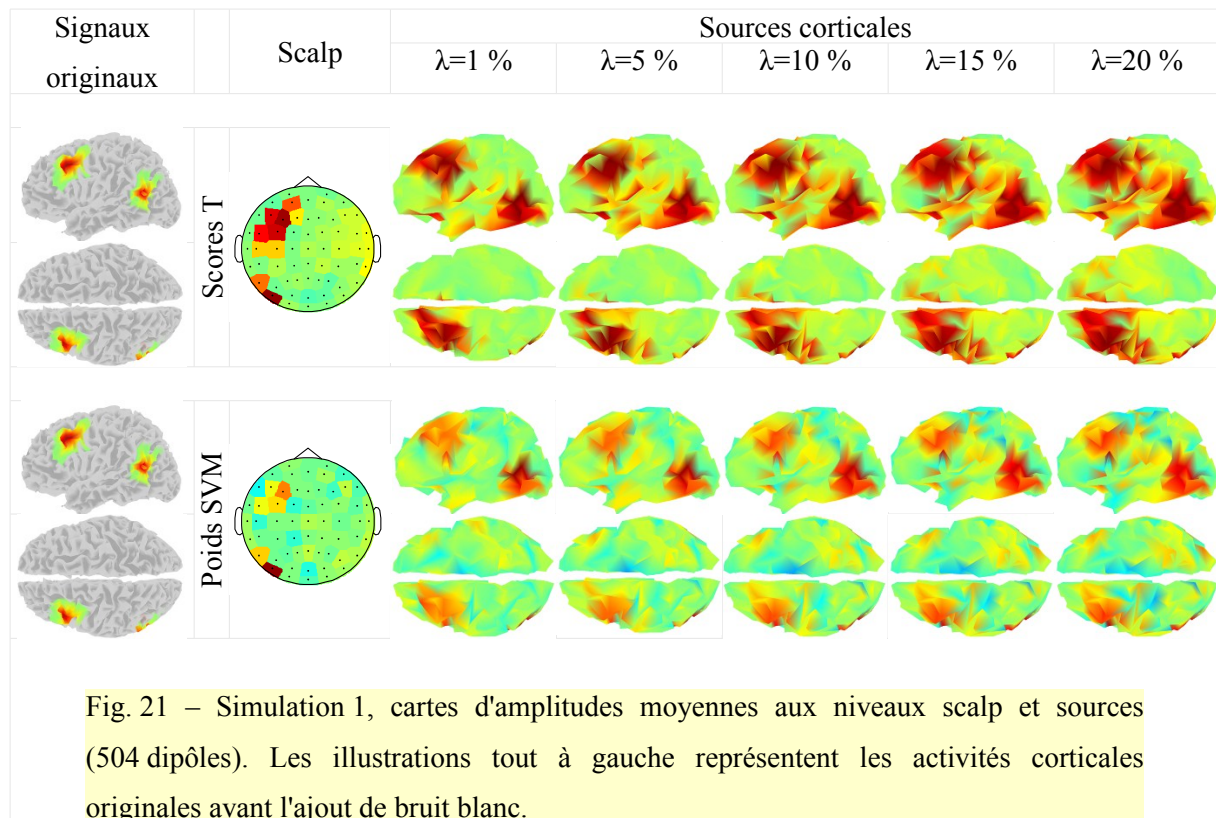
L'évaluation quantitative consiste à ordonner les poids SVM ou scores T par valeurs absolues décroissantes et à retenir le rang du premier poids ou score correspondant à un lien parmi les couples d'intérêt (la figure 20 illustre la détermination de ces couples d'intérêt). Cette procédure s'inspire des techniques de sélection de variables par lesquelles on s'intéresse aux seules variables les plus

discriminantes. Dans ce contexte, il est primordial qu'un lien parmi ceux recherchés soit « classé » parmi les plus intéressants.

La technique descriptive suit une procédure parente, basée elle-aussi sur les rangs : pour chaque source, les rangs des couples que fait cette source avec les autres sont sommés. Plus le rang total est grand, plus la source en question a été impliquée dans des phénomènes de synchronie à distance.

7.3. Résultats de la simulation

Les figures 21 et 22 présentent le pouvoir discriminant des amplitudes moyennes, quantifié de deux manières différentes, respectivement pour les simulations 1 et 2.



Quelle que soit l'échelle d'étude, la technique de contraste et la valeur du coefficient de régularisation (λ , exprimée en pourcentage de la plus grande valeur propre de la matrice directe A), les deux régions sont correctement localisées. Les poids SVM semblent rendre cette localisation plus focale et moins dépendante de λ . On note que des dipôles situés entre les deux régions cibles ont obtenu une part dans la discrimination.

Sur la figure 22, bien que l'on devine les deux régions sur le scalp, elles ne ressortent que modérément par rapport à la variabilité globale, en particulier avec les poids SVM. Au niveau cortical, une région

située plus ou moins à mi-distance des deux régions d'intérêt collecte plus d'importance dans la discrimination que les régions cibles.

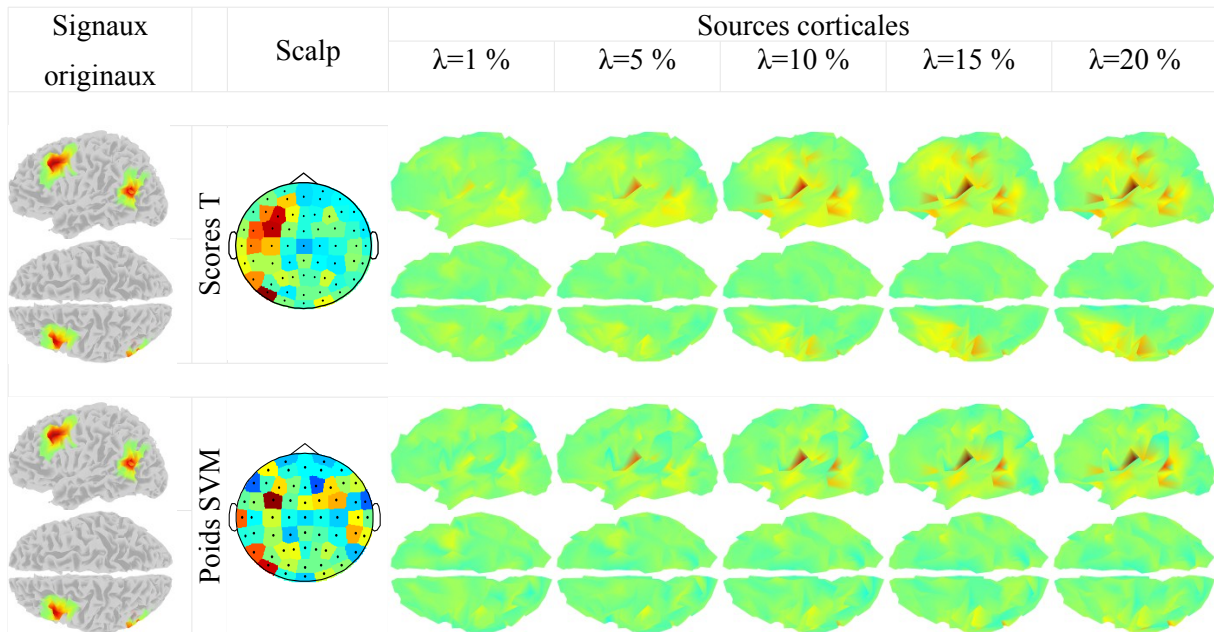


Fig. 22 – Simulation 2, cartes d'amplitudes moyennes présentées de la même manière que la figure 21.

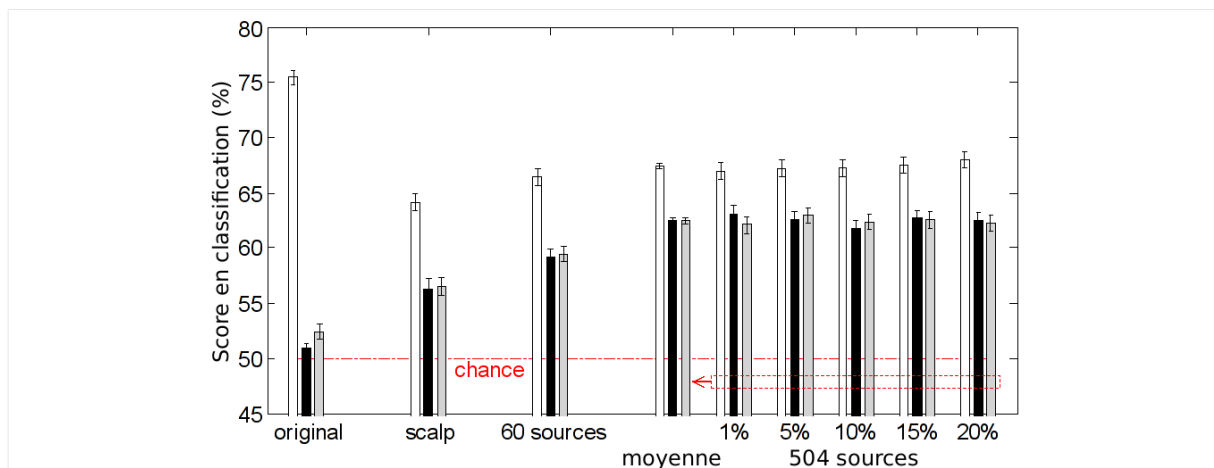
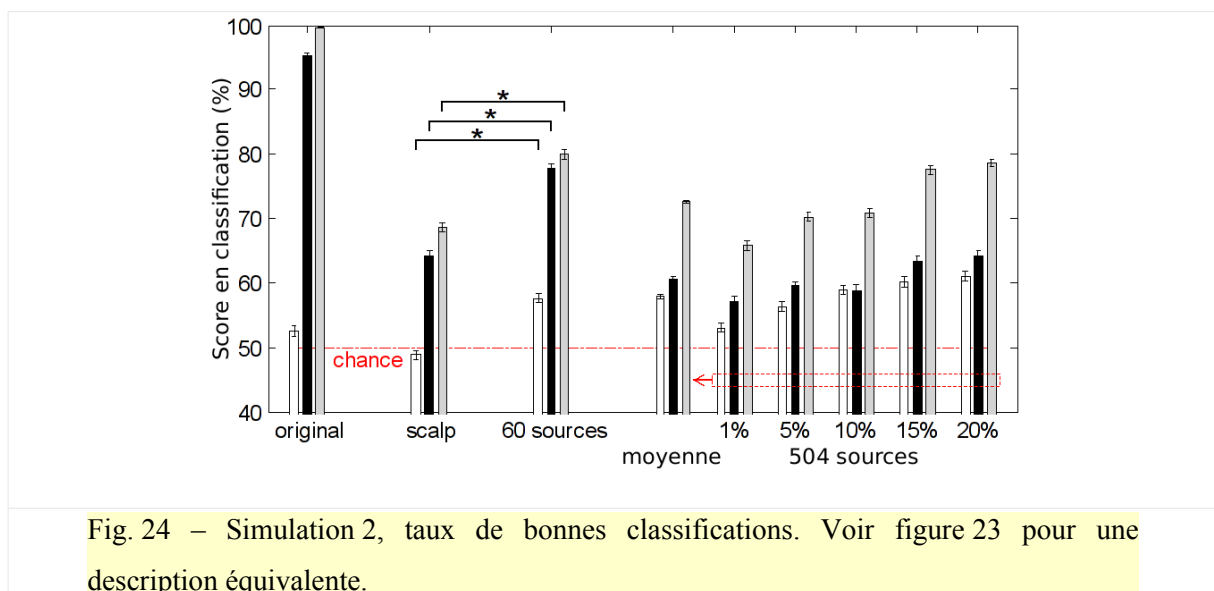


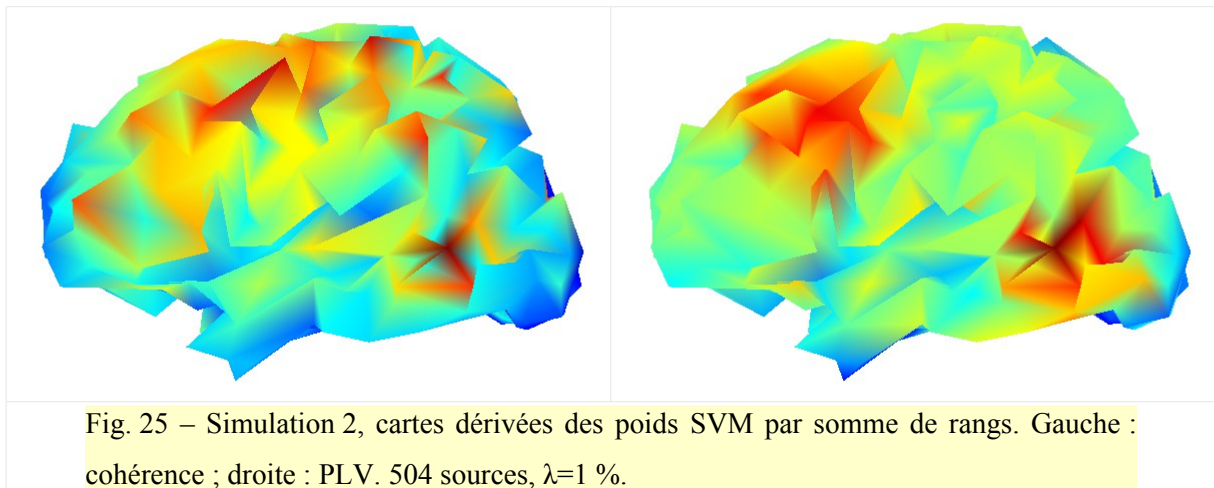
Fig. 23 – Simulation 1, taux de bonnes classifications. La couleur de barre correspond à la quantification : blanc pour l'amplitude moyenne, noir pour la cohérence et gris pour la PLV.

Sur la figure 23, le groupe de barres le plus à gauche correspond à la quantification des seules deux signaux originaux, sans ajout de bruit. Les taux de bonnes classifications illustrent donc les propriétés

des signaux en entrée de la simulation, sans les filtres que représentent la projection sur le scalp et la reconstruction des sources. Apparaissent également les taux de bonnes classifications au niveau du scalp avec ses 60 électrodes, au niveau cortical en ne prenant que 60 sources, et aussi en prenant 504 sources avec différentes valeurs pour le coefficient de régularisation (égal à λ fois la plus grande valeur singulière de la matrice directe A). La ligne discontinue rouge marque le niveau de la chance. Les différences de taux de bonnes classifications entre les niveaux scalp et 60 sources sont testés, quantification par quantification (test de Wilcoxon unilatéral pour échantillon apparié). Une étoile aurait marquée une différence significative (ici, il n'y en a aucune). Pour les reconstructions de 504 sources, les taux de bonnes classifications entretenaient une relation positive avec le coefficient de régularisation (test de Friedman), pour les amplitudes moyennes uniquement.



Sur la figure 24, les taux de bonnes classifications sur les 60 sources apparaissent significativement plus grands comparés à ceux obtenus avec les données de scalp, ce même pour les amplitudes moyennes. Sur 504 sources, toutes les quantifications ont une relation positive avec la valeur du coefficient de régularisation (tests de Friedman).



Les régions d'intérêt originales sont retrouvées par les deux quantifications de synchronies à distance, cependant les PLV permettent une estimation de ces régions à l'apparence meilleure que celle réalisée par les cohérences.

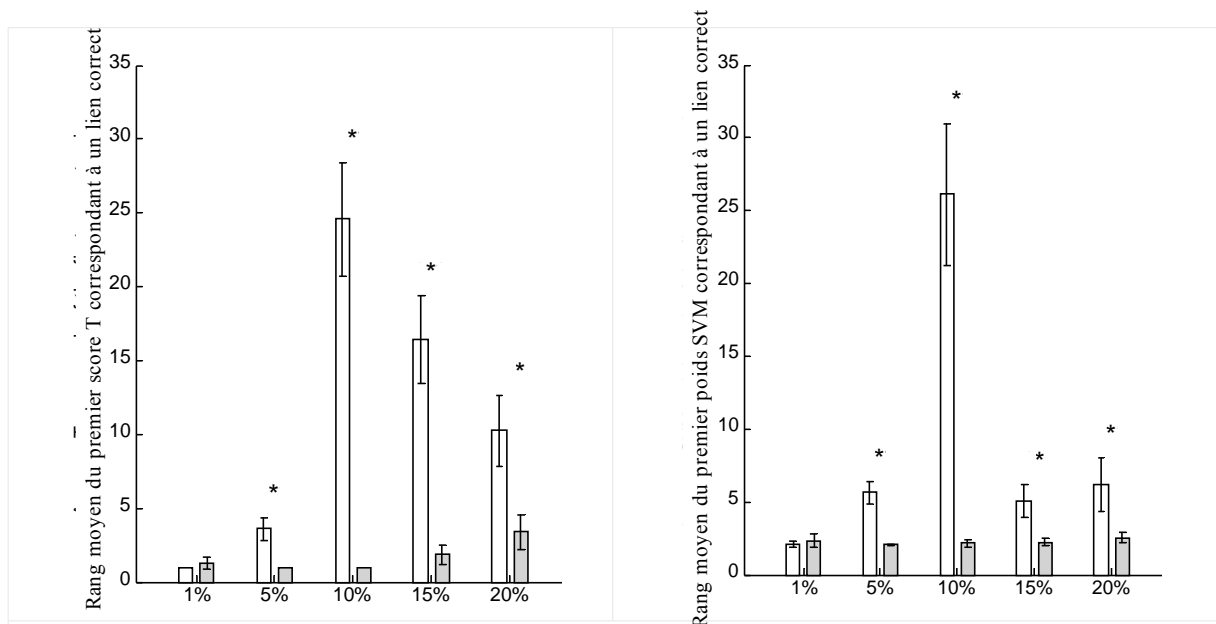


Fig. 26 – Simulation 2, quantification du pouvoir de localisation par rang moyen de la première variable correspondant à un lien correct. Abscisses : différentes valeurs pour le coefficient de régularisation utilisé à 504 sources. Barres blanches : cohérences ; barres grises : PLV. Partie de gauche : scores T ; partie de droite : poids SVM. Les étoiles marquent les situations où les PLV présentent un rang moyen significativement plus petit que celui des cohérences.

Les PLV permettent une localisation significativement plus « rapide » (si l'on imagine une procédure de recherche incrémentale qui consisterait à considérer les variables dans leur ordre d'importance) par

rapport aux cohérences, pour $\lambda=5, 10, 15$ et 20% , au sens du rang de la première variable correspondant à un lien correct.

7.4. L'expérience réelle

a) Introduction

Cette étude porte sur une expérience de plus courte durée. Il s'agit d'une tâche de compatibilité spatiale dans laquelle les stimuli étaient chacun une flèche orientée vers la gauche ou vers la droite, et positionnée dans la moitié gauche de l'écran ou dans la moitié droite. Le sujet avait à répondre à la direction de la flèche sans se laisser perturber par la position de la flèche, sachant qu'il répondait via un boîtier avec un bouton à droite et un bouton à gauche. L'effet attendu de ce type de tâche est une augmentation de temps de réponse lorsqu'il y a incompatibilité entre le caractère non pertinent du stimulus, à savoir sa position à l'écran, et la nature de la réponse, à savoir la position du bouton correspondant à la bonne réponse (plus exactement le doigt ou la main correspondants). Étant donné qu'il y a dans le même temps une incompatibilité entre les deux informations spatiales que contient le stimulus, un long débat au sujet de la contribution de l'un ou l'autre phénomène (l'incompatibilité stimulus-réponse ou la non-congruence du stimulus) dans l'augmentation du temps de réponse a animé la communauté scientifique. Il semblerait que l'incompatibilité stimulus-réponse l'ait emporté (Lu & Proctor, 1995).

b) Déroulement

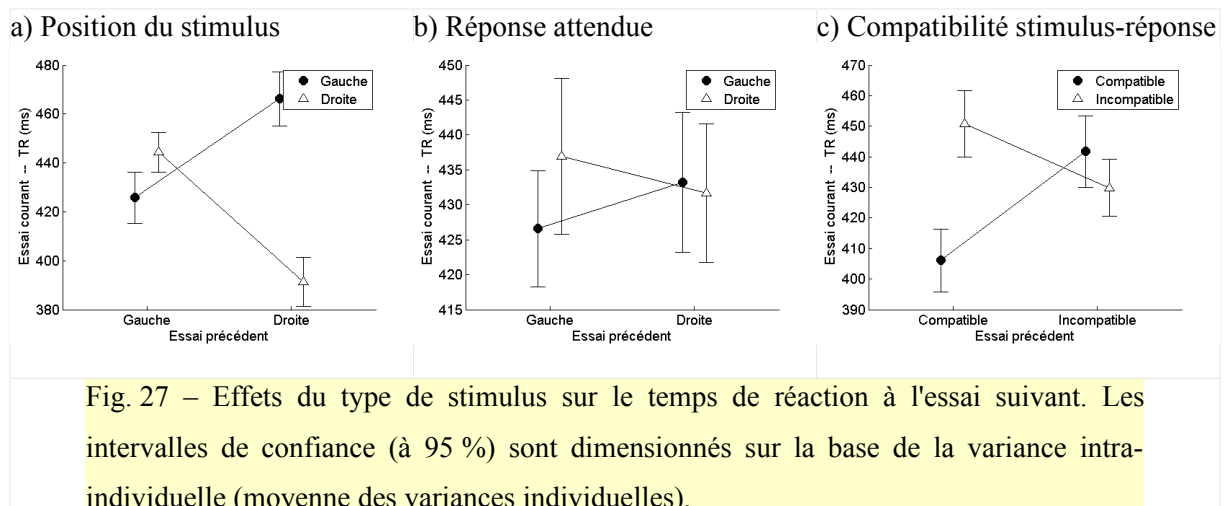
Une description détaillée du protocole expérimental est disponible ailleurs (Besserve *et al.*, 2008). Douze sujets ont participé à l'expérience qui avait lieu à la fois le matin et l'après-midi. Trois tâches cognitives différentes se succédaient, mais seule celle de compatibilité spatiale a été analysée ici. Les sujets réalisaient deux sessions d'enregistrement consécutives le matin, et deux sessions d'enregistrement consécutives l'après-midi. Ces sessions duraient chacune 10 min environ, et les sessions successives étaient espacées d'une courte période de repos d'environ 1 min. Les sujets devaient répondre aussi rapidement et correctement que possible à la direction de la flèche comme décrit précédemment, en répondant par un appui du pouce droit sur le bouton droit si la flèche était orientée à droite, du pouce gauche sur le bouton gauche si la flèche pointait à gauche. Les sujets avaient subi un entraînement auparavant. Les essais étaient au nombre d'environ 1 par seconde (intervalle inter-stimuli de 800 à 1200 ms) et étaient présentés pendant 170 ms. Pour chaque sujet, la première des quatre sessions a été omise des analyses, ceci afin de limiter la contribution d'un possible effet d'apprentissage résiduel.

7.5. Analyse du comportement

a) Observations

Les sujets ont fait très peu d'erreurs si bien que seul leur temps de réponse a été analysé et ensuite exploité dans la construction de classes liées à la performance.

Les sujets ont répondu en moyenne plus vite pour les essais avec compatibilité entre le stimulus et la réponse (une flèche à droite pointant vers la droite ou à gauche pointant vers la gauche). De même, les sujets, tous droitiers, ont répondu plus vite aux stimuli présentés à droite de l'écran qu'à ceux présentés à gauche.



De plus, le temps de réponse pour un essai était affecté par le type de l'essai précédent (Fig. 27). C'est le cas en particulier de la position du stimulus (par exemple, une flèche à gauche suivie d'une flèche à gauche donne lieu à une réponse plus rapide qu'une flèche à gauche suivie d'une flèche à droite, etc), et de la compatibilité stimulus-réponse (exemple : les sujets répondent moins vite sur un essai incompatible précédé d'un essai compatible que sur un essai compatible précédé d'un essai compatible).

b) Construction des classes

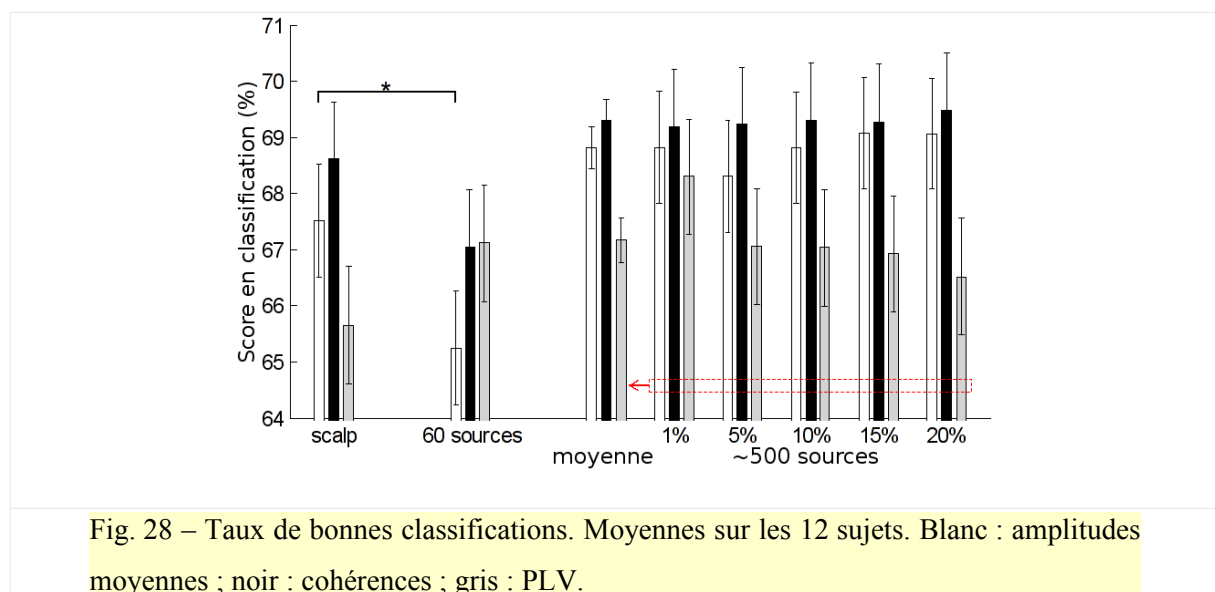
Pour chaque essai, le temps de réponse a été corrigé pour les effets observés, par centrage-réduction des temps de réponse des essais partageant les mêmes conditions expérimentales. Les différentes conditions considérées sont plus nombreuses que dans Besserve *et al.* (2008). Ensuite, pour chaque sujet, le signal a été segmenté en fenêtres de 20 s et le temps de réponse moyen a été calculé sur chaque fenêtre. Deux classes ont ensuite été construites l'une comme l'ensemble des fenêtres associées à un temps de réponse moyen bas, plus exactement inférieur au premier quartile des temps de réponse,

l'autre comme l'ensemble des fenêtres associées à un temps de réponse moyen élevé, c'est-à-dire supérieur au troisième quartile.

7.6. Analyse des données EEG

L'EEG constitué de 60 électrodes a ensuite été analysé par classification sujet-par-sujet des signaux originaux, des sources corticales reconstruites au nombre de 60 et des sources corticales reconstruites au nombre de 500 environ. Ce nombre n'était pas parfaitement égal à 500 ni égal d'un sujet à l'autre car il dépendait du maillage de la surface corticale, donc de l'IRM, et car la triangulation d'une surface est une approximation dont la qualité aurait souffert si nous avions contraint le nombre de sommets à une valeur exacte. Amplitudes moyennes, cohérences et PLV, toutes pour la seule bande 7-13 Hz, ont été utilisées indépendamment pour quantifier les activités aux trois échelles spatiales décrites.

Sur la figure 28, le seul effet significatif est une diminution des taux de bonnes classifications pour les amplitudes moyennes, entre le niveau du scalp et le niveau des 60 sources. Les cohérences offrent les meilleurs taux de bonnes classifications au niveau du scalp en comparaison avec les amplitudes moyennes et les PLV et décroissent également avec la reconstruction de sources. Seuls ceux des PLV augmentent du scalp aux 60 sources, ce de manière non significative.



7.7. Discussion

Les résultats en simulation semblent supporter l'idée que la reconstruction de source peut (ou pourrait) rendre l'information présente dans le signal EEG plus accessible aux quantifications usuelles, et par conséquent améliorer la classification. Dans d'autres études qui avaient présenté des augmentations de score de classification (multivariée), les sources étaient reconstruites et quantifiées à l'aide de mesures

de synchronie locale (Lotte *et al.*, 2007 ; Noirhomme *et al.*, 2008). Ici, une telle observation est confirmée et étendue aux mesures de synchronie à distance, du moins en simulation.

Toujours au cours des simulations, les taux de bonnes classifications au niveau des 504 sources ont exhibé un comportement particulier : plus le coefficient de régularisation était élevé, c'est-à-dire plus le lissage spatial était fort, plus les taux de bonnes classifications étaient élevés aussi. L'explication de ce phénomène réside probablement dans la redondance accrue des signaux utiles dans les sources reconstruites et le comportement multivarié du classifieur. En effet, il semblerait que plus un motif potentiellement utile à la classification est répliqué dans l'activité quantifiée, plus le nombre de variables individuellement discriminantes serait grand, et plus la discrimination à l'aide de toutes ces variables prises ensemble serait aisée, cela tant que les taux de bonnes classifications ne sont pas encore maximaux.

Les taux de bonnes classifications étant établis sur des données de test, ils ne peuvent pas être augmentés artificiellement par un éventuel effet de sur-apprentissage du fait qu'ici tous les motifs communs aux données d'apprentissage et de test sont bien ceux que l'on cherche. Il n'y a pas d'effet ajouté qui se superposerait par (mal-)chance à l'effet recherché. Cependant, la comparaison entre les niveaux scalp et cortex a été restreinte à un nombre de sources égal au nombre d'électrodes (60), cette contrainte ayant été appliquée aux vraies données.

À ce sujet, la reconstruction de sources pour les vraies données n'a pas donné lieu à une quelconque amélioration significative du taux de bonnes classifications. Tentons dans un premier temps d'expliquer cet échec.

La technique employée pour la reconstruction des sources s'étant montrée pertinente au cours des simulations, on est tenté de penser que, soit les sources qui contribuaient à la discrimination pour les données réelles n'étaient pas corticales, soit les motifs d'activité simulés n'étaient pas réalistes.

En effet, pour ce qui concerne le premier point évoqué, la modélisation dite directe de la contribution des sources dans l'activité EEG ne considérait que des sources du cortex cérébral. Peut-être des sources localisées dans des structures cérébrales profondes ont-elles aussi contribué. Mais plus vraisemblablement, ce serait des sources extra-crâniennes, c'est-à-dire les sources d'artefacts que l'on s'évertue traditionnellement à supprimer de l'activité EEG analysée, qui auraient pu être « bruitées » car « reconstruites » au mauvais endroit, par « dilution » dans des activités électriques d'origine cérébrale. En effet, il a été montré que le maintien de la performance sur une tâche classique de laboratoire nécessitait un effort mental dont les manifestations comportementales les plus apparentes seraient une activité de contraction plus importante de certains muscles du visage (van Boxtel & Jessurun, 1993 ; Waterink & van Boxtel, 1994). Ainsi l'activité électrique d'origine musculaire bruyait celle d'origine cérébrale et réciproquement, la réciproque étant d'autant plus marquée que les

sources seraient reconstruites. Le signal électrique d'origine musculaire, pourtant potentiellement utile, serait dégradé.

En ce concerne la seconde hypothèse, celle d'un manque de similarité entre les motifs d'activité utile simulés et ceux présents dans les vraies données, il est plus probable que l'« imperfection » des simulations réside dans la construction spatiale des régions d'intérêt plutôt que dans la dynamique des activités. Deux types de dynamique ont été étudiés et les états ont été construits comme deux niveaux de modulation d'un même phénomène, ceci afin de rendre les simulations plus proches des vraies données pour lesquelles les classes choisies correspondaient elles-aussi à deux niveaux différents d'une seule grandeur quantitative. En revanche, dans les simulations, les deux régions, de forme simple, d'étendue relativement vaste et situées à bonne distance l'une de l'autre, représentent un cas idéal probablement éloigné de la réalité des vraies données dans lesquelles la distribution spatiale des activités discriminantes est plus complexe. Une distribution spatiale complexe d'activités initialement décorréelées devrait donner lieu à une activité de scalp plus « mélangée ». Paradoxalement, c'est dans ce cas que l'on imagine le mieux l'intérêt des techniques de reconstruction de sources, puisque ces dernières ont été développées pour séparer les sources les unes des autres, dans une certaine mesure.

À ce sujet, la technique de l'inversion avec contrainte de norme minimale est une des techniques inverses les plus simples. Le choix de l'utiliser avait été motivé par la relative simplicité des calculs qu'elle impliquait. D'autres techniques qui s'appuieraient plus sur des caractéristiques du signal EEG, en particulier sur la dynamique du signal (Baillet & Garnero, 1997), pourraient bien améliorer grandement la qualité de la reconstruction des sources sous l'objectif d'une analyse de la dynamique des sources. Enfin, une parenthèse est présentement ouverte concernant une famille de techniques inverses dont la plus connue l'est sous le nom de sLORETA (*standardized low-resolution brain electromagnetic tomography*; Pascual-Marqui, 2002 ; voir aussi Dale & Sereno, 1993, pour une technique parente). Les solutions obtenues par ces méthodes ne diffèrent de l'inverse de norme minimale que par un facteur d'amplitude constant dans le temps, ce indépendamment pour chaque source. Aucun mélange supplémentaire n'est opéré, si bien les amplitudes moyennes n'apportent aucune information supplémentaire pour une analyse linéaire multivariée, tandis que les cohérences et les PLV sont strictement les mêmes d'une technique à l'autre. Ainsi, les résultats en classification sont parfaitement égaux (toujours d'une technique à une autre, en incluant l'inversion avec contrainte de norme minimale).

Suite au développement précédent, dans la recherche d'une explication des résultats médiocres apportés par la reconstruction de sources sur les vraies données, nous privilégions la piste de sources pas ou mal prises en compte dans le modèle initial, ce dernier ne s'intéressant qu'au seul cortex cérébral. Ce point de vue aidant, un fait intéressant se dégage des résultats en classification sur les vraies données. Premièrement, au niveau du scalp, les amplitudes moyennes et les cohérences donnent

lieu à de meilleurs taux de bonnes classifications que ceux pour les PLV, ce qui semble indiquer une importance plus grande de l'information d'amplitude que de celle de phase. Deuxièmement, les cohérences paraissent plus performantes que les amplitudes moyennes et les PLV permettent des taux de bonnes classifications qui, quoique inférieurs, talonnent ceux des amplitudes moyennes et des cohérences. Parmi les synchronies discriminantes, certaines seraient donc distantes et l'information de phase serait utile à la classification en dépit de l'observation précédente. Enfin, les taux de bonnes classifications reposant sur les amplitudes moyennes enregistrent une plus forte baisse que ceux liés aux cohérences. Du fait que les cohérences s'appuient à la fois sur des informations d'amplitude et de phase, on est tenté de penser que les taux de bonnes classifications correspondants obtenus sur le scalp doivent beaucoup à des modulations d'amplitude, et que la chute observée du taux de bonnes classifications pour les amplitudes moyennes après reconstruction de sources ait entraîné un même phénomène pour les cohérences. Parallèlement à cela, les taux de bonnes classifications pour les PLV ont augmenté, certes de manière non significative, et encore une fois, les cohérences, qui doivent avoir extrait des informations en partie commune avec celles exploitées par les PLV, ne présentent pas une aussi forte chute des taux de bonnes classifications que les amplitudes moyennes. L'impression résultante est que les amplitudes ont souffert de la reconstruction de sources, tandis que la phase en a bénéficié. Or, les modulations de phase non accompagnées de modulations d'amplitude sont plutôt la marque de phénomènes de dynamique complexe, telle que l'activité cérébrale, par opposition aux activités électriques originaires des sources extra-crâniennes. Les mécanismes de la contraction musculaire ou du mouvement du dipôle qui représente électriquement l'œil paraissent en effet beaucoup plus simple, ne serait-ce que par le moindre nombre de sources électriques sous-jacentes aux activités mesurées, ou encore par la « communication » plus limitée entre ces sources.

8 – Étude 3 : simulation de vol

8.1. Introduction

a) Situation

La fatigue mentale est fréquemment expérimentée par les opérateurs de l'aviation (Cabon *et al.*, 1993 ; Neville *et al.*, 1994). Une expérience a donc été élaborée pour l'étude de la fatigue des pilotes d'aéronefs militaires.

Cette expérience s'est déroulée à l'École de Transformation Opérationnelle (ETO) sur la base aérienne de Cazaux (BA120). Les élèves-pilotes de cette école ont réalisé une simulation de vol de près de 3 h à l'intérieur d'un simulateur d'Alphajet français. En plus de ces jeunes pilotes, l'équipe médicale de la base a été impliquée, ainsi que les moniteurs dont le rôle était de transmettre les consignes au cours de la simulation et, en temps normal, d'évaluer les élèves-pilotes.

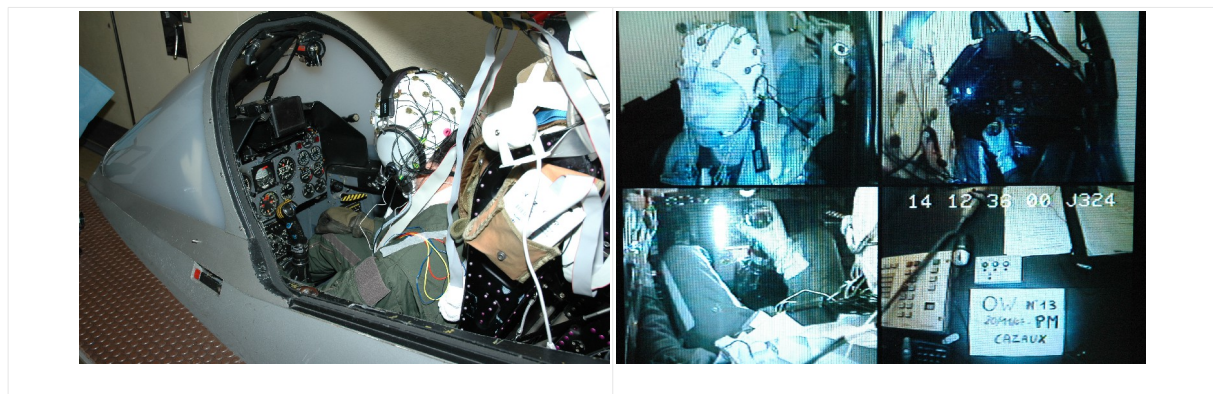


Fig. 29 – Simulateur d'Alphajet et partie de l'équipement déployé. Pendant la simulation, le pilote et les principaux actionneurs du simulateur étaient filmés, ainsi que certains des moyens de commande que le moniteur avait à sa disposition pour transmettre les ordres de manœuvre au pilote.

b) Un scénario pour définir un objectif

Imaginons une situation fictive, celle d'une mission aérienne dans laquelle la possibilité d'un abandon de la mission au moindre doute sur l'état de fatigue du pilote est envisagée.

Un aéronef décolle et entame un long vol en direction du site de sa mission. Pendant le vol, le pilote de l'aéronef n'a rien d'autre à faire que de maintenir les paramètres de vol constants. À l'approche du site de la mission, l'aéronef entre dans une zone à risque, par exemple une région sur laquelle on possède peu d'informations, mais telle que la tâche de pilotage ne change pas pour autant. Lorsque l'aéronef

arrive enfin sur le site, le pilote réalise sa mission, ce qui représente la phase d'exposition maximale au risque.

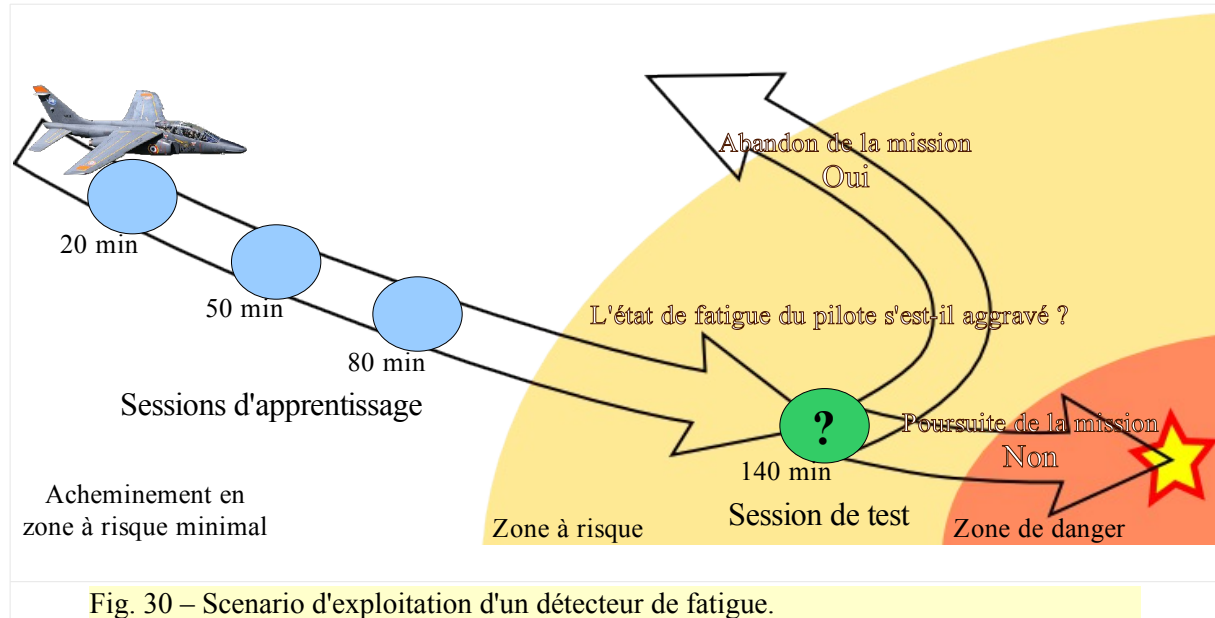


Fig. 30 – Scénario d'exploitation d'un détecteur de fatigue.

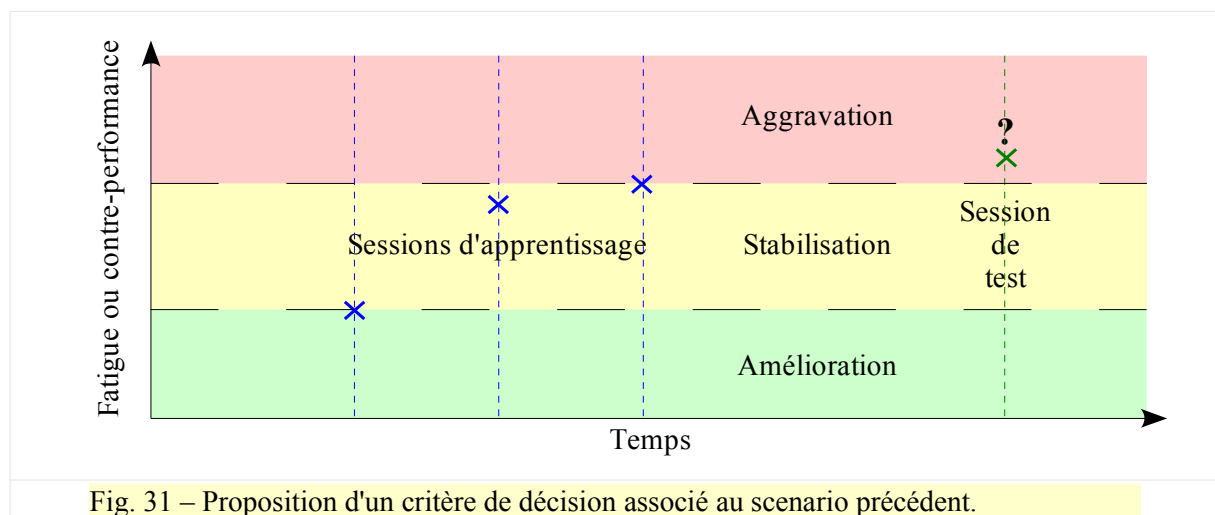
Imaginons que l'on puisse mesurer l'état de fatigue du pilote ou sa performance, au cours de la première phase du vol, par exemple en faisant réaliser au pilote par courtes périodes une tâche qui ne se différencierait pas qualitativement de la tâche principale de vol, ou tout simplement en quantifiant le respect des consignes de vol. Cela pourrait être envisageable lorsqu'aucun danger ne serait à craindre et qu'une éventuelle (légère) augmentation de la charge de travail du pilote serait sans conséquence, voire même faciliterait le maintien de l'éveil.

À l'entrée dans la zone à risque, la probabilité croissante d'un danger entraîne un changement prévu de priorités ; plus question par exemple d'occuper le pilote avec la tâche qui permettait la mesure de son état de fatigue ou de performance. C'est pourtant à ce stade qu'une ultime mesure serait souhaitable de manière à décider ou non de la poursuite de la mission, avant d'entrer dans la zone de risque maximal.

Ainsi, cette avant-dernière phase représenterait une session de test, tandis que les périodes d'enregistrement précédentes, pendant lesquelles la fatigue ou la performance étaient mesurées, représenteraient les sessions d'apprentissage. L'opération d'apprentissage aurait lieu entre temps.

Précisons tout de suite que, pour qu'un détecteur qui exploite des mesures électrophysiologiques puisse fonctionner dans de telles circonstances, il faut que la tâche de pilotage dans la zone dite « à risque » ne soit ni trop différente de celle qui précédait (pas de vol à basse altitude par exemple), ni trop stressante, les émotions provoquant d'importants changements dans l'activité de la plupart des organes d'étude.

L'enjeu est ici de prédire dans l'avenir par rapport à (et par comparaison avec) des situations où la fatigue peut être peu marquée. Sans nécessairement changer le fonctionnement du détecteur par fenêtres de temps glissantes, l'objectif est de savoir si le niveau moyen de fatigue prédit sur la session de test dépasse le niveau moyen de fatigue le plus élevé observé sur les différentes sessions d'apprentissage, ou s'il reste dans l'intervalle des valeurs observées, ou encore, pourquoi pas, s'il s'avère inférieur à la valeur précédente la plus basse (Fig. 31).



c) Quelques directions à explorer

L'intérêt de l'étude présentée ici réside dans l'adaptation, à ce contexte réaliste de simulation de vol, de la démarche méthodologique propre à la classification et à la validation. Nous espérons donner plus de sens aux résultats liés à la démarche générale, par exemple à la notion de taux de bonnes classifications, et ainsi convaincre les non-initiés du potentiel des techniques d'analyse de données employées.

Sans entrer dans les détails, trois innovations méthodologiques ont été mises en œuvre.

Premièrement, la procédure de validation a été adaptée au scénario présenté plus haut. En particulier, les données de test étaient choisies comme toujours ultérieures aux données d'apprentissage, avec un délai variable entre la fin des données d'apprentissage et le début des données de test. Cette « innovation » est à entendre par rapport aux études précédentes, car de telles pratiques ont probablement été appliquées par d'autres auteurs.

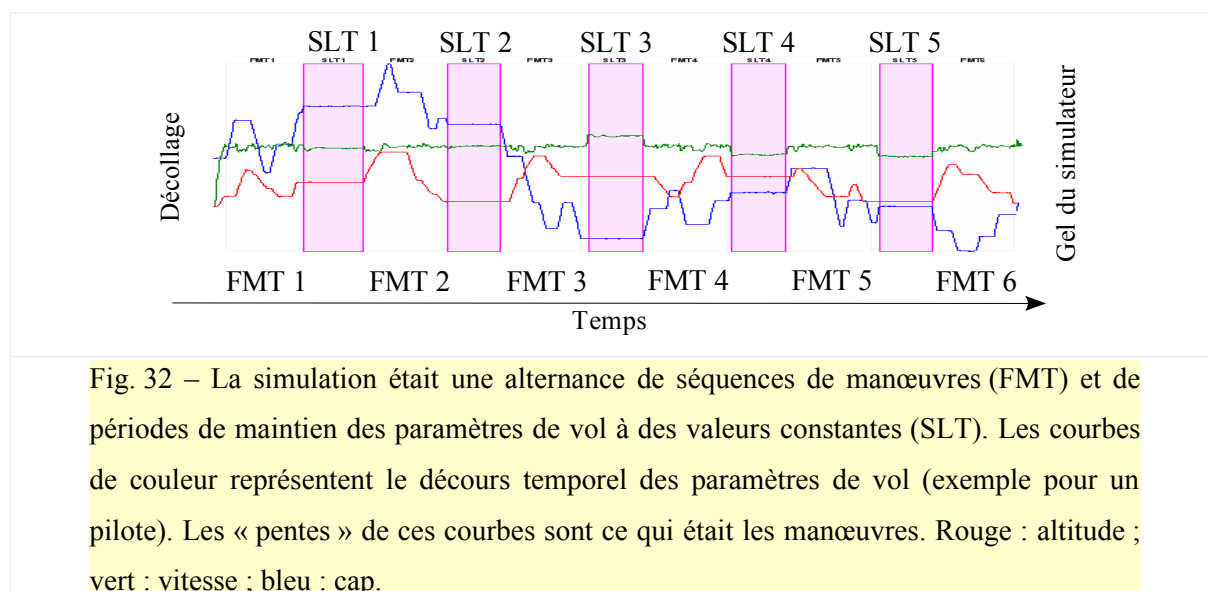
Ensuite, nous avons privilégié cette fois-ci encore une approche de prédiction de niveaux de performance. Des performances hautes ont été discriminées de performances basses, ce qui impliquait de définir une frontière entre ces deux « domaines de fonctionnement ». L'absence d'une quelconque information sur la valeur théorique de cette frontière putative et la nature relative du critère de décision

évoqué plus haut (« y-a-t'il amélioration, stabilisation ou aggravation ? ») laissent ouvert un vaste champ de manœuvre. La deuxième innovation consistait à ne pas complètement prédéterminer les classes. La procédure développée réalisait un effort d'adaptation des classes aux données électrophysiologiques de manière à ce que les premières soient mieux exprimées dans les secondes. En particulier, l'approche proposée dans cette étude laissait se former les hétérogénéités naturelles d'effectifs entre les niveaux de performance. Ainsi, les bonnes et mauvaises performances n'étaient plus nécessairement aussi nombreuses les unes que les autres, et le rapport d'effectifs choisi entre les classes était celui que les données électrophysiologiques « expliquaient » le mieux. En revanche, ces cas de classes dites déséquilibrées pouvaient entraîner des biais. Ces derniers ont été modérés grâce à des optimisations qui constituent la troisième innovation évoquée plus haut.

8.2. Méthodes

a) Données de la simulation

Treize élèves-pilotes s'entraînaient pour la première fois sur ce simulateur et totalisaient chacun environ 300 h de vol. Les pilotes étaient des hommes, dont 12 droitiers et 1 gaucher, et ont tous tenu le manche avec la main droite. Les pilotes ont déclaré avoir effectué une nuit normale avant le jour de la simulation (durée moyenne de sommeil $6,4 \pm 0,6$ h). Ils ont réalisé la même simulation les uns après les autres au rythme d'une simulation (d'un pilote) par demi-journée.



La simulation se déroulait suivant une séquence de phases de vol utilisée pour l'entraînement des élèves-pilotes (Fig. 32). Après le décollage, le pilote réalisait un segment de vol d'environ 17 min

constitué d'une série de manœuvres (*Flight Maneuvers Task*, FMT, en anglais) qui portaient chacune sur un ou deux paramètres de vol. Puis succédait un segment de vol de maintien des paramètres de vol d'une durée d'environ 10 min (*Straight and Level flying Task*, SLT, en anglais). Le pilote réalisait ensuite un deuxième FMT, puis un deuxième SLT et ainsi de suite jusqu'au sixième FMT au terme duquel la simulation prenait fin (pas d'atterrissage). Cette construction est proche de celle utilisée par Morris et Miller (1996). La simulation durait environ 2 h 45.

En dehors des manœuvres, le principal critère d'évaluation de la performance des pilotes utilisé par le moniteur qui supervisait la simulation était lié au respect des consignes sur les paramètres de vol. Les pilotes devaient maintenir chacun des paramètres de vol à l'intérieur d'une fenêtre de valeurs centrée sur la valeur de consigne donnée au pilote par le moniteur. L'écart maximal à cette valeur centrale était prédéterminé. Celui-ci, appelé tolérance, valait 5 ° pour le cap, 150 pieds pour l'altitude et 10 nœuds pour la vitesse indiquée. Une excursion hors de cette fenêtre de valeurs pendant plus de 15 s signalait une mauvaise performance.

b) Comportement

Le comportement, c'est-à-dire l'information issue des trois paramètres de vol, a fait l'objet d'une analyse préliminaire sur les 5 SLT. Le critère d'excursion en dehors de la fenêtre de valeurs autorisées pendant plus de 15 s n'était pas exploitable pour une telle analyse, car les événements que marque ce critère étaient rares.

L'écart-type de chaque paramètre de vol a été calculé sur la totalité de chacun des SLT, puis, pour chaque paramètre de vol, les écart-types ont été centrés-réduits à travers les SLT. Les mesures ont été enfin moyennées pour chaque SLT à travers les trois paramètres de vol, de manière à n'obtenir qu'une valeur par SLT. Cet indice est appelé E dans la suite du texte, par emprunt à l'indice plus complexe de Morris et Miller (1996).

$$E = \frac{1}{P} \sum_{p=1}^P \frac{E_p - \overline{E}_p}{sd(E_p)}$$

$$\text{avec } E_p = sd(X_p)$$

Où X_p est le décours temporel du paramètre de vol p au cours d'un SLT.

Chaque SLT a ensuite été segmenté en fenêtres de 10 s, 10 s semblant être une durée adéquate à la fois pour l'observation des fluctuations de la variabilité dans la tenue des paramètres de vol et pour la quantification des activités électrophysiologiques. Un indicateur de performance proche du précédent a été calculé sur chaque fenêtre de temps suivant la procédure suivante : l'écart-type de chacun des trois paramètres de vol a été calculé et divisé par la valeur de tolérance associée au paramètre de vol, puis les valeurs estimées ont été moyennées. Cet indice est appelé E_{10} dans la suite du texte.

$$E_{10} = \frac{1}{P} \sum_{p=1}^P \frac{sd(X_{p,10})}{T_p}$$

Où $X_{p,10}$ est le décours temporel du paramètre de vol p au cours d'une fenêtre de 10 s et T_p est la tolérance associée au paramètre de vol p .

Les valeurs moyennes de ce second indice sur les SLT (notées \bar{E}_{10}) ont été comparées à celles du E . Ces deux indices sont représentés sur une même figure après que chacun d'eux ait été normalisé par centrage-réduction sujet-par-sujet, étant donné que les échantillons de chaque indice sont appariés par SLT.

Une seconde représentation du E_{10} est donnée pour permettre la visualisation du décours temporel de cet indice. Afin d'en faciliter la visualisation, les séries temporelles ont été tronquées à 10 min par SLT puis ont fait l'objet d'une normalisation par centrage-réduction à travers le temps (les fenêtres de temps des SLT réunis) pour chaque pilote, et enfin elles ont été moyennées à travers les pilotes.

Ces indices sont supposés quantifier une notion d'erreur et, bien que la notion de performance soit plus souvent évoquée dans la suite du texte, une « bonne performance » correspond à une valeur petite de ces indices et une « mauvaise performance » correspond à une valeur grande.

c) Prétraitements et quantifications

L'EEG a été prétraité selon la procédure de la section 5.3.d) puis sur chacune des fenêtres de 10 s, l'amplitude moyenne pour chacune des 32 électrodes et des 3 bandes de fréquences (3-7 Hz, 7-13 Hz et 13-18 Hz) a été calculée, de même que les PLV entre paire d'électrodes aux mêmes bandes de fréquences. Les quantifications de l'EOG et de l'ECG n'ont pas été exploitées.

d) Validation

La procédure de validation (non croisée) a été décomposée en 10 configurations décrites dans le tableau de la figure 33. Cette construction repose sur la structure des données en SLT, et le principe de cette construction consiste dans la postériorité temporelle des données de test sur les données d'apprentissage. Un minimum de deux SLT formait les données d'apprentissage afin que l'algorithme d'apprentissage extraie des motifs autant reproductibles d'un SLT à un autre que possible. D'autres constructions auraient été possibles ; les deux « paramètres » qui ont guidé ces choix étaient la quantité des données d'apprentissage et le délai entre la fin dans le temps des données d'apprentissage et le début des données de test.

Pour un SLT de test qui suivait le dernier SLT d'apprentissage, l'intervalle était en théorie de 17 min. Lorsqu'un SLT séparait celui de test et le dernier d'apprentissage, l'intervalle était en théorie de 44 min. Enfin, lorsque deux SLT s'intercalaient, l'intervalle était de 71 min.

Configuration	SLT				
	1	2	3	4	5
1	A	A	T		
2	A	A		T	
3	A	A			T
4	A	A	A	T	
5	A	A	A		T
6	A	A	A	A	T
7		A	A	T	
8		A	A		T
9		A	A	A	T
10			A	A	T

Légende : A Apprentissage T Test

Fig. 33 – Configurations de répartition des données entre ensemble d'apprentissage et ensemble de test.

e) Construction des classes

Deux classes étaient définies comme les ensembles de fenêtres de temps de 10 s pour lesquelles le E_{10} était supérieur (classe des « mauvaises performances » ou -1) ou inférieur (classe des « bonnes performances » ou +1) à un seuil.

Le choix du seuil faisait l'objet d'une procédure automatique complexe. Neuf valeurs pour ce seuil étaient prédéfinies comme égales respectivement aux neuf premiers déciles du E_{10} mesuré sur les seules données d'apprentissage. Une procédure de validation croisée permettait la sélection d'un seuil parmi les neuf. La construction de classes correspondante était ensuite appliquée aux données de test.

Selon le seuil sélectionné pour la construction des classes, le rapport d'effectif d'une classe sur l'autre pouvait atteindre 9 : 1. Des classes ainsi déséquilibrées peuvent induire des biais dans l'apprentissage du classifieur car dans une telle situation, un classifieur peut atteindre de meilleurs taux de bonnes classifications en répondant systématiquement en faveur de la classe majoritaire plutôt qu'en cherchant réellement à séparer les classes. En effet, les quantifications utilisées permettent un niveau initialement inconnu de séparabilité des classes, potentiellement inférieur à la proportion de la classe majoritaire. Deux optimisations ont donc été intégrées à la procédure pour limiter cet effet.

La première optimisation consistait à pondérer les erreurs modélisées dans la formulation du problème d'apprentissage du SVM. Pour rappel, l'algorithme d'apprentissage du SVM est le suivant :

$$\begin{aligned} \min_{\mathbf{w}, b} c \cdot \sum_i \epsilon_i + \frac{\|\mathbf{w}\|^2}{2} \\ (\langle \mathbf{w} | \mathbf{x}_i \rangle + b) \cdot y_i \geq 1 - \epsilon_i, \forall i \\ \epsilon_i \geq 0, \forall i \end{aligned}$$

Les ϵ_i sont des erreurs que l'algorithme cherche à minimiser. Il a déjà été proposé d'utiliser un c différent pour chacune des deux classes (Veropoulos *et al.*, 1999 ; Morik *et al.*, 1999). Nous proposons de manière équivalente de réécrire le terme $\sum_i \epsilon_i$ par $\langle a | \epsilon \rangle$ où a est un vecteur tel que

$a_i = \frac{n}{2n_{y_i}}$ et n_{y_i} est l'effectif de la classe y_i du point x_i . Ainsi $\sum_i a_i = 1$ donc la somme des a_i pour tout i correspondant à une même classe ne dépend pas de la classe. Dans le même temps, c conserve sa signification initiale.

Pour exploiter plus en profondeur cette idée de pondération des erreurs, le cumul des poids des erreurs a été équilibré également entre la première moitié de chaque SLT d'apprentissage et la deuxième moitié correspondante. Pour cela, les SLT ont été découpés en deux parts d'effectifs égaux. Ainsi, pour k SLT utilisés en apprentissage, $a_i = \frac{n}{4kn_{y_i, z_i}}$ où n_{y_i, z_i} est l'effectif de la classe y_i dans le demi-segment z_i auquel appartient le point x_i .

Cette procédure de pondération a en revanche nécessité de sélectionner les variables de manière à en réduire le nombre, ceci afin que les ϵ_i ne soient pas tous nuls. En effet, lorsque le nombre de variables est supérieur au nombre d'éléments à classer (les fenêtres de temps), il est probable que les données d'apprentissage soient linéairement séparables, qu'il n'y ait donc aucune erreur d'adéquation aux données.

L'étape de sélection de variables précédait celle d'apprentissage. Le T de Student (sans permutation) a été estimé pour chaque variable à l'aide des données d'apprentissage, puis les variables ont été sélectionnées par T décroissants en valeur absolue, au nombre plus ou moins arbitraire du tiers du nombre de fenêtres de temps disponibles en apprentissage.

f) Apprentissage

Avant l'apprentissage du classifieur, une procédure de validation croisée, déjà mentionnée, a été appliquée aux seules données d'apprentissage. Pour n SLT utilisés en apprentissage, la validation croisée opérait sur $2n$ groupes de fenêtres de temps, chacun représentant un demi-SLT (première ou deuxième moitié d'un SLT).

Lors de chaque itération de cette validation croisée « imbriquée », 9×5 classifieurs étaient appris, un pour chacune des 9 seuils (ou constructions de classes) et chacune des 5 valeurs suivantes du paramètre c du SVM : 10 puissances -2, -1, 0, 1 et 2. Ce dernier paramètre a été optimisé car les

données d'apprentissage n'étaient a priori pas linéairement séparables suite à la réduction du nombre de variables.

Chaque classifieur appris était évalué, toujours à l'intérieur de la validation croisée, par le critère suivant :

$$\frac{1}{2} \left(\frac{VP}{FN + VP} + \frac{VN}{VN + FP} \right)$$

avec VP le nombre de vrais positifs, VN le nombre de vrais négatifs, FP le nombre de faux positifs (prédits comme positifs mais en réalité négatifs) et FN le nombre de faux négatifs (Fig. 34).

		\sum_i		Réponses attendues	
		y_i	+1	-1	
Prédiction	+1	VP	FP		
	-1	FN	VN		

Fig. 34 – Matrice de confusion ; illustration de la notion de vrais/faux positifs/négatifs. y_i est l'étiquette de classe correspondant à la fenêtre quantifiée x_i . Elle vaut +1 pour la classe des « bonnes performances » (la région inférieure au seuil) et -1 pour la classe des « mauvaises performances » (la région supérieure au seuil).

Ce critère à maximiser est une moyenne de la spécificité $VN/(VN+FP)$ et de la sensibilité $VP/(FN+VP)$. Il est égal au taux de bonnes classifications lorsque les deux classes (+1 et -1) ont le même effectif, mais rend égales les parts de bonnes classifications des deux classes. Il pénalise tout particulièrement le cas où le classifieur aurait attribué toutes les fenêtres de validation à l'une des deux classes. En effet, le critère attribuerait à un tel classifieur la valeur correspondant au niveau de la chance (50 %).

Le classifieur qui maximisait ce critère sensibilité-spécificité était sélectionné ce qui avait pour conséquence de sélectionner un seuil et donc une construction de classes appliquée ensuite aux données de test, et une valeur pour le paramètre c .

Une fois ces deux paramètres (seuil et c) sélectionnés, un nouveau classifieur était appris sur la totalité des données d'apprentissage puis testé sur le SLT de test, tout ceci sur la base de la construction de classes associée au seuil sélectionné.

g) Évaluation de la classification

Une double évaluation du produit de la classification a été menée.

D'un côté, les taux de bonnes classification et les valeurs de sensibilité-spécificité ont été calculés sur les données de test et leur moyennes et intervalles de confiance associés ont été rapportés. Ce sont des

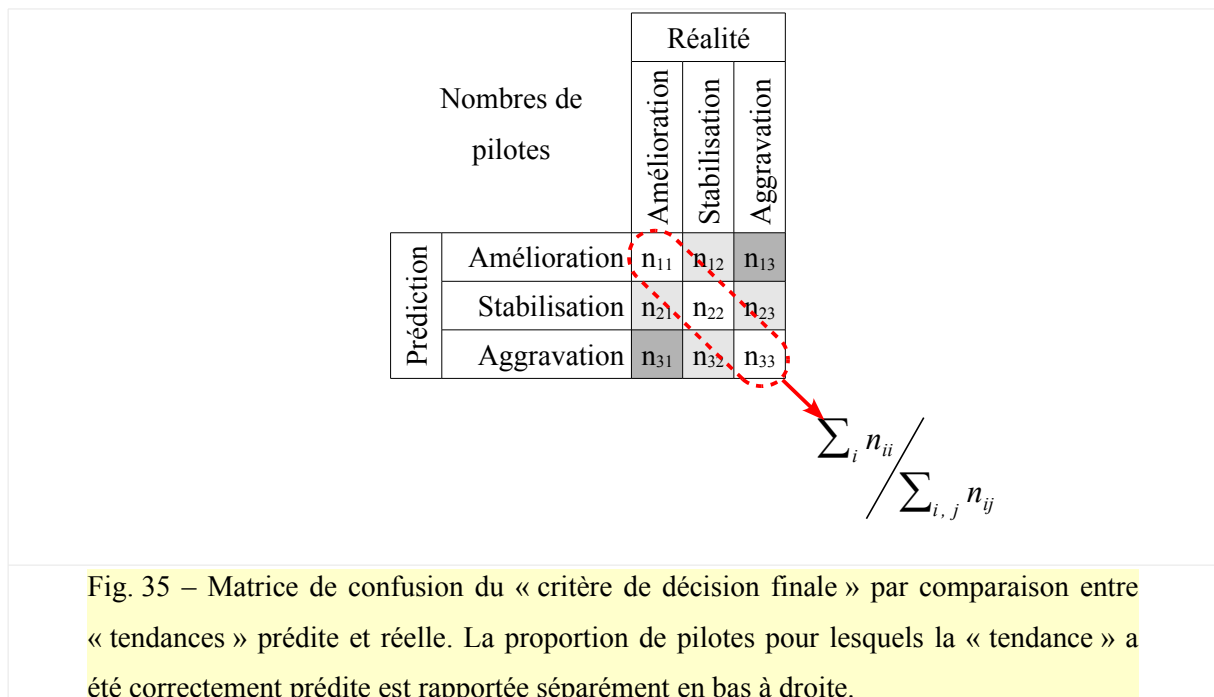
mesures « classiques » de comparaison des prédictions du classifieur avec les valeurs réelles correspondantes.

D'autre part, un « critère de décision finale » consistait à correctement prédire la « tendance » de l'état de « fatigue » du pilote comme schématisé figure 31. Le but est de prendre une seule décision en fin de SLT quant à l'évolution globale de la proportion de chaque classe.

Pour cela, la proportion de la classe des « bonnes performances » dans les prédictions du classifieur était comparée aux proportions réelles de cette même classe dans les SLT d'apprentissage. Trois cas (ou tendances) étaient considéré(e)s : la proportion mesurée sur les données de test pouvait être soit supérieure au maximum mesuré sur les données d'apprentissage (« amélioration » de l'état du pilote), soit inférieure au minimum (« aggravation » de l'état de « fatigue » du pilote), soit entre les deux (« stabilisation »).

Ensuite, la même procédure était appliquée au taux réel sur le SLT de test et enfin, tout comme pour la prédiction de la classe d'une fenêtre de temps, la tendance prédite était comparée à la tendance réelle. La compatibilité des tendances prédite et réelle est appelée ici « critère de décision finale ». Il peut y avoir égalité, incompatibilité mineure (par exemple une « amélioration » prise pour une « stabilisation ») ou incompatibilité majeure (« amélioration » plutôt que « aggravation » et réciproquement).

Les matrices de confusion 3x3 présentées dans la section Résultats ont la structure suivante :



8.3. Résultats

E et \bar{E}_{10} présentent une même évolution au cours de la simulation à l'exception du SLT 5 sur lequel le E décroît tandis que le \bar{E}_{10} augmente (Fig. 36).

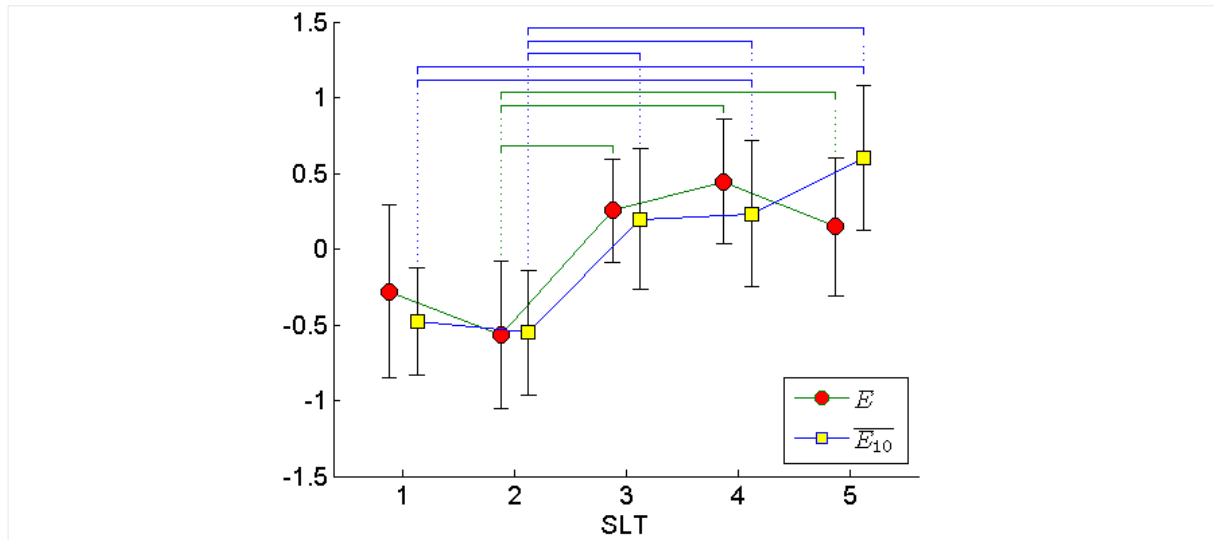


Fig. 36 – Performance des pilotes sur les SLT au cours de la simulation de vol (E et \bar{E}_{10}). Les barres verticales représentent les intervalles de confiance tandis que les barres horizontales marquent la présence d'une différence significative entre paire de SLT.

Le décours temporel moyen de l'indice comportemental utilisé pour la quantification de la « performance » à des fins de classification (le E_{10}) est maximal en début de SLT. Sur les deux premiers SLT, il décroît puis se stabilise (Fig. 37). Sur les SLT suivants, il décroît sur une plus longue période de temps et reste plus irrégulier par la suite.

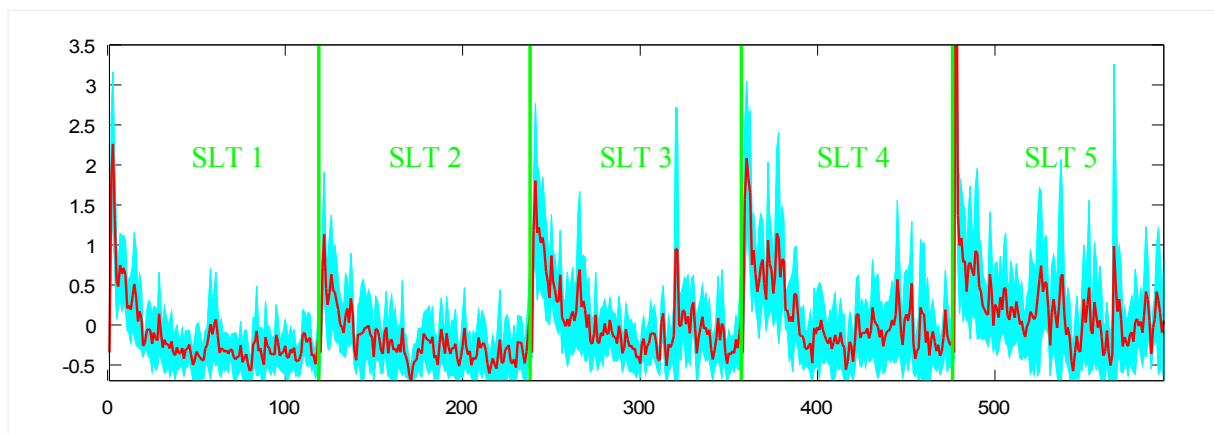


Fig. 37 – Décours temporel du E_{10} . Les abscisses représentent les fenêtres de 10 s. Les barres verticales vertes délimitent les SLT, la courbe rouge est la moyenne à travers les pilotes et la région cyan est l'intervalle de confiance.

Les résultats de classification sont donnés tableaux 38 et 39.

Tableau 38, on remarque une certaine constance des taux de bonnes classifications, c'est-à-dire une certaine robustesse des résultats, malgré le délai entre la fin (dans le temps) des données d'apprentissage et le début des données de test.

La sensibilité-spécificité est significativement au dessus du niveau de la chance (50 %) pour les configurations 5, 6 et 9. Ses valeurs sont toujours inférieures aux taux de bonnes classifications correspondants.

Quelque soit la configuration, on trouve plus de pilotes pour lesquels la partie basse du critère comportemental (les « bonnes performances ») est dominante sur les données d'apprentissage, en fréquence de fenêtres de temps et donc en durée totale, malgré une variabilité importante.

SLT				
1	2	3	4	5
	0,63 ± 0,20	74,6 ± 7,9 %	72,1 ± 8,5 %	73,2 ± 9,2 %
		53,6 ± 4,1 %	51,7 ± 2,2 %	53,4 ± 4,3 %
	0,77 ± 0,14		70,6 ± 5,9 %	69,1 ± 5,0 %
			52,4 ± 3,2 %	55,7 ± 4,4 %
	0,75 ± 0,16			72,8 ± 3,7 %
				57,5 ± 5,7 %
	0,67 ± 0,17		68,1 ± 9,1 %	69,8 ± 7,7 %
			52,2 ± 3,1 %	52,3 ± 4,4 %
	0,74 ± 0,15			71,4 ± 6,3 %
				56,4 ± 5,6 %
		0,65 ± 0,17		70,9 ± 8,5 %
				54,3 ± 5,4 %

Fig. 38 – Résultats de la classification. Chaque colonne représente un SLT. Les SLT en grisé ont servi à l'apprentissage. Les SLT suivants (à droite) ont été un-à-un utilisés en test. Dans la région grisée est rapporté le seuil sélectionné moyen (proportion de la classe des « bonnes performances » sur les données d'apprentissage) avec son intervalle de confiance. Pour chaque SLT de test, deux informations sont données : en haut, le taux moyen de bonnes classifications ; en bas, la moyenne sensibilité-spécificité.

Le taux de pilotes pour lesquels la prédiction de la « tendance » était correcte est maximal (13/13) pour les configurations 2, 4, 8 et 10 (tableau 39).

Dans les prédictions de « tendance », aucune erreur « majeure » n'a été relevée.

Les matrices de confusion permettent de constater qu'il y a, en considérant la population des pilotes, plus souvent « aggravation », par rapport aux périodes exploitées pour l'apprentissage, qu'« amélioration » ou « stabilisation ».

SLT																															
1	2	3	4	5																											
		<table border="1"> <tr><td>4</td><td></td><td></td></tr> <tr><td></td><td></td><td></td></tr> <tr><td></td><td>1</td><td>8</td></tr> </table> <p>12/13</p>	4							1	8	<table border="1"> <tr><td>4</td><td></td><td></td></tr> <tr><td></td><td></td><td></td></tr> <tr><td></td><td></td><td>9</td></tr> </table> <p>13/13</p>	4								9	<table border="1"> <tr><td>4</td><td></td><td></td></tr> <tr><td></td><td></td><td></td></tr> <tr><td></td><td>1</td><td>8</td></tr> </table> <p>12/13</p>	4							1	8
4																															
	1	8																													
4																															
		9																													
4																															
	1	8																													
			<table border="1"> <tr><td>3</td><td></td><td></td></tr> <tr><td></td><td></td><td></td></tr> <tr><td></td><td></td><td>10</td></tr> </table> <p>13/13</p>	3								10	<table border="1"> <tr><td>2</td><td></td><td></td></tr> <tr><td>1</td><td></td><td></td></tr> <tr><td></td><td></td><td>10</td></tr> </table> <p>12/13</p>	2			1					10									
3																															
		10																													
2																															
1																															
		10																													
				<table border="1"> <tr><td>2</td><td></td><td></td></tr> <tr><td></td><td></td><td></td></tr> <tr><td></td><td>1</td><td>10</td></tr> </table> <p>12/13</p>	2							1	10																		
2																															
	1	10																													
			<table border="1"> <tr><td>2</td><td></td><td></td></tr> <tr><td>2</td><td></td><td></td></tr> <tr><td></td><td></td><td>9</td></tr> </table> <p>11/13</p>	2			2					9	<table border="1"> <tr><td>4</td><td></td><td></td></tr> <tr><td></td><td></td><td></td></tr> <tr><td></td><td></td><td>9</td></tr> </table> <p>13/13</p>	4								9									
2																															
2																															
		9																													
4																															
		9																													
				<table border="1"> <tr><td>2</td><td>1</td><td></td></tr> <tr><td></td><td></td><td></td></tr> <tr><td></td><td></td><td>10</td></tr> </table> <p>12/13</p>	2	1							10																		
2	1																														
		10																													
				<table border="1"> <tr><td>3</td><td></td><td></td></tr> <tr><td></td><td></td><td></td></tr> <tr><td></td><td></td><td>10</td></tr> </table> <p>13/13</p>	3								10																		
3																															
		10																													

Fig. 39 – Critère de décision finale. La structure du tableau est parente de celle figure 38. Les résultats pour chaque SLT de test sont rapportées par la matrice de confusion des sujets pour le « critère de décision finale » et le taux de sujets pour lesquels la prédiction de la « tendance » était correcte (cf figure 35). Les cases vides des matrices de confusion correspondent à des effectifs nuls.

Pour rappel, les résultats présentés ont été obtenus par comptage des « bonnes performances » plutôt que des « mauvaises performances ». Il est possible d'appliquer une même procédure aux taux de

« mauvaises performances », mais les résultats apparaissaient alors beaucoup moins intéressants ($5,6 \pm 0,9$ « tendances » correctement prédites contre $12,3 \pm 0,4$ dans le cas présenté).

8.4. Discussion

a) Classification

Tout d'abord, si l'on reprend l'objectif factice de la prédiction d'une tendance du E_{10} entre le début, c'est-à-dire les SLT utilisés pour l'apprentissage du classifieur, et la fin, c'est-à-dire le SLT de test, on constate qu'il y a une grande majorité de succès. En effet, le critère de décision finale est atteint pour 11 pilotes sur les 13 dans le cas le moins favorable. Dans certaines configurations il n'y a même eu aucune erreur.

Ces résultats intéressants associés à des taux de bonnes classifications moins élevés s'expliquent en établissant un parallèle avec la notion de taux de transfert d'information (voir section 4.3.a). Pour augmenter la probabilité d'une bonne décision, plutôt que ne considérer qu'une seule réponse du classifieur, on peut capitaliser des réponses successives et trancher en faveur de la réponse majoritaire, entre autres stratégies possibles. Tout dépend du délai que le décideur peut se permettre pour prendre sa décision. Ici, il a été proposé de prendre une décision au terme du SLT de test.

On remarque en revanche qu'il n'y a pas de « proportionnalité » évidente entre les valeurs moyennes des taux de bonnes classifications ou des moyennes sensibilité-spécificité et la proportion de tendances correctement prédites. Cela est dû au fait que le critère de décision finale n'implique pas qu'une synthèse des réponses du classifieur mais aussi un partitionnement en trois modalités lié à des informations supplémentaires.

Considérons la configuration la plus semblable au scénario présenté figures 30 et 31. Il s'agit de la cinquième, dans laquelle les trois premiers SLT ont formé l'ensemble d'apprentissage d'un classifieur testé ensuite sur le dernier SLT. L'intervalle théorique de 44 min entre la fin des données d'apprentissage et le début des données de test n'a pas réellement mis le classifieur en difficulté puisqu'au final, pour 12 sujets sur les 13, la tendance a été correctement prédite.

Ces résultats ont été obtenus par l'emploi d'une procédure pour le moins élaborée et il serait souhaitable à présent de comparer quantitativement cette dernière aux procédures plus conventionnelles, en particulier par l'usage de constructions « fixes » de classes équilibrées puis déséquilibrées. C'est la perspective à plus court terme de ce travail inachevé.

On peut s'interroger sur la problématique des classes déséquilibrées car on observe dans cette étude que la moyenne sensibilité-spécificité affiche des valeurs très faibles car proches du niveau de la chance. De toute évidence, c'est un critère très sévère. On pourrait en tester d'autres. L'usage de procédures imbriquées de validation croisée est fréquent, de même que le choix d'un critère autre que

la maximisation du taux de bonnes classifications, pour l'optimisation de paramètres, de manière à obtenir un taux de bonnes classifications maximal en validation.

La difficulté à prédire la tendance à partir du taux de mauvaises performances (plutôt que de bonnes performances) peut s'expliquer par la précision accrue dans le comptage de grands effectifs. En effet, encore une fois, plus on a de réponses de la part du classifieur, plus une synthèse de ces réponses sera robuste aux quelques erreurs qui ont pu se glisser parmi les réponses. En outre, compte tenu du seuil sélectionné en moyenne élevé, les bonnes performances étaient probablement encore majoritaires même dans le dernier SLT où la moyenne de l'indice était la plus élevée.

Tout cela laisse tout de même penser qu'un critère plus complexe que le simple taux de bonnes performances pourrait permettre une augmentation du taux de tendances correctement prédites, en considérant aussi le taux de mauvaises performances au moins dans certains cas. En effet, pour certains couples configuration-pilote, le seuil sélectionné descendait jusqu'à 0,1 donc les mauvaises performances étaient définies de manière à être dominantes en effectif, à l'inverse de la majorité pour qui le seuil sélectionné était 0,9. Et en effet, on remarque que les seuls sujets pour lesquels la tendance a été mal prédite présentaient en réalité soit une amélioration, soit une stabilisation de leur état.

b) Quantification du comportement

La question du sens et de la pertinence des divers indices comportementaux utilisés ici est ouverte. Dans la littérature, l'écart-type des paramètres de vol est fréquemment considéré comme une mesure de performance du pilote (Hubbard, 1987 ; Bisson *et al.*, 1993 ; Morris & Miller, 1996), bien que cette mesure devrait idéalement être utilisée de pair avec d'autres (Hubbard, 1987 ; Johnson & Rantanen, 2005). Les mesures de performance sont développées plus souvent dans le but d'évaluer le niveau des pilotes que de suivre en temps réel leur performance « instantanée ». Morris et Miller ont étudié une telle variable sur la durée des SLT pour en observer les fluctuations au cours d'une simulation. C'est donc l'échelle de temps de référence pour cette étude, d'où l'effort de comparaison entre le E et le E_{10} , c'est-à-dire entre les échelles de temps du segment de vol et de la fenêtre de 10 s.

Le E et le \bar{E}_{10} ont des dynamiques semblables à l'exception donc du dernier SLT. Il semblerait donc que les deux différences majeures entre le E et le E_{10} ne prêtent pas tant à conséquences que ce que l'on aurait pu craindre. Une différence était, au niveau de chaque paramètre de vol, l'absence de centrage et l'usage de la tolérance comme facteur réducteur en remplacement de l'écart-type à travers les SLT. Cette différence délibérée était initialement motivée par deux remarques. D'une part, nous avons déjà remarqué lors de nos premières analyses que la différence n'était pas considérable et donc que les valeurs de tolérance imposées par les moniteurs pouvaient être plus ou moins interprétées comme des facteurs de conversion entre les unités dans lesquelles sont exprimées les paramètres de vol et une unité de « difficulté », en quelque sorte. En outre, l'opération de centrage-réduction utilisée

pour le calcul du E présente ce qui pourrait être considéré comme un inconvénient : les valeurs calculées pour les différents SLT sont interdépendantes si bien que, si nous avons par exemple ajouté un sixième SLT, la valeur du E sur le premier SLT serait différente. Le E_{10} et le \bar{E}_{10} en revanche ne présentent pas cette relativité.

Notre conclusion est que la ressemblance de dynamique entre E et le \bar{E}_{10} apporte un support à la fois au E_{10} et au E .

c) Dynamique du comportement

Si l'on considère les motifs du E et du \bar{E}_{10} , la performance tend à décroître à partir du troisième SLT. Pour le E , les différences significatives concernent en revanche le deuxième SLT avec chacun des trois suivants. Bien que qualitativement comparables, les différences entre SLT du \bar{E}_{10} étaient plus évidentes que celles du E car aucune des différences n'étaient significatives pour le E avec un test bilatéral plutôt qu'unilatéral. Le premier SLT présente des valeurs d'indices plus élevées (performances moins bonnes) que sur le deuxième SLT, ce qui pourrait être attribuable à un phénomène d'apprentissage peu surprenant en soi lorsqu'on se rappelle que les pilotes utilisaient le simulateur pour la première fois.

Si l'on considère à présent le E_{10} , en particulier la figure 37, trois états ou phases sont possibles : variabilité décroissante, variabilité stable et faible, et variabilité irrégulière. L'irrégularité marquée dans les trois derniers SLT peut faire directement suite à la décroissance systématique en début de SLT.

La décroissance en début de SLT peut être attribuée soit à une longue stabilisation (2 min ou plus) suite à la dernière manœuvre du FMT qui précède, soit à un changement comportemental d'apaisement – les pilotes se feraient plus calmes, soit à l'effet bénéfique du repos que représente le SLT par rapport au FMT précédent.

L'irrégularité de la variabilité dans la tenue des paramètres de vol, moins bien délimitée, pourrait être due soit à une fatigue liée à la monotonie, soit à une tolérance faiblissante à l'inconfort dont tous les pilotes se sont plaints à l'issue de la simulation ; ils déclaraient avoir mal aux fesses à cause du siège (type MK4).

d) Perspectives

De cette discussion, il ressort qu'un volume important d'hypothèses restent à tester. Certaines ont déjà été évoquées, mais bien d'autres restent à formuler.

Il est par exemple possible de se lancer dans la recherche d'un meilleur indice comportemental, encore faudrait-il avoir un critère de comparaison entre indices. Une possibilité mise de côté est offerte par certaines des mesures effectuées de fatigue subjective, passées sous silence jusque là. Une autre

possibilité est d'exploiter le cadre d'étude offert par la procédure de classification-validation. Ce cadre permet de considérer avec plus de perspectives qu'auparavant la possibilité d'employer un critère comportemental différent d'un pilote à un autre, voire, puisque l'intérêt d'un tel critère serait la définition d'états mentaux, de considérer des régions prédéfinies dans l'espace engendré par plusieurs variables comportementales. Enfin, le simple critère de scores élevés en classification est en soi exploitable puisqu'il quantifie plus ou moins en quoi les niveaux de la variable comportementale sont sous-tendus par des motifs dans les quantifications électrophysiologiques, autrement dit il exprime quelle « réalité » prennent les classes dans l'espace des quantifications.

La procédure de classification peut également faire l'objet de nombreuses optimisations. L'idée d'une construction adaptative des classes est nouvelle, cette étude étant la première à proposer une procédure du genre. On pourrait par exemple sélectionner une taille de fenêtres parmi un ensemble de durées prédéfinies de la même manière que pour le seuil entre classes.

On pourrait également pondérer les erreurs dans l'apprentissage du SVM en suivant une autre stratégie, par exemple en faisant intervenir le temps écoulé à l'intérieur de chaque SLT de manière à pénaliser les mauvaises performances au début ou à la fin de chaque SLT, selon l'interprétation que l'on donne aux unes et aux autres.

Enfin, d'autres quantifications des signaux électrophysiologiques pourraient être intégrées. Ce point, commun à toutes les études, est laissé au chapitre de conclusion.

9 – Conclusion et perspectives

9.1. Retour sur les principaux résultats

Trois études ont été menées et décrites dans cette thèse, chacune attachée à une expérience différente. La première étude portait sur la problématique de la classification de types de quantifications électrophysiologiques divers. Les données analysées étaient celles d'une tâche dite de *task switching*. La seconde étude présentée traitait de la reconstruction des sources corticales de l'activité EEG appliquée à des données issues d'une tâche de compatibilité spatiale. Enfin, la troisième expérience était une simulation de vol pour laquelle l'évolution de l'état du pilote était prédite en fin de simulation par comparaison avec l'état du même pilote lors de périodes antérieures.

Le tableau de la figure 40 récapitule les principaux résultats obtenus.

Par ailleurs, les rôles avaient été partagés ainsi : la problématique de la fatigue mentale était traduite dans la construction des classes, la physiologie l'était dans les quantifications, et l'analyse des données, c'est-à-dire l'instrument optique au travers duquel les résultats ont été présentés, l'était dans le choix des techniques de classification et des procédures de validation.

9.2. Quelles données physiologiques ?

La première étude a permis l'observation de faits compatibles avec des réalités physiologiques connues. Tout d'abord, l'EEG est la technique la mieux adaptée pour l'étude d'états mentaux, mais l'ECG et l'EOG apportent chacune une information alors que l'on aurait pu croire que non. Ensuite, les variations des variables d'EOG sont cohérentes entre elles, du moins au regard de la littérature qui décrit une augmentation de la fréquence et de la durée des clignements associées à une diminution d'amplitude, manifestations attribuées à une levée de l'inhibition de l'activité palpébrale.

Le paradigme de *task switching* a également provoqué des changements assez globaux sur l'EEG, avec une augmentation en amplitude moyenne dans la bande 7-13 Hz et une diminution d'activité dans la bande 13-18 Hz avec la fatigue. Si l'on compare avec les observations faites pour la tâche de compatibilité spatiale de la deuxième expérience, observations décrites dans Besserve *et al.* (2008), on retrouve une diminution avec la baisse de performance en 13-18 Hz localisée en pariétal.

En revanche, pour la bande 7-13 Hz, on trouve plutôt une légère diminution, à l'exception des régions périphériques postérieures. La baisse de vigilance que l'on suppose être marquée par l'augmentation en 7-13 Hz ne serait donc pas commune aux deux situations de fatigue, à moins qu'il n'y ait pas eu fatigue dans l'expérience de compatibilité spatiale entre les situations de bonnes et de mauvaises performances.

Expérience Conception	<i>Task switching</i>	Compatibilité spatiale	Simulation de vol
Construction des classes	<ul style="list-style-type: none"> • Approche « bloc a vs bloc b » • Blocs différenciés par le taux d'erreurs, le temps passé à la tâche et la fatigue subjective • Fenêtres de temps de 4 à 30 s. 	<ul style="list-style-type: none"> • Prédiction de niveaux de performance • Performance = temps de réponse • Fenêtres de temps de 20 s 	<ul style="list-style-type: none"> • Prédiction de niveaux de performance • Performance = variabilité des paramètres de vol • Fenêtres de temps de 10 s
Quantifications	<ul style="list-style-type: none"> • Amplitudes moyennes (EEG) • <i>Phase locking values</i> (EEG) • Diverses quantifications de l'ECG et de l'EOG 	<ul style="list-style-type: none"> • Amplitudes moyennes • PLV • Cohérences 	<ul style="list-style-type: none"> • Amplitudes moyennes • PLV
Analyse de données	<ul style="list-style-type: none"> • Validation croisée • Interprétation des poids SVM 	<ul style="list-style-type: none"> • Validation croisée • Exploitation des poids SVM pour la localisation des activités simulées 	<ul style="list-style-type: none"> • Validation par configurations choisies • Sélection de paramètres (seuil entre classes, c) • Équilibrage des classes • Méta-classification
Comportement et fatigue subjective	<ul style="list-style-type: none"> • γ taux d'erreurs • γ fatigue subjective 	<ul style="list-style-type: none"> • Effet de la compatibilité stimulus-réponse et de la relation avec le stimulus précédent 	<ul style="list-style-type: none"> • E et E_{10} rapportent une même information à différentes échelles temporelles
Électrophysiologie	<ul style="list-style-type: none"> • EOG : γ de l'inhibition des clignements • ECG : γ HR et γ HRV • EEG : γ en 7-13 Hz, γ en 13-18 Hz • ERN (γ en 3-7 Hz en FCz) 	<ul style="list-style-type: none"> • (Besserve et al., 2008) : γ en 3-7 Hz, γ en 13-18 Hz en pariétal 	
Classification	<ul style="list-style-type: none"> • EEG meilleure que EOG et ECG • EOG et ECG > niveau de la chance • EEG+EOG+ECG > EEG • Taux de bonnes classifications > 90 % 	<ul style="list-style-type: none"> • Apport de la reconstruction de sources en simulation mais pas sur les vraies données • Vraies données : taux de bonnes classifications \approx 70 % 	<ul style="list-style-type: none"> • Prédiction de tendance correcte pour au moins 11 pilotes sur 13 et jusqu'à 13 pilotes sur 13 • Taux de bonnes classifications \approx 70%
Résultats Expérience	<i>Task switching</i>	Compatibilité spatiale	Simulation de vol

Fig. 40 – Tableau récapitulatif des spécificités expérimentales et principaux résultats.

Les synchronies locales de la bande 3-7 Hz présentent également des différences notables entre les deux études. Dans l'expérience de *task switching*, un motif en FCz évoque une activité de type *error related negativity* (ERN ; van Veen & Carter, 2002) en lien avec l'augmentation du taux d'erreurs entre les deux classes, tandis que dans l'expérience de compatibilité spatiale, une augmentation d'activité avec la baisse de performance irradie la totalité du scalp, évoquant plutôt une levée d'inhibition des réseaux de la mémoire à long terme, inhibition qui aurait pu favoriser le traitement des stimuli visuels. Ce dernier phénomène, propre à l'expérience de compatibilité spatiale, évoque donc le mécanisme de levée d'inhibition que l'on postule dans l'expérience de *task switching* au vu des manifestations observées dans l'activité palpébrale. Ainsi, la levée d'inhibition pourrait être un dénominateur commun aux situations de fatigue mentale.

On peut discuter en revanche la méthode de cette comparaison, les mesures étant différentes d'un travail à l'autre. En effet, alors que les résultats décrits dans Besserve *et al.* (2008) reposent sur des mesures univariées, l'information descriptive pour l'expérience de *task switching* s'appuie sur une technique multivariée par laquelle la contribution des synchronies locales était considérée en même temps que celle des synchronies à distance et des quantifications de l'EOG et de l'ECG.

En ce qui concerne la reconstruction de sources, cette technique n'a pas permis de gain en classification de niveaux de performance. Son intérêt réside plus ici dans la localisation de motifs directement sur le cortex cérébral. De manière générale, la reconstruction de sources pourrait être fort utile dans des situations telles que la classification d'états mentaux qui ne se différencieraient que par des modulations de l'activité du cortex cérébral, au vu des résultats intéressants en simulation.

À présent, en ce qui concerne les perspectives de ces études dans le cadre de l'étude de la fatigue mentale, l'horizon le plus vaste est celui des quantifications alternatives. En effet, pour l'EEG, on pourrait exploiter d'autres quantifications de synchronie locale comme des amplitudes moyennes relatives ou des rapports d'amplitudes moyennes. De même, le champ des mesures de couplages dynamiques entre sites distants est très vaste. Une notion de plus en plus mise en avant dans la littérature au sujet de ces quantifications est la directionnalité du couplage : de combien et de quelle manière l'activité de tel site pilote celle de l'autre ? Il y a ensuite les quantifications de structure du graphe des liens dynamiques. Ce sont des mesures pour le moins dérivées, qui nécessitent la construction d'un graphe ou réseau fonctionnel à l'aide de mesures de couplage par exemple, mais qui semblent être porteuses d'information sur l'état cérébral.

Une autre perspective est ouverte par la reconstruction de sources en intégrant d'autres types de sources dans le modèle direct et en comparant les différentes techniques de résolution du problème inverse. Cela permettrait probablement un apport pour l'étude et la classification d'états de fatigue mentale, mais aussi pour tout le champ des interfaces cerveau-machine dans lesquelles l'activité des

muscles péricraniaux et des yeux, en plus d'être présente et donc gênante en l'état actuel des objectifs et des techniques, pourrait être ainsi soit mieux éliminée, soit mieux exploitée.

9.3. La fatigue mentale comme objet d'étude

Cette thèse a été le terrain d'un questionnaire sur la manière de représenter la fatigue mentale, dans la quête de variables d'étude qui présenteraient des fluctuations semblables à celles de la fatigue. La question a été réduite à la construction d'états (de niveaux) de fatigue mentale. Comment les provoquer ? Comment les repérer ?

L'expérience de *task switching* a fait l'objet d'une approche « bloc *a* versus bloc *b* ». Ces blocs ont été sélectionnés de manière à ce qu'ils s'opposent l'un à l'autre en matière de taux d'erreurs moyen. Cette approche est probablement très parente de l'approche « début *versus* fin » exploitée par quelques auteurs (Trejo *et al.*, 2005 ; ten Caat *et al.*, 2008), car la tâche avait été choisie pour que le temps passé à la tâche induise une fatigue mentale et donc éventuellement une baisse de performance.

La fatigue était en effet provoquée par le temps passé à la tâche, mais le temps passé à la tâche ne fournissait pas pour autant un bon critère pour la construction des états de fatigue mentale. C'est là une critique qui a été amplement développée dans la discussion section 6.6.b) . La crainte formulée concerne le risque probablement élevé qu'un phénomène puisse superposer ses effets « par chance » à ceux de la fatigue sans être lui-même en lien avec la fatigue.

Dans une approche « début *versus* fin », on pourrait séparer les effets en deux composantes, selon qu'ils sont reproductibles ou non sous la condition de fatigue mentale, en réalisant de nombreuses sessions séparées par des périodes de récupération totale, par exemple en différents jours. L'objectif serait de ne pas se contenter d'observer des variations dans les variables d'étude entre le début et la fin de chaque session, mais d'extraire la part de ces variations qui serait corrélée avec les variations de la fatigue subjective ou de toute autre mesure première de la fatigue mentale. On pourrait ainsi cibler la cause de la part non reproductible des variations entre début et fin, de l'étudier et d'en rejeter la contribution ou de concevoir des protocoles expérimentaux qui supprimeraient cette (ou ces) cause(s). Cela nécessite en revanche des sujets suffisamment volontaires pour réaliser plusieurs fois une tâche longue et pénible. Il faut donc soit rémunérer les sujets, soit s'orienter vers des situations expérimentales réelles dans lesquelles la réalisation en elle-même de la tâche présente un intérêt pour le sujet. L'expérience de simulateur est un bon exemple, car les moniteurs reconnaissaient l'utilité de confronter les élèves-pilotes à des durées de vol plus longues que ce à quoi ils étaient entraînés. Lorsque la fatigue représente un coût, donner à l'opérateur des repères par rapport à sa propre fatigue, lui permettre de connaître ses limites et les éventuelles contraintes liées aux tâches longues, est un argument relativement bien reçu.

À présent, concernant le comportement des sujets au cours des expériences de *task switching* et de simulateur, nous avons remarqué une discrète similitude dans la durée passée à la tâche au terme de laquelle la dégradation de la performance ralentissait. Dans l'expérience de *task switching*, cette durée se situait dans le temps du quatrième bloc sur lequel le taux d'erreurs et d'autres variables comportementales atteignaient un maximum, soit environ entre 60 et 90 min, un bloc durant souvent plus de 20 min. Dans l'expérience de simulateur, le *E* présentait une augmentation très suivie par l'ensemble des pilotes entre le deuxième et le troisième SLT, soit en théorie entre 54 min (la fin du deuxième SLT en excluant le décollage) et 81 min (la fin du troisième SLT), puis variait moins jusqu'à la fin de la simulation.

Ce phénomène ne semble pas dû à des rythmes ultradiens tels que le BRAC, *Basic Rest Activity Cycle*, dont la période varie entre 90 et 150 min (Guerrien & Nansy-Danney, 2003). Notre hypothèse serait plutôt que la fatigue serait à l'origine de cette caractéristique de la dynamique de la performance au cours d'une tâche, ce qui pourrait valider l'approche de prédiction de niveaux de performance. Cela n'est pas trivial, puisqu'il semblerait que la sensation subjective de fatigue ne soit pas toujours corrélée à des critères comportementaux pris seuls.

L'approche de prédiction de niveaux de performance a été exploitée sur les expériences de compatibilité spatiale et de simulation de vol afin d'obtenir des classes moins ségréguées dans le temps. Des bonnes performances alternaient avec des mauvaises performances. Cependant, nous avons observé dans l'expérience de compatibilité spatiale que les taux de bonnes classifications les plus élevés correspondaient aux situations où les bonnes performances se concentraient plutôt au début de la période d'enregistrement et les mauvaises performances plutôt à la fin. Ajoutée au fait que les taux de bonnes classifications étaient en moyenne relativement bas (inférieurs à 60 % pour plusieurs sujets), cette observation nous a fait douter de la pertinence de la construction de classes (pour rappel, basée sur le temps de réponse moyen sur 20 s) appliquée à cette expérience. Ce n'est que dans l'expérience de simulation de vol que la construction des classes prend un sens plus tangible grâce à l'intérêt de la démarche méthodologique guidée par des objectifs concrets.

En ce qui concerne cette dernière expérience, nous souhaitons à présent obtenir des réponses à de nombreuses hypothèses, énoncées pour une part dans le chapitre précédent, même s'il est possible que les conclusions à venir ne puissent concerner que le pilotage d'aéronef.

9.4. Le potentiel de l'analyse de données par apprentissage

Tout au long de cette thèse, le classifieur utilisé était invariablement le SVM linéaire. Ce choix n'était pas critique au départ du fait que toutes les techniques de classification se valent, d'une certaine manière. Il s'avère que, face à des problématiques particulières, des adaptations peuvent être faites

comme la pondération des erreurs proposée dans l'étude concernant l'expérience de simulation de vol. Il peut donc être utile de maîtriser une technique quitte à n'utiliser qu'elle.

Par exemple, au risque de me montrer un peu trop insistant sur ce point, le paramètre c du SVM ne nécessite pas d'être « réglé » dans des situations où il n'a pas de raison d'être, par exemple lorsque les données d'apprentissage sont linéairement séparables.

Plus important est l'évaluation du produit de l'apprentissage, du classifieur ou détecteur. Elle est permise par une quantification de la précision du classifieur sur des données différentes de celles qui ont servi à l'apprentissage, et par une procédure de partage des données entre ensemble d'apprentissage et ensemble de test.

Le taux de bonnes classifications en validation (ou test) se justifie pleinement lorsque les classes sont équilibrées comme c'était le cas ou presque dans les études correspondant aux expériences de *task switching* et de compatibilité spatiale. C'est l'indicateur le plus utilisé.

Pour la validation, une procédure de validation croisée a été utilisée sur les expériences de *task switching* et de compatibilité spatiale, et une procédure inspirée d'un scénario d'utilisation d'un détecteur a été proposée et exploitée sur l'expérience de simulation de vol.

Les boîtes à outils pour la classification et autres programmes comparables ont fleuri et leur usage a été simplifié. Ainsi, certains voudraient voir leur utilisation se généraliser tout comme cela a été fait avec succès pour les tests statistiques univariés. Le fil de réflexion tissé au cours des études présentées dans ce document serait plutôt matière à modération au sujet de la pratique « sauvage » de telles techniques, car parallèlement à la diffusion des outils de classification, un usage réfléchi des procédures de validation paraît important.

10 – Références

- Aaronson L., Tell C., Cassmeyer V., Neuberger G., Pallikkathayil L. & Pierce J. Defining and measuring fatigue *Journal of Nursing Scholarship*, **1999**, *31*, 45-50.
- Akerstedt T. Consensus statement: fatigue and accidents in transport operations *Journal of Sleep Research*, **2000**, *9*, 395-395.
- Akerstedt T. Works hours and sleepiness/Horaires de travail et somnolence *Neurophysiologie Clinique/Clinical Neurophysiology*, **1995**, *25*, 367-375.
- Astolfi L., Cincotti F., Mattia D., Marciani M., Baccala L., de Vico Fallani F., Salinari S., Ursino M., Zavaglia M., Ding L., Edgar J., Miller G., He B. & Babiloni F. Comparison of different cortical connectivity estimators for high-resolution EEG recordings *Human Brain Mapping*, **2007**, *28*, 143-157.
- Baillet S., Mosher J. & Leahy R. Electromagnetic brain mapping *IEEE Signal Processing Magazine*, **2001**, *18*, 14-30.
- Bartels, M.N. Fatigue in cardiopulmonary disease *Physical Medicine and Rehabilitation Clinics of North America*, **2009**, *20*, 389-404.
- Benington J., Kodali S. & Craig Heller H. Stimulation of A1 adenosine receptors mimics the electroencephalographic effects of sleep deprivation *Brain Research*, **1995**, *692*, 79-85.
- Benjamini Y. & Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing *Journal of the Royal Statistical Society*, **1995**, *57*, 289-300.
- Besserve M., Philippe M., Florence G., Laurent F., Garnero L. & Martinerie J. Prediction of performance level during a cognitive task from ongoing EEG oscillatory activities *Clinical Neurophysiology*, **2008**, *119*, 897-908.
- Bills A. Fatigue in mental work *Physiological Reviews*, **1937**, *17*, 436-453.
- Bisson R., Neville J.J., Boll P.A., French J., Ercoline W.R., McDaniel R.L. & Storm W.F. Digital flight data as a measure of pilot performance associated with fatigue from continuous operations during the Persian Gulf conflict *AGARD, Nutrition, Metabolic Disorders and Lifestyle of Aircrew*, **1993**.
- Boksem M. & Tops M. Mental fatigue: costs and benefits *Brain Research Reviews*, **2008**, *59*, 125-139.

- Bonnefond A., Rogé J. & Muzet A. Behavioural reactivation and subjective assessment of the state of vigilance -- application to simulated car driving *International Journal of Occupational Safety and Ergonomics*, **2006**, *12*, 221-229.
- van Boxtel A. & Jessurun M. Amplitude and bilateral coherency of facial and jaw-elevator EMG activity as an index of effort during a two-choice serial reaction task *Psychophysiology*, **1993**, *30*, 589-604.
- Brown I. Driver fatigue Driver fatigue *Human Factors*, **1994**, *36*, 298-314.
- ten Caat M., Lorist M., Bezdán E., Roerdink J. & Maurits N. High-density EEG coherence analysis using functional units applied to mental fatigue *Journal of Neuroscience Methods*, **2008**, *171*, 271-278.
- Cabon P., Coblenz A., Mollard R. & Fouillot J. Human vigilance in railway and long-haul flight operation *Ergonomics*, **1993**, *36*, 1019-1033.
- Castell L., Yamamoto T., Phoenix J. & Newsholme E. The role of tryptophan in fatigue in different conditions of stress *Adv. Exp. Med. Biol.*, **1999**, *467*, 697-704.
- Chaudhury, A. & Behan, P.O. Fatigue and basal ganglia *Journal of Neurological Sciences*, **2000**, *179*, 34-42.
- Congedo M., Lotte F. & Lécuyer A. Classification of movement intention by spatially filtered electromagnetic inverse solutions *Physics in Medicine and Biology*, **2006**, *51*, 1971-1989.
- Corbetta M. & Shulman G. Control of goal-directed and stimulus-driven attention in the brain *Nature Reviews Neuroscience*, **2002**, *3*, 201-215.
- Curcio G., Casagrande M. & Bertini M. Sleepiness: evaluating and quantifying methods *International Journal of Psychophysiology*, **2001**, *41*, 251-263.
- De Vico Fallani F., Astolfi L., Cincotti F., Mattia D., Tocci A., Grazia Marciani M., Colosimo A., Salinari S., Gao S., Cichocki A. & Babiloni F. Extracting information from cortical connectivity patterns estimated from high resolution EEG recordings: a theoretical graph approach *Brain Topography*, **2007**, *19*, 125-136.
- Devinsky O., Morrell M. & Vogt B. Contributions of anterior cingulate cortex to behaviour *Brain*, **1995**, *118*, 279-306.
- Fishbain D., Cutler R., Cole B., Lewis J., Smets E., Rosomoff H. & Rosomoff R. Are patients with chronic low back pain or chronic neck pain fatigued? *Pain Medicine*, **2004**, *5*, 187-195.

- Floru R., Cail F. & Elias R. Psychophysiological changes during a VDU repetitive task *Ergonomics*, **1985**, 28, 1455-1468.
- Franaszczuk P., Blinowska K. & Kowalczyk M. The application of parametric multichannel spectral estimates in the study of electrical brain activity *Biological Cybernetics*, **1985**, 51, 239-247.
- Gandevia S. Spinal and supraspinal factors in human muscle fatigue *Physiological Reviews*, **2001**, 81, 1725-1789.
- Huang M., Mosher J. & Leahy R. A sensor-weighted overlapping-sphere head model and exhaustive head model comparison for MEG *Physics in Medicine and Biology*, **1999**, 44, 423-440.
- Hubbard D. Inadequacy of root mean square error as a performance measure *Proceedings of the 4th ISAP, Columbus, OH*, **1987**.
- Hämäläinen M. & Sarvas J. Realistic conductor geometry model of the human head for interpretation of neuromagnetic data *IEEE Transactions on Biomedical Engineering*, **1989**, 36, 165-171.
- Ingre M., Akerstedt T., Peters B., Anund A. & Kecklund G. Subjective sleepiness, simulated driving performance and blink duration: examining individual differences *Journal of Sleep Research*, **2006**, 15, 47-53.
- Jackson C. & Earl L. Prevalence of fatigue among commercial pilots *Occupational Medicine (Oxford, London)*, **2006**, 56, 263-268.
- Jasper H. The ten-twenty electrode system of the International Federation *Electroencephalography and Clinical Neurophysiology*, **1958**, 10, 371-375.
- Johnson N.R. & Rantanen E.M. Objective pilot performance measurement: a literature review and taxonomy of metrics *13th International Symposium on Aviation Psychology, Dayton, OH*, **2005**.
- Kaipio J. & Somersalo E. Antman, S.; Marsden, J. & Sirovich, L. (ed.) Statistical and Computational Inverse Problems *Springer New York*, **2005**, 160.
- Kamoussi B., Amini A. & He B. Classification of motor imagery by means of cortical current density estimation and Von Neumann entropy *Journal of Neural Engineering*, **2007**, 4, 17-25.
- Koelewijn T., van Schie H., Bekkering H., Oostenveld R. & Jensen O. Motor-cortical beta oscillations are modulated by correctness of observed action *NeuroImage*, **2008**, 40, 767-775.
- Kronegg J., Voloshynovskiy S. & Pun T. Analysis of bit-rate definitions for brain-computer interfaces *International Conference on Human-Computer Interaction*, **2005**.
- Lachaux J., Rodriguez E., Martinerie J. & Varela F. Measuring phase synchrony in brain signals *Human Brain Mapping*, **1999**, 8, 194-208.

- Lorist M., Kernell D., Meijman T. & Zijdwind I. Motor fatigue and cognitive task performance in humans *The Journal of Physiology*, **2002**, *545*, 313-319.
- Lorist M., Klein M., Nieuwenhuis S., de Jong R., Mulder G. & Meijman T. Mental fatigue and task control: planning and preparation *Psychophysiology*, **2000**, *37*, 614-625.
- Lotte F., Lécuyer A. & Arnaldi B. FuRIA: a novel feature extraction algorithm for brain-computer interfaces using inverse models and fuzzy region of interest *Proc. of the 3rd IEEE-EMBS international Conference on Neural Engineering*, **2007**, 175-178.
- Lu C.-H. & Proctor R.W. The influence of irrelevant location information on performance: a review of the Simon and spatial Stroop effects *Psychonomic Bulletin & Review*, **1995**, *2*, 174-207.
- Luk W., Zhang Y., White T., Lue F., Wu C., Jiang C.-G. & Zhang L. Adenosine: a mediator of interleukin-1beta-induced hippocampal synaptic inhibition *The Journal of Neuroscience*, **1999**, *19*, 4238-4244.
- Luu P., Tucker D. & Makeig S. Frontal midline theta and the error-related negativity: neurophysiological mechanisms of action regulation *Clinical Neurophysiology*, **2004**, *115*, 1821-1835.
- Marcora S., Staiano W. & Manning V. Mental fatigue impairs physical performance in humans *J. Appl. Physiol.*, **2009**, *106*, 857-864.
- Meeusen R., Watson P., Hasegawa H., Roelands B. & Piacentini M. Central fatigue: the serotonin hypothesis and beyond *Sports Med.*, **2006**, *36*, 881-909.
- Mion G. & Ricouard S. Repos de sécurité : quels enjeux ? *Annales Françaises d'Anesthésie et de Réanimation*, **2007**, *26*, 638-648.
- Morik K., Brockhausen P. & Joachims T. Combining statistical learning with a knowledge-based approach - a case study in intensive care monitoring *Proceedings of ICML*, **1999**.
- Morris T. & Miller J. Electrooculographic and performance indices of fatigue during simulated flight *Biological Psychology*, **1996**, *42*, 343-360.
- Mosher J., Leahy R. & Lewis P. EEG and MEG: forward solutions for inverse methods *IEEE Transactions on Biomedical Engineering*, **1999**, *46*, 245-259.
- Neville K., Bisson R., French J., Boll P. & Storm W. Subjective fatigue of C-141 aircrews during Operation Desert Storm *Human Factors*, **1994**, *36*, 339-349.
- Nickel P. & Nachreiner F. Sensitivity and diagnosticity of the 0.1-Hz component of heart rate variability as an indicator of mental workload *Human Factors*, **2003**, *45*, 575-590.

- Noirhomme Q., Kitney R. & Macq B. Single-trial EEG source reconstruction for brain-computer interface *IEEE Transactions on Biomedical Engineering*, **2008**, *55*, 1592-1601.
- Nozaki S., Tanaka M., Mizuno K., Ataka S., Mizuma H., Tahara T., Sugino T., Shirai T., Eguchi A., Okuyama K., Yoshida K., Kajimoto Y., Kuratsune H., Kajimoto O. & Watanabe Y. Mental and physical fatigue-related biochemical alterations *Nutrition*, **2009**, *25*, 51-57.
- Nykopp T. Statistical modelling issues for the adaptive brain interface *Helsinki University of Technology, Dep. of Electrical and Comm. Eng.*, **2001**.
- Ogilvie R. The process of falling asleep *Sleep Medicine Reviews*, **2001**, *5*, 247-270.
- Ota T., Toyoshima R. & Yamauchi T. Measurements by biphasic changes of the alpha band amplitude as indicators of arousal level *International Journal of Psychophysiology*, **1996**, *24*, 25-37.
- Pearson R. & Byars G. The development and validation of a checklist for measuring subjective fatigue *Air University, School of Aviation Medicine, USAF, Randolph AFB, Texas*, **1956**.
- Porkka-Heiskanen T., Alanko L., Kalinchuk A. & Stenberg D. Adenosine and sleep *Sleep Medicine Reviews*, **2002**, *6*, 321-332.
- Qin L., Ding L. & He B. Motor imagery classification by means of source analysis for brain-computer interface applications *Journal of Neural Engineering*, **2004**, *1*, 135-141.
- Rodriguez E., George N., Lachaux J., Martinerie J., Renault B. & Varela F. Perception's shadow: long-distance synchronization of human brain activity *Nature*, **1999**, *397*, 430-433.
- Rogers R. & Monsell S. Costs of a predictable switch between simple cognitive tasks *Journal of Experimental Psychology: General*, **1995**, *124*, 207-231.
- Salahuddin L., Jaegeol C., Myeong G.J. & Desok K. Ultra short term analysis of heart rate variability for monitoring mental stress in mobile settings *29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 2007*, **2007**, 4656-5659.
- Saporta G. Probabilités, analyse des données et statistique *Editions Technip*, **2006**.
- Shen J., Barbera J. & Shapiro C. Distinguishing sleepiness and fatigue: focus on definition and measurement *Sleep Medicine Reviews*, **2006**, *10*, 63-76.
- Silbernagl S. & Despopoulos A. Atlas de poche de physiologie *Flammarion Médecine*, **1997**.
- Stern J., Boyer D. & Schroeder D. Blink rate: a possible measure of fatigue *Human Factors*, **1994**, *36*, 285-297.
- Task Force N. Heart rate variability *European Heart Journal*, **1996**, *17*, 354-381.

- Trejo L., Kochavi R., Kubitz K., Montgomery L., Rosipal R. & Matthews B. Caldwell, J. & Wesensten, N. (ed.) Measures and models for predicting cognitive fatigue *Proceedings of SPIE, Biomonitoring for Physiological and Cognitive Performance During Military Operations*, **2005**, 5797, 105-115.
- Varela F., Lachaux J., Rodriguez E. & Martinerie J. The brainweb: phase synchronization and large-scale integration *Nature Reviews Neuroscience*, **2001**, 2, 229-239.
- van Veen V. & Carter C.S. The anterior cingulate as a conflict monitor: fMRI and ERP studies *Physiology & Behavior*, **2002**, 77, 477-482.
- Veropoulos K., Campbell C. & Cristianini N. Controlling the sensitivity of support vector machines *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, **1999**.
- Volterra A. & Steinhäuser C. Glial modulation of synaptic transmission in the hippocampus *Glia*, **2004**, 47, 249-257.
- Wallstrom G., Kass R., Miller A., Cohn J. & Fox N. Automatic correction of ocular artifacts in the EEG: a comparison of regression-based and component-based methods *International Journal of Psychophysiology*, **2004**, 53, 105-119.
- Watanabe A., Kato N. & Kato T. Effects of creatine on mental fatigue and cerebral hemoglobin oxygenation *Neurosci. Res.*, **2002**, 42, 279-285.
- Waterink W. & van Boxtel A. Facial and jaw-elevator EMG activity in relation to changes in performance level during a sustained information processing task *Biological Psychology*, **1994**, 37, 183-198.
- Wolpaw J., Birbaumer N., McFarland D., Pfurtscheller G. & Vaughan T. Brain-computer interfaces for communication and control *Clinical Neurophysiology*, **2002**, 113, 767-791.

10.1. Publications et communications relatives à la thèse

- Besserve M., Jerbi K., **Laurent F.**, Baillet S., Martinerie J., Garnero L., Classification methods for ongoing EEG and MEG signals, *Biological Research*, 2007, 40(4): 415-437.
- Besserve M., Philippe M., Florence G., **Laurent F.**, Garnero L., Martinerie J., Prediction of performance level during a cognitive task from ongoing EEG oscillatory activities, *Clinical Neurophysiology*, 2008, 119(4): 897-908.
- Laurent F.**, Besserve M., Florence G. et Martinerie J., Apport de la reconstruction de sources en EEG pour la détection d'états de fatigue mentale, *GRETSI'07*, Troyes, France, 2007.

Laurent F., Besserve M., Garnero L., Philippe M., Florence G., Martinerie J., Source reconstruction and synchrony measurements for revealing functional brain networks and classifying mental states, *International Journal of Bifurcation and Chaos*, sous presse.

Laurent F., Guillard M., Martinerie J., Florence G., Changes in flight parameter-related behavior with increased mental fatigue, *Aerospace Medical Association's 80th Annual Scientific Meeting*, Los Angeles, California, 2009.

Laurent F., Valderrama M., Besserve M., Guillard M., Lachaux J.P., Martinerie J., Florence G., Extending the « brainweb » to the « bodyweb » for monitoring mental fatigue, *Neuroimage*, soumis.

Laurent F., Guillard M., Van Beers P., Martinerie J., Florence G., Prediction of pilot performance level during a flight simulation from ongoing EEG oscillatory activities, en préparation.