Thèse de l'université de lyon

Délivrée par

université claude bernard lyon 1

diplôme de doctorat

Ma spécialité

Décodage des intentions et des représentations motrices chez l'homme : analyse multi-échelle et application aux interfaces cerveau-machine

par

Etienne Combrisson

Thèse soutenue le 09/2016 devant le jury composé de :

Mme	Erika Raté	Université à la Menthe	(Rapporteur)
M.	Jacques Ouille	Université à la Fraise	(Rapporteur)
M.	Henri Zoto	Laboratoire laborieux	(Rapporteur)
M.	Jean File	Indienne	(Directeur)
	etc.		

À Isabelle et Didier, mes deux parents,
qui ont tout donné pour que ceci me soit un jour possible.
Merci

Remerciements

Je voudrais tout d'abord exprimer mes plus profonds remerciements à...
AHÂÂÂH !
Je conclurai en remerciant de tout cœur (l'être aimé).

Montréal, le 15 juillet 2016.

Table des matières

Table des matières vi

Liste des figures viii

Notations 1

I Introduction générale 3

1 Présentation de la thématique 7
- 1.1 Interface Cerveau-Machine : définition et objectifs . . 7
  - 1.1.1 Interactions naturelles avec l'environnement 7
  - 1.1.2 ICM : une communication alternative 7
  - 1.1.3 Composantes d'une ICM 8
  - 1.1.4 Acquisition de l'activité neuronale 8
  - 1.1.5 Neuro-réhabilitation et handicap moteur 10
- 1.2 État de l'art des interfaces cerveau-machine 11
  - 1.2.1 Les différents types d'Interface Cerveau-Machine 11
  - 1.2.2 ICM non-invasives 11
  - 1.2.3 ICM invasives 11
- 1.3 Apprentissage machine : applications aux neurosciences 12
- 1.4 Encodage et décodage moteur : bases physiologiques . 12
- 1.5 Intention, exécution et imagerie motrice 12
- 1.6 Delayed task : protocole expérimental 13

2 Objectifs de la thèse 15
- 2.1 Décodage cérébrale à partir d'activité intracrânienne 15
- 2.2 Exploration et amélioration des features 15
- 2.3 Comparatif des classifieurs 16
- 2.4 Exploration des régions non-motrices 16

3 Méthodologie 17
- 3.1 Extraction des features 17
  - 3.1.1 Pré-requis 17
  - 3.1.2 Puissance spectrale 20
  - 3.1.3 Phase 21
  - 3.1.4 Phase-amplitude coupling 21
- 3.2 Apprentissage supervisé 26
  - 3.2.1 Labellisation et apprentissage 27
  - 3.2.2 *Training, testing* et validation-croisée 27
  - 3.2.3 Classifieurs 29

3.2.4 Évaluation de la performance de décodage 31
3.2.5 Seuil de chance et évaluation statistique de la performance de décodage 31
3.2.6 Du single au multi-features 32
3.2.7 Généralisation temporelle 35
3.3 Configuration pour débuter 37
3.4 Implémentation des méthodes présentées 37

4 Données expérimentales 41
4.1 Données intracrâniennes 41
4.1.1 Acquisition 41
4.1.2 Avantages et limitations des données intracrâniennes . . 41
4.1.3 Prétraitements 42
4.2 Données d'étude 43
4.2.1 Données *Center-out* 43
4.2.2 Autres données 45

5 Ouverture 47

II Étude 1 : niveau de chance et évaluation statistique des résultats de classification par apprentissage supervisé 49

III Étude 2 : encodage de l'intention et de l'exécution motrice 63

IV Étude 3 : décodage des directions de mouvement pendant et avant l'exécution de mouvement de membres supérieurs 95

V Étude 4 : optimisation des paramètres de la bande gamma 99

VI Étude 5 : décodage des émotions 103

Conclusion générale 107

A Annexes 108
A.1 Comparatif de methodes PAC (Tort et al., 2010) 109
A.2 Pipeline standard de classification 110
A.3 Comparatif de classifieurs (Pedregosa et al., 2011) . . . 112
A.4 Exemple de schéma d'implantation 114

Bibliographie 115

Liste des figures

1.1 Schéma d'une Interface Cerveau-Machine (Pfurtscheller et al., 2008) 8
1.2 Méthodes d'acquisition de l'activité cérébrale (Waldert et al., 2009) 10
1.3 Pipeline général d'un Interface Cerveau-Machine 11
1.4 Contrôle d'un bras robotisé 12
1.5 Comparatif 13

2.1 Mécanismes du couplage phase-amplitude 15
2.2 Localisation des aires sensorimotrices 16

3.1 *"One-tailed"* et *"two-tailed"* test 19
3.2 Exemple de représentation temps-fréquence de puissance normalisées z-score (Ossandon et al., 2011) 21
3.3 Densité de probabilité d'une distribution d'amplitudes en fonction de tranches de phases 22
3.4 **(A)** Exemple de cartes temps-fréquence phase locked sur le β, **(B)** Exemple de comodulogramme 25
3.5 Labellisation de données 27
3.6 Exemple d'une cross validation 3-folds 29
3.7 Principe du Linear Discriminant Analysis (Lotte et al., 2007) 29
3.8 Principe du Support Vector Machine (Lotte et al., 2007) . . . 30
3.9 Principe du k-Nearest Neighbor (Weinberger et al., 2005) . . 30
3.10 Entraînement puis test d'un classifieur linéaire 30
3.11 Calcul de l'acuité de décodage 31
3.12 Exemple d'une *Forward feature selection* appliquée sur six features 34
3.13 Exemple d'une *Backward feature elimination* appliquée sur six features 35
3.14 Exemple de décodage temporel (Waldert et al., 2008). Ici, l'auteur décode 4-directions de mouvements de la main dans le temps. A chaque instant, un classifieur est créé, entraîné puis testé à ce même instant. 36
3.15 Exemple de généralisation temporelle (King and Dehaene, 2014) 37

4.1 Détails cliniques des sujets ayant participé à la tâche *Center-out* 43
4.2 Implantation intracrâniale et couverture corticale de six sujets épileptiques ayant passés la tâche *Center-out* 44
4.3 Descriptif de la tâche *Center-out* 45

A.1 Comparatif de methodes PAC (Tort et al., 2010) 109
A.2 Pipeline standard de classification 110
A.3 Comparatif de classifieurs (Pedregosa et al., 2011) 112
A.4 Exemple de schéma d'implantation 114

Notations

Général

ICM	Interface Cerveau-Machine
BCI	Brain Computer Interface

Enregistrements

EEG	Électroencéphalographie
MEG	Magnétoencéphalographie
SUA	Single Unit Activity
MUA	Multi Unit Activity
SEEG	Stéréoélectroencéphalographie
ECoG	Électrocorticographie

Features

PAC	Phase Amplitude Coupling

Classifieurs

LDA	Linear Discriminant Analysis
SVM	Support Vector Machine
RF	Random Forest
KNN	k-Nearest Neighbor
NB	Naive Bayes

Première partie

Introduction générale

BLABLABLABKIBLABLOU INTRO ...

L'objectif de cette thèse a été de ...

Totalité des méthodes explorées durant ma thèse sont présentes dans une toolbox python appelée brainpipe, libre d'accès et de droit.

Présentation de la thématique

1

En 1964, Dr. Grey Walter connecte des électrodes directement dans le cortex moteur d'un patient et lui demande de presser un bouton pour faire avancer un rétro-projecteur. En même temps, il enregistre l'activité neuronale de telle sorte que elle aussi, puisse le faire avancer. Là où l'expérience devient remarquable, c'est que le rétro-projecteur avance avant que le patient ne presse le bouton ! Tout l'appareil musculaire du sujet est court-circuité et le contrôle se fait sans mouvement. Contrôler par la *pensée*, un sujet de science fiction qui devient une réalité.
Cette anecdote décrite par Graimann et al. (2009), permet de placer la naissance des Interface Cerveau-Machine (ICM) dans l'histoire. C'est le point d'entrée qui a ensuite conduit une grande diversité de chercheurs à se passionner pour ce sujet.

1.1 Interface Cerveau-Machine : définition et objectifs

1.1.1 Interactions naturelles avec l'environnement

Pour interagir avec son environnement, l'individu se sert des voies de communications naturelles, c'est-à-dire via son système nerveux et musculaire. Le processus de communication débute par une intention qui active certaines régions dans le cerveau. Il en résulte un signal cérébrale qui est ensuite envoyé par le système nerveux périphérique en directions des muscles. C'est ce processus simplifié qui permet à une personne d'interagir avec ce qui l'entoure.

1.1.2 ICM : une communication alternative

Une ICM (ou BCI en anglais pour *Brain Computer Interface*) est un autre système de communication où les voies naturelles sont cout-circuitées. Au lieu de passer par le système nerveux puis musculaire, le signal cérébral est directement intercepté au niveau du cerveau et va ensuite être transformé en commandes. Une ICM est donc un système permettant de traduire une activité neuronale en commande extérieure. Le terme *traduire* est à prendre au sens linguistique c'est-à-dire que les signaux cérébraux forment un langage, composé de règles, de motifs ou *pattern*, que l'on va essayé de décoder (via un ordinateur) pour les transformer en opérations. D'où le terme "Interface Cerveau-Machine ".

1.1.3 Composantes d'une ICM

Pfurtscheller et al. (2008) et (Graimann et al., 2009) introduisent quatre éléments qui composent une ICM :

1. Enregistrer l'activité directement depuis le cerveau. Cet enregistrement pourra être invasif ou non-invasif (cf. 1.2.1)
2. Générer un retour ou *feedback* pour l'utilisateur
3. L'enregistrement et le *feedback* doivent être en temps réel
4. Enfin, l'interface doit être contrôlable par l'utilisateur, de manière active, via un ensemble d'intentions.

A titre d'exemple et pour illustrer ce dernier point, un utilisateur pourrait par exemple décider de bouger un curseur de souris sur un écran en imaginant des mouvements soit de la main gauche soit de la main droite.

FIGURE 1.1 – *L'activité neuronale est enregistrée (Signal acquisition) puis nettoyée (Preprocessing). Ensuite, on extrait des motifs ou patterns qui caractérisent la commande que souhaite envoyer le sujet (Feature extraction). Enfin, la machine tente de reconnaître ces motifs (Classification) et de les transformer en commande (Application interface) (Pfurtscheller et al., 2008)*

1.1.4 Acquisition de l'activité neuronale

L'acquisition de l'activité neuronale constitue la première étape d'une ICM. Deux paramètres permettent de décrire ces différentes techniques :

1. L'accessibilité : c'est-à-dire si cette technique est plus ou moins facile à mettre en place.
2. La qualité : c'est-à-dire le rapport signal sur bruit (RSB)

Si la technique d'enregistrement n'est pas facilement accessible, cela pourrait limiter le nombre de sujets potentiel pour l'étude. La qualité quant à elle va permettre d'avoir accès à des signaux cérébraux de plus ou moins grande qualité ce qui va influer sur la performance et sur les limitations d'une ICM.
Les différents types d'acquisition peuvent être différenciés par leur degrés de pénétration dans le corps. Les enregistrements dits *invasifs* vont nécessiter une intervention chirurgicale mais donnent accès à des signaux de

grande qualité. Les enregistrements *non-invasifs* sont globalement plus faciles à mettre en place car ne nécessitant aucune chirurgie mais la qualité du signal est moindre car l'activité neuronale est enregistrée en dehors de la boîte crânienne.
Au sein de ces deux catégories, on trouvera un ensemble de méthodes d'acquisition qui se différencient par la taille des populations de neurones qu'elles enregistrent ou par le type de signal.

Enregistrements invasifs

On parlera d'intracrânien pour les enregistrements à l'intérieur de la boîte crânienne puis d'intracortical pour des électrodes implantées dans le cortex (Jerbi et al., 2009, Engel et al., 2005). Étant donné que les techniques ci-dessous nécessitent une intervention chirurgicale,

Single Unit Activity (SUA) et **Multi Unit Activity (MUA)** : ces micro-électrodes déposées directement au contact de neurones, disposent de la plus haute résolution et enregistrent des décharges neuronales. SUA et MUA se distinguent par la taille des populations enregistrées.

Stéréoélectroencéphalographie (SEEG) : macro-électrodes enregistrant des populations plus larges que les micro-électrodes. contrairement à la SUA ou MUA où l'on peut compter le nombre de fois qu'un ou plusieurs neurones déchargent, la SEEG enregistre des potentiels électriques (μV).

Électrocorticographie (ECoG) : grille flexible composée d'une matrice d'électrodes. Cette grille est ensuite déposée à la surface corticale. Parce que cette méthode n'est pas intracortical elle est dite *semi-invasive*.

Enregistrements non-invasifs

Les méthodes *non-invasives* représentent le but ultime de l'implémentation concrète d'une Interface Cerveau-Machine . En effet, ne nécessitant aucune intervention chirurgicale, leur utilisation est très répandue. Actuellement, bien qu'ayant une bonne résolution temporelle permettant ainsi de capturer les phénomènes courts, elles souffrent encore d'un manque de résolution spatiale et d'un rapport signal sur bruit plus faible que les techniques invasives.

Électroencéphalographie (EEG) : c'est la technique la plus utilisée dans le domaine des ICM pour son aspect pratique et portatif. On dispose à la surface de la boîte crânienne un ensemble d'électrodes qui enregistrent l'activité électrique émanant de populations larges de neurones. On trouve même denouveaux systèmes sans fils, prometteurs.

Magnétoencéphalographie (MEG) : autre méthode *non-invasive* mais nettement moins portable que l'EEG, la MEG enregistre les champs magnétiques résultant de l'activité électrique du cerveau. Ces champs magnétiques peuvent être jusqu'à 10 milliards de fois plus faibles que le champs

magnétique terrestre. L'aspect inamovible de la MEG limite son utilisation pour les ICM mais on trouve quand même quelque études ayant testé son utilisation (Mellinger et al., 2007, Waldert et al., 2008).

FIGURE 1.2 – *Méthodes d'acquisition de l'activité cérébrale (Waldert et al., 2009) classées par invasivité. La figure indique également la taille des populations de neurones enregistrés ainsi que la résolution spatiale intimement liée à l'invasivité.*

1.1.5 Neuro-réhabilitation et handicap moteur

Prélever directement l'activité neuronale et donc, *bypasser* les voies naturelles, permet de s'affranchir d'éventuelles limitations physiologiques. C'est pourquoi les ICM représentent un enjeu majeur pour la réhabilitation motrice ou handicap moteur ou encore pour la communication palliative. Les applications concrètes des Interface Cerveau-Machine visent donc les personnes disposant de leurs capacités cognitives mais qui sont privées de facultés motrices.
Par exemple, le *syndrome d'enfermement* ou *locked-in syndrome* qui surviennent majoritairement à la suite d'un accident vasculaire cérébrale. Les personnes touchées par ce syndrome sont pleinement consciente de leur corps, de l'environnement, ils peuvent ressentir les sensations de toucher et de douleur mais n'ont plus de facultés motrices, hormis peut-être, les mouvements de paupières ou des yeux. Un autre exemple est celui de la la sclérose latérale amyotrophique (ou SLA) qui est une dégénérescence des neurones moteur. Progressivement, les personnes atteintes de SLA perdent l'usage des bras, des jambes, de la parole des muscles faciaux et enfin de la déglutition. Toutefois, a l'instar du *locked-in syndrome*, la conscience et les facultés cognitives demeurent intactes.

SLA et *locked-in syndrome* ne sont que deux exemples expliquant l'intérêt social du développement des ICM. Redonner un peu de contrôle ou établir un canal de communication avec les familles des patients expliquent l'engouement qui existe depuis maintenant plus de 40 ans pour les ICM.

1.2 État de l'art des interfaces cerveau-machine

Mettre des ICM plus récentes voir faire un schéma récapitulatif à mettre en annexe. Good idea

(Bekaert et al., 2009)

Figure 1.3 – *Pipeline général d'un Interface Cerveau-Machine*

1.2.1 Les différents types d'Interface Cerveau-Machine

Okokok, ref vers ces différents types

ICM *synchrones* et *asynchrones*

Intro blabla
ICM *synchrones* : description
ICM *asynchrones* : description

ICM *invasives* et *non-invasives*

Blabla intro
ICM *invasives* : description
ICM *non-invasives* : description
La partie qui va suivre introduit les différents types d'enregistrement cérébraux.

1.2.2 ICM non-invasives

P300 speller

1.2.3 ICM invasives

(Hochberg et al., 2012)

FIGURE 1.4 – *Contrôle d'un bras robotisé*

1.3 Apprentissage machine : applications aux neurosciences

okok

1.4 Encodage et décodage moteur : bases physiologiques

Pomper les résultas de rodrigo + besserve + inclure schéma global du cerveau

1.5 Intention, exécution et imagerie motrice

La figure 1.5 représente une trajectoire d'un processus markovien de saut, avec les notations associées. (Hanakawa et al., 2008)

FIGURE 1.5 – *Comparatif*

1.6 DELAYED TASK : PROTOCOLE EXPÉRIMENTAL

okok

Objectifs de la thèse

2

2.1 Décodage cérébrale à partir d'activité intracrânienne

Expliquer que le terme motif utilisé en intro se transforme ici en feature ou attribut. Ce qui nous permet d'effectuer une succession de mouvements identiques c'est que il y a dans l'AN des pattern permettant de reproduire des mouvements. D'ailleurs, c'est grâce à ça que l'on peut se perfectionner dans un sport ou en zik car il y a apprentissage.

- Exemple d'un schéma d'implantation + IRM
- Bipolarisation : débruitage et augmentation de la spécificité (article Karim)
- Extraction de features (ici on pourrait mentionner que le deep learning pourrait marcher sur les données brutes)

2.2 Exploration et amélioration des features

Rôle physiologique du phase-amplitude coupling

(Hyafil et al., 2015)

Figure 2.1 – *Mécanismes du couplage phase-amplitude*

2.3 Comparatif des classifieurs

Expliquer que, chaque classifier possède une méthodologie propre permettant de répondre à des types de données différentes (en fonction des hypothèses de fonctionnement de chacun des classifiers)

2.4 Exploration des régions non-motrices

(Van Langhenhove et al., 2008)

Figure 2.2 – *Localisation des aires sensorimotrices*

Méthodologie 3

Cette partie méthodologique sera divisée en deux grandes sous parties visant à présenter :

Inclure une meilleure description des figures

1. L'extraction des features : présentation des méthodes utilisées dans le cadre de l'extraction d'attributs issus de l'activité neuronale. De manière générale, nous avons étudiés des attributs spectraux comprenant :
   — Phase et puissance spectrale
   — Attributs de couplage
2. Le machine learning : présentation des principaux algorithmes testées dans le cadre du décodage de l'activité neuronale

3.1 Extraction des features

Comme nous l'avons décrit précédemment, l'objectif du décodage de l'activité neuronale est d'arriver à extraire des signaux cérébraux une information suffisamment pertinente pour pouvoir discriminer différents types de classes (exemple : mouvement vers la gauche Vs droite).
Tout les attributs testés dans le cadre de cette thèse sont des attributs spectraux, donc issus de bandes de fréquences. La plupart de ces outils partagent donc une partie méthodologique commune à savoir, le filtrage. De plus, la plupart sont extraits en utilisant la transformée d'Hilbert. Pour éviter une redondance à travers les attributs, nous allons tout d'abord introduire quelques pré-requis.

3.1.1 Pré-requis

Cette partie ne devrait pas figurer dans l'extraction des features

Filtrage

L'intégralité des filtrages dans cette thèse ont été effectués avec la fonction *eegfilt* (qui a ensuite été reproduite pour le passage à python). De plus, afin d'éviter tout phénomène de déphasage, le fonction *filtfilt* a été systématiquement utilisée afin que le filtre soit appliqué dans les deux sens. Si cette dernière fonctionnalité n'est pas forcément indispensable dans le cadre d'un calcul de puissance, elle est absolument nécessaire pour un calcul de couplage phase-amplitude .
L'ordre du filtre présenté au dessus dépend de la fréquence de filtrage. Il

a systématiquement été calculé en utilisant la méthode décrite par Bahramisharif et al. (2013) :

$$FiltOrder = N_{cycle} \times f_s / f_{oi} \tag{3.1}$$

où f_s est la fréquence d'échantillonnage, f_{oi} est la fréquence d'intérêt et N_{cycle} est un nombre de cycles définit par $N_{cycle} = 3$ pour les oscillations lentes et $N_{cycle} = 6$ pour les oscillations rapides.

Transformée d'Hilbert

Transformée permettant de passer un signal temporel $x(t)$ du domaine réel au domaine complexe. Le signal peut ensuite s'écrire $x_H(t) = a(t)e^{j\phi(t)}$ où $a(t)$ est l'amplitude et $\phi(t)$, la phase. Cette transformation est particulièrement exploitée car le module de $x_H(t)$ permet de récupérer l'amplitude et la phase est obtenue en prenant l'angle de $x_H(t)$.

Transformée en ondelettes

La transformée en ondelettes (Tallon-Baudry et al., 1997, Worrell et al., 2012) permet de décomposer un signal dans le domaine temps-fréquence. La décomposition en ondelettes d'une fonction f est définie par :

$$f(a,b) = \int_{-\infty}^{\infty} \mathrm{f}(x)\overline{\psi}_{a,b}\mathrm{d}x \tag{3.2}$$

Où ψ est appelé ondelette mère dont la définition générale est donnée par $\psi_{a,b} = \frac{1}{\sqrt{a}}\Psi(\frac{x-b}{a})$ où a est le facteur de dilatation et b le facteur de translation. Le choix de l'ondelette mère s'est porté sur l'ondelette de Morlet qui est très largement utilisée à travers la littérature et définie par :

$$w(t, f_0) = A\mathrm{e}^{-t^2/2\sigma_t^2}\mathrm{e}^{2i\pi f_0 t} \tag{3.3}$$

Où $\sigma_f = 1/2\pi\sigma_t$ et $A = (\sigma_t\sqrt{\pi})^{-1/2}$. L'ondelette de Morlet est caractérisée par le ratio constant $r = f_0/\sigma_f$ que nous avons fixé égale à 7 comme suggéré par Tallon-Baudry et al. (1997).
Cette décomposition peut être comparée à la transformée courte de Fourier qui décompose le signal en une somme de combinaisons linéaire de sinus et de cosinus mais part du principe qu'il existe une régularité dans le signal permettant une telle décomposition. La transformée en ondelettes résout plusieurs limitations :

— Elle permet d'obtenir l'énergie d'un signal dans le temps, ce qui permet une bien meilleure exploration des phénomènes.
— Le rapport constant r permet d'obtenir des ondelettes dont la résolution fréquentielle varie en fonction des fréquences et permet une meilleure coïncidence avec la définition des bandes physiologiques (Bertrand et al., 1994)

Tout les attributs qui vont être maintenant présentés, utilisent les méthodes décrites ci-dessus.

Évaluation statistique à base de permutations

Pour une distribution de permutations construite à partir de deux sous-ensembles A et B et comportant N observations et pour une valeur p prédéfinie, on pourra conlure que :

— $A > B$ si A est parmi les $N - N \times p$ derniers échantillons (*"One-tailed test upper tail"*)

— $A < B$ si A est parmi les $N \times p$ premiers échantillons (*"One-tailed test lower tail"*)

— $A \neg B$ si A est soit inférieur aux $(N \times p)/2$ premiers échantillons soit supérieur aux $(N - N \times p)/2$

FIGURE 3.1 – *"One-tailed" et "two-tailed" test*

Grâce à cette méthode d'évaluation statistique, nous pourrons par exemple conclure si l'on a une augmentation, une diminution ou une différence statistique entre une valeur de puissance et la puissance contenue dans une période de baseline. Dernière précision, on comprend ainsi que pour obtenir une valeur p il faut que la taille de la distribution N soit au moins de $1/p$.

Hyperplan

Un hyperplan est un espace de co-dimension 1. Donc, dans un espace $3D$, l'hyperplan est un plan (dimension $2D + 1$). De manière générale, un espace de dimension N possède un hyperplan de dimension $N - 1$

$$dim_{ESPACE} = dim_{HYPERPLAN} + 1 \tag{3.4}$$

3.1.2 Puissance spectrale

Méthodes explorées

Le calcul de la puissance spectrale a été approché par deux méthodologies et qui ont été utilisés à des fins différentes :

— La transformée d'Hilbert : souvent exploité dans le cadre du décodage ainsi que pour garder une uniformité entre les attributs de phase et couplage phase-amplitude basés eux aussi sur cette transformée.

— La transformée en ondelettes : principalement utilisée pour la visualisation des cartes temps-fréquence à cause de l'adaptation des ondelettes aux bandes physiologiques.

Normalisation

On utilise la normalisation pour observer l'émergence d'un phénomène par rapport à une période définie comme baseline. A travers la littérature, quatre grands types de normalisation sont rencontrés :

1. Soustraction par la moyenne de la baseline
2. Division par la moyenne de la baseline
3. Soustraction puis division par la moyenne de la baseline
4. Z-score : soustraction de la moyenne puis division par la déviation de la baseline

La normalisation z-score est certainement la plus fréquemment rencontrée à travers la littérature. Le choix du type de normalisation dépend du type de données utilisées. Dans le cadre de nos données, *3.* était clairement la plus adaptée pour la visualisation. En revanche, dans le cadre de la classification, nous obtenions systématiquement de meilleurs résultats sans normalisation.

Évaluation statistique

La fiabilité statistique de la puissance a été évaluée en comparant chaque valeur de puissance à la puissance contenue dans une période définie comme baseline. Pour ce faire, nous avons testé deux approches :

1. Permutations : les valeurs de puissance et de baseline sont aléatoirement mélangées à travers les essais. Puis, on normalise cette puissance. En répétant cet procédure N fois, on obtient une distribution qui peut ensuite être utilisée pour en déduire la valeur p de la véritable puissance (cf : *pré-requis*)
2. "Wilcoxon signed-rank test" : ordonne les distances entre les paires de puissances (vraie valeur, baseline) (Demandt et al., 2012, Rickert, 2005, Waldert et al., 2008)

FIGURE 3.2 – *Exemple de représentation temps-fréquence de puissance normalisées z-score (Ossandon et al., 2011)*

3.1.3 Phase

L'extraction de la phase se fait de la même manière que pour le Couplage phase-amplitude , en prenant l'angle de la transformée d'Hilbert d'un signal filtré. La significativité peut être évaluée en utilisant le test de Rayleigh (Jervis et al., 1983, Tallon-Baudry et al., 1997). Point de vue pratique, cela correspond à la fonction *circ_rtest* de la toolbox Matlab *CircStat* (Berens and others, 2009)

3.1.4 Phase-amplitude coupling

Le calcul du Phase-amplitude coupling ne se limite pas uniquement à la méthode. En réalité, pour obtenir une estimation fiable sur des données réelles, il est indispensable de suivre les trois étapes suivantes :

1. Estimation de la véritable valeur de PAC. Il existe plusieurs méthodes.
2. Calcul de "*surrogates*" : on va calculer des PAC déstructurés. Idem, il existe de nombreuses méthodes
3. Correction du véritable PAC par les "*surrogates*". Cette correction, qui est en faite une normalisation, aura pour but de soustraire à l'estimation du PAC de l'information considérée comme bruitée.

Les sous-parties suivantes présenteront de manières succinctes les principales méthodes rencontrées dans la littérature, ainsi que différents types de corrections applicables.

Méthodologie du phase-amplitude coupling

Il existe une large variété de méthodes pour calculer le PAC, ce qui complique son exploration. Toutefois, il n'existe pas de consensus sur une méthode plus polyvalente qu'une autre, chacune possédant ses points forts et limitations. Pour aller un peu plus loin, et présenter quelques méthodes, il est nécessaire d'introduire quelques variable. Soit $x(t)$, une série temporelle de données de taille N. Pour cette série temporelle, on souhaite savoir si la phase extraite dans une bande de fréquence $f_\phi = [f_{\phi_1}, f_{\phi_2}]$ est couplée avec l'amplitude contenue dans $f_A = [f_{A_1}, f_{A_2}]$. Pour cela, on va tout

d'abord extraire $x_\phi(t)$ et $x_A(t)$ les signaux filtrés dans ces deux bandes. Enfin, la phase $\phi(t)$ est obtenue en prenant l'angle de la transformée d'Hilbert de $x_\phi(t)$ tandis que l'amplitude $a(t)$ est obtenue en prenant le module de la transformée d'Hilbert de $x_A(t)$.

1. Mean Vector Length-Modulation Index :
   Cette méthode à été introduite par Canolty et al. (2006) et consiste à sommer, à travers le temps, le complexe formé de l'amplitude des hautes fréquences avec la phase des basses fréquences. L'équation est donnée par :
   $$MVL = |\sum_{j=1}^{N} a_(j) \times e^{j\phi(j)}| \tag{3.5}$$

2. Kullback-Leibler divergence :
   A l'origine, la divergence de Kullback-Leibler (KLD), qui est issue de la théorie de l'information, permet de mesurer les dissimilarités entre deux distributions de probabilités. Ainsi, pour pouvoir utiliser cette mesure dans le cadre du PAC, Tort et al. (2010) propose une solution élégante qui consiste à générer une distribution de densité probabilités de l'amplitude (DPA) en fonction des valeurs de phase et d'ensuite utiliser le KLD pour comparer cette distribution à la densité de probabilité d'une distribution uniforme (DPU). Plus la DPA s'éloigne de la DPU, plus le couplage entre l'amplitude et la phase est consistant.
   Pour construire la DPA, l'astuce consiste à couper le cercle trigonométrique en N tranches (dans l'article il est proposé de couper en 18 tranches de 20°). Puis, si on prend l'exemple de la tranche $[0, 20°]$, on va chercher tout les instants temporels où la phase prend des valeurs comprises entre $[0, 20°]$ $(t, \phi(t) \subset [0, 20°])$. On prend ensuite la moyenne de l'amplitude pour ces valeurs de t et on répète cette procédure pour chacune des tranches de phase. On obtient ainsi la densité d'amplitudes en fonction des valeurs de phase. Il ne reste plus qu'à normaliser cette distribution par la somme des amplitudes à travers les tranches et on récupère une distribution de densité de probabilités. La figure 3.3 (Tort et al., 2010) présente un exemple de DPA en fonction de tranches de phase.

FIGURE 3.3 – *Densité de probabilité d'une distribution d'amplitudes en fonction de tranches de phases*

Le calcul de la divergence de Kullback-Leibler est ensuite appliqué

pour mesurer les dissimilarités entre la DPA et la DPU et c'est cette mesure qui servira d'estimation du couplage phase-amplitude :

$$D_{KL}(P,Q) = \sum_{j=1}^{N} P(j) \times \log \frac{P(j)}{Q(j)} \tag{3.6}$$

où $P(j)$ est la densité de probabilité de $a(t)$ en fonction de $\phi(t)$ et $Q(j)$ est la densité de probabilité d'une distribution uniforme.

3. Height Ratio
   La méthode du Height Ratio (Lakatos, 2005) est extrêmement proche du Kullback-Leibler divergence . En effet, l'amplitude sera binée de la même façon en fonction des tranches de phase. La mesure du PAC est ensuite donnée par :

$$hr = (f_{max} - f_{min})/f_{max} \tag{3.7}$$

   où f_{max} et f_{min} sont respectivement le maximum et le minimum de la de la densité de probabilité de l'amplitude en fonction des valeurs de phase.

4. Normalized Direct Phase-Amplitude Coupling
   Le Normalized Direct Phase-Amplitude Coupling , qui n'est pas une des méthodes les plus fréquemment rencontrées, présente toutefois une avantage certain. En plus de fournir une estimation fiable du couplage phase-amplitude , Ozkurt (2012) démontre l'existence d'un seuil à partir duquel on peut considérer l'estimation du PAC comme étant statistiquement fiable. La beauté de cette méthode, c'est que ce seuil statistique, qui est une fonction de la valeur p désirée, ne dépend que de la taille de la série temporelle. Ce qui rend son utilisation particulièrement simple.
   Pour estimer le PAC, une des hypothèses ayant permis d'aboutir à ce seuil statistique est de devoir normaliser l'amplitude par un z-score dénotée $\tilde{a}(t)$. L'estimation du PAC est quasiment identique au MVL puisque c'est en réalité le carré de celle-ci. Enfin, pour une valeur p désirée, l'article introduit le seuil statistique :

$$x_{lim} = N \times [erf^{-1}(1-p)]^2 \tag{3.8}$$

   où erf^{-1} est la fonction d'erreur inverse. On déduira que l'estimation PAC est significative si et seulement si cette valeur est deux fois supérieur à ce seuil.

5. Autres méthodes : Tout les algorithmes présentés ci-dessus ont été testés, implémentés et comparés. En complément, voici une liste non exhaustive d'autres méthodes existantes :

— *Phase Locking Value (PLV)* (Cohen, 2008, Penny et al., 2008) : détournement du PLV proposé par Lachaux et al. (1999) qui mesure la synchronie de phase entre deux électrodes. Cette méthode va comparer la phase des basses fréquences avec la phase de l'amplitude des hautes-fréquences.
— *Generalized Linear Model (GLM)* (Penny et al., 2008) : outil décrit comme adapté aux données courtes et bruitées.
— *Generalized Morse Wavelets (GMW)* (Nakhnikian et al., 2016) : basée sur des ondelettes, semble particulièrement utile dans le cadre de l'exploration des données.
— *Oscillatory Triggered Coupling (OTC)* (Dvorak and Fenton, 2014, Watrous et al., 2015) : issue d'une détection de maximums des hautes fréquences.

Correction du phase-amplitude coupling et évaluation statistique

Nous avons vu dans la section précédente différentes méthodes permettant de calculer un Couplage phase-amplitude . Toutefois, celui-ci peut être largement améliorée en faisant une estimation du PAC contenu dans le bruit des données. Une fois que cette estimation sera faite, on pourra retrancher ce PAC bruité à la valeur initiale. Tout comme il existe plusieurs méthodes de PAC, les équipes de recherche proposent à tour de rôle de nouvelles méthodes. Parmi elles, on peut citer :
— *Time-lag* : proposée par Canolty et al. (2006), on introduit un délai sur l'amplitude compris entre $[f_s, N - f_s]$ où f_s est la fréquence d'échantillonnage et N est le nombre de points de la série temporelle
— *Shuffling des couples [phase,amplitude]* : ici, on mélange aléatoirement les essais de phase et d'amplitude (Tort et al., 2010)
— *Swapping temporel d'amplitudes (ou de phase)* : on mélange aléatoirement les essais d'amplitude puis on recalcule le PAC avec la phase originale (Bahramisharif et al., 2013, Lachaux et al., 1999, Penny et al., 2008, Yanagisawa et al., 2012)

Ces trois méthodes produisent une distribution de *surrogates*. On pourra ensuite appliquer un z-score à la véritable estimation en utilisant la moyenne et la déviation de cette distribution. Enfin, l'évaluation statistique se fait également à partir de cette distribution (cf : *pré-requis*)
A ma connaissance, il n'existe pas de comparatif entre ces corrections et je n'ai jamais rencontré d'articles mentionnant que l'on ne puisse pas combiner les méthodes de PAC avec les différentes corrections. En revanche, ce qui est relaté c'est que le *time-lag* nécessite des données longues dû à l'introduction de ce délai temporel.

Comparatif des méthodes

Penny et al. (2008) ont comparé plusieurs méthodes dont le *MVL*, *PLV* et le *GLM* et Tort et al. (2010) ont complété cette étude avec d'autres méthodes (cf. A.1). Enfin, Canolty and Knight (2010) a fait une review qui comprend un descriptif très instructif.

Représentation du phase-amplitude coupling

Comparée à la puissance, l'exploration du PAC peut s'avérer plus complexe dû à sa dimensionnalité plus grande. Il existe donc des outils et des méthodes destinées à simplifier cette exploration et à visualiser ces résultats.
Exemple concret, si on cherche à connaître les modulations de puissance contenue dans un signal, on peut représenter une carte temps-fréquence . Pour le PAC, idéalement on voudrait visualiser les phases, les amplitudes et le temps mais ces trois dimensions empêche une représentation simple. On peut donc avoir recours à différents types de représentations complémentaires :

— Puissance phase-locked : cette représentation permet de faire émerger l'existence d'un couplage, pour une phase donnée, et d'observer sa durée. Pour cela, on aligne les phase en détectant le pic le plus proche de l'instant temporel étudié. On calcul les cartes temps-fréquence que l'on va ensuite moyenner après les avoir recalées de la même façon que les phases (c'est-à-dire avec la même latence).
— Comodulogramme : pour une tranche temporelle définie, on représente les valeur de PAC pour différentes valeurs de phase et d'amplitude

FIGURE 3.4 – ***(A)*** *Exemple de cartes temps-fréquence phase locked sur le β,* ***(B)*** *Exemple de comodulogramme*

La figure 3.4 (de Hemptinne et al., 2013) met en évidence que la représentation des cartes temps-fréquence phase-locked **(A)** est limitée d'une part, par la phase sur laquelle on choisit de recaler et d'autre part cette méthode est également limité par l'instant où l'on choisit de recaler. Pour la figure **(B)**, le calcul du PAC se faisant à travers la dimension temporelle, on a aucune idée de l'évolution du couplage dans le temps.

Phase-amplitude coupling : résolution temporel ?

Comment peut-on savoir si un ensemble de musiciens jouent ensemble, en rythme ? L'approche traditionnelle consiste à dire que, en fonction de la prestation du groupe, on sera en mesure de dire si ils étaient en rythme ou

non. Donc on focalise notre attention sur chaque instant du morceau et on analyse chaque note, chaque décalage. Cela signifie aussi que toute notre attention a été mobilisée par l'analyse du rythme et finalement, on passe à côté de la musique. Notre attention au détail nous a écarté du morceau global. On pourrait dire que l'on a écrasé la dimension temporelle du morceau. Une autre approche consiste à assister à toute les répétitions du fameux groupe. Ce faisant, on est capable de dire si d'une manière générale les musiciens ont tendance à jouer ensemble. Ainsi, le jour d'une représentation, toute notre attention peut rester uniquement sur le concert. On garde donc la dimension temporelle.
C'est par ce changement de positionnement face au problème de résolution temporelle que Voytek et al. (2013) introduit le Event Related Phase-Amplitude Coupling . L'approche traditionnelle du PAC nécessitant de connaître un nombre de cycles afin d'en déduire l'existence ou non du couplage, et donc perdre la dimension temps, l'article propose de calculer le PAC à travers les essais (ou répétitions). Pour un jeu de données de M essais de longueur N, on extrait respectivement les phases et les amplitudes $\phi_M(t)$ et $a_M(t)$ puis, pour chaque point temporel, on calcul la corrélation à travers les essais (corrélation linéaire-circulaire (Berens and others, 2009) qui se fait entre l'amplitude et des sinus/cosinus de la phase). Il en résulte une valeur de corrélation pour chaque instant et donc, de couplage.

3.2 Apprentissage supervisé

Le travail effectué durant cette thèse s'est exclusivement porté sur l'apprentissage supervisé. Celui-ci consiste à apprendre à la machine à reconnaître des événements qui ont été labellisé au préalable (cf. 3.2.1). A contrario, l'apprentissage non supervisé laisse la machine apprendre par elle-même. En pratique, l'apprentissage se fait sur des attributs. Par exemple, pour différencier des chats et des chiens, ou pourra utiliser l'angle formé par le sommet des oreilles. Les attributs doivent contenir une information pertinente permettant de différencier les classes. Enfin, les algorithmes de classification vont se servir de ces attributs pour définir une frontière entre les classes étudiées. A ce stade, il semble important de préciser que l'utilisation des outils d'apprentissage machine peut s'orienter (globalement) suivant deux axes :

1. Optimisation des attributs : on travail sur un raffinement des attributs afin que ceux-ci soient les plus performants possibles pour séparer les classes
2. Optimisation des paramètres de classification : on considère une base de données comme étant fixe, définitive, optimale et l'on va faire varier les différents paramètres liés à l'apprentissage machine (classifeurs, cross-validation...). C'est le cas des compétitions *BCI* où tout le monde travail sur une même base de données.

Bien sûr, ces deux axes peuvent être cumulés. Dans le cadre de cette thèse, le machine learning a été utilisé comme outil de validation d'hypothèses donc essentiellement porté sur l'optimisation des attributs. Le raffinement

des paramètres de classification a également été étudié, mais, au final, il ne constitue pas la majeure partie de l'étude.

Un schéma classique d'analyse peut-être décrit par :

1. Labellisation des données
2. Constitution de données d'entraînement (*training*) et de test (*testing*)
3. Choix d'un classifieur puis entraînement de celui-ci sur les données *training*
4. Test de ce classifeur entraîné sur les données *testing* et évaluation de la performance
5. Évaluation statistique de cette acuité de décodage

3.2.1 Labellisation et apprentissage

La labellisation c'est le fait d'associer à chaque événement l'appartenance à une classe ou à une condition. C'est par ce procédé que l'on va pouvoir apprendre ensuite au classifieur à identifier les classes. Par exemple, considérons *up* et *down* deux classes qui reflètent des mouvements de la main vers le haut ou vers le bas. On va donc construire un vecteur $y_{direction}$ qui labellise chaque essais avec direction effectuée (ce vecteur peut aussi être booléen ou contenir des entiers. L'essentiel est que à chaque classe soit attribué une valeur qui lui est propre). Ce vecteur y est appelé *vecteur label*, qui vient labelliser chaque essais d'un vecteur d'attributs x.

$$y_{direction} = \begin{pmatrix} up \\ down \\ down \\ \vdots \\ up \end{pmatrix}, y_{bool} = \begin{pmatrix} 0 \\ 1 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, x = \begin{pmatrix} x_{trial_1} \\ x_{trial_2} \\ x_{trial_3} \\ \vdots \\ x_{trial_N} \end{pmatrix} \tag{3.9}$$

D'où le nom apprentissage supervisé. Finalement, l'apprentissage machine se fera grâce à ce vecteur label y et cette matrice d'attributs x. Ce qui nous amène directement aux notions de *training set* et de *testing set*.

x												
y	0	0	0	1	1	1	2	2	2	3	3	3

FIGURE 3.5 – *Labellisation de données*

3.2.2 *Training, testing* et validation-croisée

Cette section est sans aucun doute la plus importante pour le machine learning puisque c'est elle qui assure la conformité méthodologique.
Un bon exemple pour comprendre cette partie est celui des contrôles de mathématiques. Avant l'examen, l'étudiant s'entraîne sur une série d'exercices. C'est la phase de *training* . D'ailleurs, plus il s'entraîne, plus ses chances de réussir à l'examen sont grandes. Le jour du contrôle, le professeur teste l'étudiant sur une série de nouveaux exercices en lien avec ce

qu'il a étudié. C'est le *testing* . Ici, c'est un test parfait puisque l'étudiant est naïf sur le contenue de l'examen ce qui veut dire que l'on teste ses capacités mathématiques pures. Toutefois, il peut arriver durant la scolarité que l'on soit testé sur des exercices que l'on a déjà vu dans la phase de *training* . Dans ce cas, la moyenne des notes des étudiants est généralement beaucoup plus élevée puisque l'on ne teste plus des capacités mathématiques, mais la capacité à restituer un apprentissage.

Training set, *testing set* et naïveté

Pour en revenir à la question du machine learning, on définit un une partie des données pour entraîner la machine. Ensuite, on teste cette machine entraînée sur un nouveau jeu de données de test. Il est essentiel d'avoir une séparation stricte entre des données définies comme *training* et des données de *testing* afin d'assurer la naïveté du classifieur. Même si cela peut parître évident, nous verrons que ça n'est pas toujours aussi facile que ça.
Se pose maintenant la question de comment l'on choisit de couper les données en *training* et *testing* . Une méthode serait de prendre une partie des données de manière aléatoire, de la définir comme *training* et sur tester sur les données restantes. Toutefois, ce choix ne représenterai qu'une partie des données. Une méthode plus exhaustive et plus rigoureuse consiste à utiliser une validation-croisée (ou cross-validation).

Validation-croisée

La validation-croisée (CV) est une procédure permettant de séparer les données en *training* et *testing* . Pour comprendre comment cela fonctionne, prenons un ensemble composé de N échantillons. Il existe plusieurs de CV mais de manière générale, toutes dérivent du même principe qui est la cross-validation k-Fold (Efron and Tibshirani, 1994, Kohavi and others, 1995). On coupe les N échantillons en k paquets de tailles égales (ou proches). Ensuite, le classifieur est entraîné sur $k-1$ paquets puis on le test sur le paquet restant. Cette procédure est ensuite appliquée k fois afin que chaque paquet passe au *testing* . On dira que la cross-validation est *stratified* si la proportion de classes représentées au sein de chaque dossier est approximativement uniforme à travers les folds. on pourra aussi rencontrer le terme *shuffle* si il y a un mélange supplémentaire. Tout cela nous emmène à des CV k-fold, k-fold stratified, k-fold shuffle ou encore k-fold stratified shuffle.
Concernant le nombre de folds, on rencontre en générale 3 valeurs à travers la littérature : 3-folds, 5-folds ou 10-folds (Latinne et al., 2001, Yanagisawa et al., 2009, Besserve et al., 2007, Waldert et al., 2008). Un cas particulier, mais si le nombre de folds $k = N$, ça revient à entraîner la machine sur $N-1$ échantillons tester sur celui qui a été isolé et on répète cette procédure N fois. C'est ce que l'on appelle le *Leave-One-Out* . Toutefois cette dernière possède une grande variance et peut conduire à des estimations non fiables (Efron and Tibshirani, 1994, Kohavi and others, 1995).
Un autre cas particulier, est celui du *Leave-p-Subject-Out* (Vidaurre et al., 2009, Lajnef et al., 2015) qui consiste à entraîner sur p sujets et tester sur

les sujets restants. Cette procédure est particulièrement exigeante puisqu'elle nécessite d'avoir une certaine reproductibilité entre les sujets. Cette validation-croisée est fréquente avec des données EEG mais impossible à mettre en œuvre pour la sEEG à cause de l'implantation unique de chaque sujet.

FIGURE 3.6 – *Exemple d'une cross validation 3-folds*

3.2.3 Classifieurs

1. Linear Discriminant Analysis (LDA)
   Le LDA (Fisher, 1936) est un classifieur linéaire. Pour un problème à deux classes, le LDA tente de trouver un hyperplan qui va maximiser la distance entre les classes tout en minimisant la variance inter-classes. Ce classifieur fait l'hypothèse que les données sont normalement distribuées avec la même co-variance. Un problème multi-classes pouvant être transformée en multiple bi-classes, le LDA tente de trouver un hyperplan séparant la classe du reste (*One-vs-All*)

FIGURE 3.7 – *Principe du Linear Discriminant Analysis (Lotte et al., 2007)*

2. Support Vector Machine (SVM)
   Le SVM (Boser et al., 1992, Cortes and Vapnik, 1995, Vladimir and Vapnik, 1995) utilise également un hyperplan pour séparer deux classes. Toutefois, cet hyperplan optimal est trouvé en maximisant les marges (ou distance) entre ce plan et les attributs les plus proches. Le SVM possède une particularité, il utilise un noyau qui peut permettre de résoudre les problèmes linéaire (*linear SVM*) mais également les problèmes non-linéaire en projetant les données dans un espace de dimension supérieure (*kernel trick*). Un noyau que l'on retrouve assez régulièrement est le *Radial Basis Function (RBF)* (Burges, 1998). Les problèmes multi-classes peuvent également être traités en *One-vs-All*

FIGURE 3.8 – *Principe du Support Vector Machine (Lotte et al., 2007)*

3. k-Nearest Neighbor (KNN)
   Pour un nouveau point de testing, le KNN (Fix and Hodges Jr, 1951) mesure la distance avec les k plus proches voisins et déduit la classe de ce point en fonction des classes de ces k-voisins (l'attribution de la classe se fait donc par vote)

FIGURE 3.9 – *Principe du k-Nearest Neighbor (Weinberger et al., 2005)*

4. Naive Bayes (NB)
   Le NB (Fukunaga, 1990) est un classifieur probabiliste. Une des hypothèses du NB est que les données dans les classes doivent être normalement distribuées et indépendante.

La figure A.3 en annexe, issue de l'excellentissime librairie python scikit-learn dédiée au machine learning, illustre le comportement de chaque classifieur façe à trois types de données. D'autres informations détaillées à propos des classifieurs peuvent être trouvées dans Lotte et al. (2007), Wieland and Pittore (2014), Wu et al. (2008)

FIGURE 3.10 – *Entraînement puis test d'un classifieur linéaire*

3.2.4 Évaluation de la performance de décodage

La question qui se pose maintenant, c'est comment évaluer la performance de décodage. Pour cela, on peut par exemple utiliser le *Decoding accuracy* ou le *roc*

1. Decoding accuracy (DA)
   L'utilisation (DA) est ce que l'on retrouve le plus fréquemment. Le calcul est simple, on compare les véritables labels avec les labels prédits par le classifieur. En faisant la somme des labels correctement prédits divisé par le nombre d'essais, on obtient un ratio qui correspond à l'acuité de décodage. Le plus souvent, ce ratio est ensuite exprimé en pourcentage. Le taux d'erreurs peut-être calculé en prenant $1 - DA$.

y_{TRUE}	$y_{PREDICTED}$
0	0
1	2
2	2
3	3
0	0
1	0
2	2
3	2
0	0
1	2
2	0
3	3

8/12 => 75%

FIGURE 3.11 – *Calcul de l'acuité de décodage*

2. Receiver operating characteristic (ROC)
   Une autre méthode pour évaluer la performance de décodage est l'utilisation de l'aire sous la courbe (AUC) ROC (Ling et al., 2003, Huang and Ling, 2005, Bradley, 1997). Celle-ci prend en compte le nombre d'essais correctement et incorrectement classifiés et pourrait donc prendre davantage de valeur possible comparé au Decoding accuracy .

3.2.5 Seuil de chance et évaluation statistique de la performance de décodage

De manière théorique, le seuil de chance est donné par $1/c$ où c est le nombre de classes. Par exemple, un problème à quatre classes donne un seuil de chance de 25%. Toutefois, ce seuil de chance est atteint pour un nombre de sample n infinis. En pratique, nous travaillons avec un nombre réduis de données, parfois même, avec très peu de sample. Dans ce cas, on peut obtenir des DA très élevés qui pourtant, ne sont pas pertinents. Les méthodes présentées ci-dessous ont pour but de trouver le seuil de chance associé à un jeu de donnée et de trouver pas la même occasion, la valeur p.

1. Loi binomiale
   En faisant l'hypothèse que l'erreur de classification suit une distri-

bution binomiale cumulative, on peut utiliser la loi suivante pour en déduire la probabilité de prédire au moins z fois la classe c :

$$P(z) = \sum_{i=z}^{n} \binom{n}{i} \times \left(\frac{1}{c}\right)^{i} \times \left(\frac{c-1}{c}\right)^{n-1} \tag{3.10}$$

2. Permutation
   Les permutations présentent l'avantage d'être calculées à partir des données (*data driven*). Ojala and Garriga (2010) nous renseigne sur les différents types de permutations possibles dans le cadre du décodage :
   (a) *Full permutation* : les données sont mélangés
   (b) *Shuffle y* : le vecteur de label est mélangé. C'est la procédure la plus fréquemment rencontrée.
   (c) *Intra-class shuffle* : les données sont mélangées à travers la dimension $features$ (colonne) et ce, à l'intérieur de chaque classe.

   Autant les méthodes (a) et (b) nous renseigne véritablement sur la consistance d'un décodage par rapport aux données, autant la méthode (c) donne des informations un peu différentes. En effet, en cas de décodage non-significatif, on pourra soit conclure qu'il n'y a pas de consistance dans les attributs à l'intérieur des classes, soit que le classifieur est incapable d'utiliser cette l'inter-dépendance. Ojala and Garriga (2010) précise que dans ce cas, il n'est pas nécessaire d'utiliser un classifieur compliqué et qu'un classifieur simple devrait suffire.

Cette partie est volontairement synthétique puisqu'elle a fait l'objet d'une publication scientifique (cf. II).
Un pipeline standard de classification est proposé en annexe (cf. A.2).

3.2.6 Du single au multi-features

Dans les sections précédentes, nous avons vu comment extraire des attributs de l'activité neuronale et comment les classifier. C'est ce que l'on appelle le *single feature* (SF), c'est-à-dire que l'on évalue la performance de chaque attribut séparément. Cette approche permet de constituer un set de features pertinents et répond à des questions neuro-scientifique. Cette démarche de SF a donc un but exploratoire.
La question que l'on peut maintenant se poser, c'est quelle performance de décodage puis-je obtenir si je combine ces attributs et dans quel cas est-ce utile ? C'est le multi-attributs (ou *multi-features* (MF)). Tout d'abord, le MF est utilisé lorsqu'il y a soit un désir soit un besoin de performances accrue. Par exemple, on utilisera le MF dans les compétitions de décodage ou tout simplement, pour une BCI où la performance est essentielle. Si l'on construit un système de bras robotisé piloté par activité neuronale, on comprend sans peine que celui-ci doit être le plus efficace possible et donc, le MF s'impose. Le dernier cas où l'on rencontre du MF, et ce n'est pas le cas le plus glorieux, c'est le cas où il y a un besoin de pallier à des résultats de SF assez faibles. La littérature expose des Decoding accuracy toujours plus hauts, des méthodes toujours plus complexes et donc, pour

publier correctement un article, il faut avoir des résultats au-moins aussi perspicaces.
Le multi-features c'est donc l'utilisation de multiples attributs pour aboutir à une classification et ce, sans sélection particulière. Individuellement, les attributs d'un même set n'auront pas la même performance. Certains seront des bons marqueurs et d'autres, n'ajouteront pas ou peu d'information. Donc en combinant ces features, il est probable que l'acuité de décodage soit moins bonne que la performance en attribut unique. Pour cela, on pourra donc utiliser des algorithmes de sélections de marqueurs (*feature selection*). Le but de cette sélection est de trouver dans un set d'attributs, un sous-ensemble dont la performance groupée est meilleure que la performance individuelle.
Cette sélection est une procédure exigeante où le risque de sur-apprentissage est grand. C'est la raison pour laquelle cette sélection doit être mise à l'intérieur d'une cross-validation . Donc on définis un set de *training* et de *testing* grâce à la validation croisée, puis sur le *training* , on lance la *feature selection*. On aboutit à un sous-ensemble de marqueurs qui va servir à entraîner le classifieur. Ensuite, on sélectionne ce subset dans le *testing* et on test le classifieur avec ce subset. Toute ceci étant enfin répété pour chaque *fold* de la cross-validation . A la vue de cette procédure, deux problèmes émergent :

— La sélection d'attributs se faisant à l'intérieur des folds de la cross-validation , on peut très bien aboutir à des listes d'attributs différentes. Pour obtenir une information finale, on pourra donc parler des attributs les plus fréquemment choisis. Par exemple, si la sélection se fait dans un cross-validation 10-folds, on pourra dire que le feature 1 a été choisi 7/10, le feature 2, 3/10...
— En fonction de la sélection choisie et de la cross-validation , le pipeline complet peut être très (très) lourd et long.

Les mécanismes de *feature selection* peuvent être regroupés en deux grandes familles (Guyon and Elisseeff, 2003, Liu et al., 2008, Das, 2001) : les *Filter methods* et les *Wrapper methods*.

Filter methods

Ces méthodes sont basées sur un critère et sont indépendantes du classifieur. Parmi elles, on retrouve des outils de corrélation, d'information mutuelle ou encore de statistiques. Ces derniers outils évaluent la contribution de chaque feature de manière indépendante sans tenir compte de la corrélation entre ces features. Pour résoudre ce problème, Yu and Liu (2004), Ding and Peng (2005) introduisent le *minimal-redundancy-maximal-relevance* qui en plus de trouver les features les plus pertinents, va permettre d'éliminer ceux qui sont redondants.
Pour terminer, ces méthodes sont effectivement indépendantes de l'algorithme de classification mais elles peuvent s'avérer optimales pour tel ou tel classifieur (ex : l'utilisation du critère de Fisher pour filtrer les features est très performant lorsqu'il est ensuite associé au Linear Discriminant Analysis (Duda et al., 2001)).

Wrapper methods

Contrairement aux méthodes de filtrage, les *wrapper* utilisent le classifieur comme outil de sélection. Le premier inconvénient que l'on peut d'ors et déjà leur reprocher, c'est que le résultat final sera donc classifieur-dépendant, donc difficile pour la généralisation.
Parmi ces *Wrapper methods*, on peut citer :

1. Sélection exhaustive : on teste toutes les combinaisons de features possibles puis on sélectionne la meilleure. Procédure qui ne peut être faisable qu'en présence d'un jeu de données particulièrement restreint.
2. Sélection sur la statistique de décodage : on utilise le classifieur pour évaluer l'acuité de décodage de chaque feature séparément pour en déduire une valeur p (cf : 3.2.5). Enfin, on sélectionne les features dont la valeur p est inférieur à un seuil désiré.
3. Sélection séquentielle : processus où l'on va ajouter/enlever des features de manière séquentielle jusqu'à atteindre un décodage optimal. Ce type de sélection se fait suivant deux directions :
   (a) *Forward feature selection* (FFS) : la première étape consiste à évaluer la performance de chaque attributs. On sélectionne le meilleur que l'on va ensuite combiner en couple avec tout les features restant. On sélectionne le meilleur couple puis on teste les combinaisons des meilleures triplettes... On continu tant que la performance s'améliore. Si le DA d'une étape i est inférieur au DA de l'étape $i-1$, on considère le nouveau subset de features à $i-1$.

FIGURE 3.12 – *Exemple d'une Forward feature selection appliquée sur six features*

   (b) *Backward feature elimination* (BFE) : la philosophie est la même que pour un *forward*. On classifie d'abord les N features pris ensemble, puis on enlève à tour de rôle chaque marqueur. On sélectionne le subset composé de $N-1$ features ayant fournit

le meilleur résultat, puis on enlève de nouveau chaque feature... L'algorithme s'arrête de la même façon que le *forward*.

FIGURE 3.13 – *Exemple d'une Backward feature elimination appliquée sur six features*

De manière générale, il est rapporté que la FFS converge plus rapidement que la BFE (Guyon and Elisseeff, 2003). Toutefois, la FFS tombe plus facilement dans des minimums locaux et donc, mène à un décodage moins bon. En effet, la *forward* sélectionne pas-à-pas les meilleurs attributs, elle est donc moins ensembliste que la *backward*.

Les méthodes de filtrage demandent moins de ressources et représentent donc un premier choix pour les larges sets de données. En revanche, elles peuvent ne pas déceler les phénomènes de complémentarité entre features. Pour cette dernière raison, les méthodes de wrapper fournissent en général de meilleurs résultats (Chai and Domeniconi, 2004).

3.2.7 Généralisation temporelle

L'introduction du *single-feature* faite plus haut était une présentation générique, c'est-à-dire que celle-ci est vraie quelque soit les features étudiés. On pourra donc classifier des attributs de puissance, de PAC, de phase, d'entropie... On peut également envisage l'étude un seul marqueur mais dans sa dimension temporelle. En effet, cela consiste à entraîner et tester un classifieur à différents instants temporels pour voir si le décodage varie dans le temps. Une des limitations de cette utilisation d'un classifieur est que, à chaque instant, celui-ci change. Donc on ne peut inférer aucune généralisation. Pour envisager une généralisation, il faut entraîner le classifieur à un instant puis le tester à travers toute la dimension temporelle restante. Dans ce cas, on pourra parler de généralisation mais reste encore le problème du choix de l'instant temporel qui servira à entraîner le prédicateur.

FIGURE 3.14 – *Exemple de décodage temporel (Waldert et al., 2008). Ici, l'auteur décode 4-directions de mouvements de la main dans le temps. A chaque instant, un classifieur est créé, entraîné puis testé à ce même instant.*

Pour répondre à cette question, King and Dehaene (2014) introduisent une idée particulièrement esthétique visant à généraliser le comportement d'un classifieur à travers le temps. Sans trop de surprise, ils ont nommé cette méthode la *généralisation temporelle*. Elle permet de répondre à deux limitations :

— Comment généraliser le comportement d'un classifieur lors d'une étude temporelle ?

— Comment choisir l'instant qui servira à entraîner le classifieur et quel impact ce choix aura-t-il sur le reste du décodage temporel ?

La méthodologie consiste à prendre un instant i, entraîner le classifieur et tester celui-ci sur tout les instants. Puis, on prend l'instant suivant $i+1$, on entraîne un nouveau classifieur et on teste... Et on répète cette procédure pour tout les instants. On obtient ainsi une représentation $2D$ où, par convention, l'axe des ordonnées matérialise l'endroit où le classifieur a été entraîné (*Training time*) et l'axe des abscisses pour tester ce prédicateur sur le reste de la dimension temporelle (*Generalization time*). La couleur permettra de signaler la performance de décodage.

FIGURE 3.15 – *Exemple de généralisation temporelle (King and Dehaene, 2014)*

Il y a aussi des interprétations neuroscientiques possibles. Comme le fait d'entraîner dans une phase et de tester dans une autre. Si ça marche, ça veut aussi dire que ces phases partagent des corrélats neuronaux. A voir si j'inclue ceci.

3.3 Configuration pour débuter

Dans la jungle des méthodes, il peut parfois être difficile de s'y retrouver. Cette section a pour objectif de fournir une liste de méthodes conseillées pour débuter. Rien ne dit que ce sont les meilleures méthodes mais elles ont le mérite d'avoir fait leurs preuves, que ce soit dans cette thèse et surtout, dans la littérature. Gardons à l'esprit que les meilleures méthodes dépendent des données mais certaines, sont plus polyvalentes.

Couplage phase-amplitude :
Kullback-Leibler divergence avec swapping des essais de phase et d'amplitude (Tort et al., 2010)

Classification :

1. Cross-validation : 10-folds
2. Classifieur : Support Vector Machine avec noyau linéaire (Vladimir and Vapnik, 1995, Lotte et al., 2007)
3. Évaluation statistique : permutations (Ojala and Garriga, 2010, Combrisson and Jerbi, 2015)
4. Multi-features : *Backward feature elimination* (Guyon and Elisseeff, 2003)

3.4 Implémentation des méthodes présentées

A force d'implémenter différentes méthodes afin de les tester, il se trouve que cet amas de fonction a formé assez naturellement une *toolbox*. Avec un peu de mise en forme, une uniformisation à travers les fonctions et la constitution d'une aide, nous avons décidé de mettre cet ensemble

d'outils *online* et en libre accès.
La toolbox s'appelle **brainpipe** , entièrement codée en Python. Le choix de python a également été mûrement réfléchis. La première version était en Matlab. Toutefois, Python présente les avantages d'être libre, beaucoup plus souple et intuitif et bien plus puissant. De plus, la communauté s'est appropriée ce langage et est particulièrement réactive. Quant à la documentation, elle est excellente et permet d'apprivoiser ce langage très rapidement. Autre point fort qui est une conséquence du libre, c'est la multitudes d'outils existants. Par exemple, pour la classification, il existe *scikit-leanr* dont la qualité est sans comparaison possible avec Matlab. Enfin, les derniers outils à la pointe (tel que *tensorflow* de Google) ne seront jamais proposé en Matlab. Le Julia (qui est un autre langage de programmation) semble particulièrement prometteur mais, étant très jeune et ne disposant que d'une petite communauté, il a été écarté. Autre point, les outils que nous avons présenté dans cette partie méthode peuvent nécessiter énormément de ressources. **brainpipe** a donc entièrement été développé pour le calcul parallèle qui est intégré par défaut. A titre de comparaison, le calcul du PAC sur tout un jeux de données a pris une semaine Matlab, contre une journée en Python.

brainpipe est dédiée à la classification de signaux cérébraux. La construction globale peut être résumée avec les points suivants :

— *Pre-processing* : pré-traitement de données (filtrage, re-référencement, informations anatomiques en fonction de coordonnées MNI...)
— *Features* : extraction d'une large variété de features (amplitude, puissance, phase, PAC, PLV, ERPAC, entropie...)
— *Classification* des signatures cérébrales : repose sur *scikit-learn* pour l'implémentation des algorithmes de classification. **brainpipe** sert à faire le lien entre l'analyse de données neuro-scientifiques et le *machine-learning*. Il y a également un module dédié au *multi-features*.
— *Statistics* : tout les attributs disposent des méthodes statistiques ainsi que pour la classification et la correction multiple.
— *Visualization* : enfin, un ensemble d'outils de visualisation sont également disponible afin de faciliter les représentations graphiques.

Liens vers **brainpipe** :

— Téléchargement : `https://github.com/EtienneCmb/brainpipe`
— Documentation en ligne : `https://etiennecmb.github.io/`

1. *ipywksp* : pour les personnes désirant ou utilisant déjà python, un environnement de développement particulièrement agréable et puissant s'impose de plus en plus. Ce projet s'appelle Jupyter. Pour l'introduire brièvement, c'est un notebook dans lequel le code est intégré dans des cellules qui peuvent être lancées individuellement. Une des lacunes de ce notebook pour les utilisateurs Matlab, c'est le manque d'un *workspace* pour voir les variables en cours. *ipywksp* est un petit module, également développé dans cette thèse qui vient combler ce manque. Toujours en libre accès : `https://github.com/EtienneCmb/ipywksp`

2. *visbrain* : enfin, un dernier module Python qui est en cours de développement mais qui proposera une interface graphique pour toute visualisation nécessitant un cerveau MNI $3D$ (connectivité, projection corticale...). Il existe de nombreux logiciels existants déjà. Le but de celui-ci sera de permettre une intégration totale avec **brainpipe** . Le projet débute, mais le lien suivant permet de suivre son évolution : `https://github.com/EtienneCmb/visbrain`

Données expérimentales 4

Durant cette thèse, l'exploration s'est faite chez l'Homme par le biais, principalement, de données de type Stéréoélectroencéphalographie (SEEG). Ces données rares de très grande qualité (cf. 4.1.2) ont été acquise avant le début de la thèse, ce qui a permis de rentrer dans le vif du sujet très rapidement, après une période d'acclimatation aux différents traitements, propres à ce type d'enregistrement. D'autres types tel que l'EEG, la MEG ou les micro-électrodes ont également été approchés mais de manière ponctuelle, comme ce fût le cas dans l'Étude 1 (cf. II) ou dans différentes collaborations. Toutefois, étant donné que le temps consacré à ces données ne représente qu'une faible portion du travail total, nous allons ici nous concentrer uniquement sur l'intra.

Pour commencer, nous verrons ce que l'analyse de la Stéréoélectroencéphalographie a de particulier (richesse des données, qualité, les traitements associés, les avantages et les limitations). Enfin, nous verrons concrètement les enregistrements qui ont été utilisés dans le cadre de cette thèse.

4.1 Données intracrâniennes

4.1.1 Acquisition

La première question que l'on est en droit de se poser, c'est comment est-il possible de travailler, chez l'Homme, avec des enregistrements qui nécessitent une implantation invasive, c'est-à-dire dans le cortex ? Certaines personnes présentent des formes agressives d'épilepsies, pouvant s'avérer pharmacorésistantes. En fonction de la localisation du foyer épileptogène, les méfaits engendrés par les décharges épileptiques peuvent être variés. Dans ce cas, il est nécessaire de localiser ce foyer avec, de préférence, des techniques non-invasives telles que l'EEG ou la MEG. Mais si ces dernières ne permettent pas une localisation précise le patient sera implanté avec des macro-électrodes comme la SEEG pour tenter de localiser puis d'enlever ce foyer par intervention chirurgicale. Cette implantation a un second objectif, déterminer quel est le rôle fonctionnel de la structure lésée (rôle moteur, langage, vision...). C'est dans ce contexte que les chercheurs proposent au patient de participer à une étude scientifique.

4.1.2 Avantages et limitations des données intracrâniennes

Le paragraphe précédent met en exergue la rareté de ces données. De plus, ce type d'acquisition enregistre l'activité cérébrale d'une population

relativement restreinte de neurones. En conséquence, on peut espérer que ce petit groupe s'active dans des processus précis et ainsi, étudier des phénomènes fins. Enfin, le rapport signal sur bruit (RSB) de la SEEG est excellent, ce qui doit permettre l'étude de processus, même en essais unique là où d'autres types d'enregistrements auront besoin d'une large banque d'essais avant de pouvoir constater l ?émergence d'un phénomène.
La SEEG présente toutefois quelques limitations que l'on peut nuancer. Le problème majeur est certainement la généralisation d'un phénomène ou la reproductibilité à travers les sujets. La pathologie est propre à chaque patient, donc son implantation aussi. Ce qui signifie qu'il n'y a aucune chance que plusieurs sujets présentent rigoureusement la même implantation. Pour contourner cette limitation, ou pourra utiliser :

— Des régions d'intérêt (ROI) : on va regrouper les électrodes des différents sujets par "proximité" en faisant l'hypothèse que celles-ci s'activent de façon similaire face à un processus. Ces ROI pourront être par exemple les gyrus ou les aires de Brodmann. Bien sûr, ce que l'on gagne on généralisation, on le perd en précision.
— Projection corticale : autour de chaque électrode, on définit une sphère d'intérêt (généralement 10 millimètres de rayon) puis on prend l'intersection de ces sphères avec la surface du cerveau. Cette technique permet une visualisation des activités proches de la surface à travers les sujets mais on perd de la lisibilité sur ce qui se passe en profondeur.

Une utilisation combinée des ROI et de la projection corticale permet de palier, au moins partiellement, au problème de reproductibilité inter-sujets. Autre limitation, on ne dispose que d'une couverture partielle puisque le neuro-chirurgien implante une quantité limitée d'électrodes. Ce dernier point est traité en augmentant le nombre de sujets. Enfin, la dernière limitation que l'on soulèvera ici, concerne le fait de travailler sur un cerveau "malade" empêchant donc une généralisation à des sujets sains. On limite ce problème par un ensemble de prétraitements (Jerbi et al., 2009) décris dans le prochain paragraphe.
Un dernier point que l'on peut argumenter à la fois comme avantage ou limitation, c'est de ne pas pouvoir contrôler l'implantation pour étudier un phénomène précis. Par exemple, si l'on analyse l'encodage moteur, on s'attendrait à concentrer les efforts sur le cortex moteur primaire ou prémoteur. Or l'implantation SEEG peut très bien contenir du frontal, du pariétal, du temporal. Là où finalement on peut considérer ça comme un avantage, c'est que l'on a accès à un ensemble de structures jugées non-primordiales mais dont l'ajout pourrait permettre la compréhension d'un processus de manière plus globale.

4.1.3 Prétraitements

Le premier prétraitement appliqué a été une réjection des sites bruités ou présentant une activité pathologique, c'est-à-dire des décharges épileptiques. C'est par cette inspection manuelle que l'on augmente le potentiel de généralisation aux sujets sains.
Autre prétraitement, les données peuvent être bipolarisées comme c'est le cas dans de nombreuses études (Bastin et al., 2016, Ossandon et al., 2011,

Jerbi et al., 2009). La bipolarisation part du principe que deux sites proches enregistrent des activités neuronales différentes mais que toute source de bruit, ou influence de sources lointaines, se retrouvera sur ces deux sites. La technique de bipolarisation consiste donc à soustraire les activités neuronales de sites proches ce qui a pour effet de supprimer la partie commune, le bruit. Par exemple prenons une électrode contenant les sites $k9$, $k10$, $k11$ et $k12$. Après bipolarisation, on considérera les sites matérialisés par $k10 - k9$, $k11 - k10$ et $k12 - k11$. Les bénéfices de la bipolarisation peuvent être résumés par :

1. Limitation des influences des sources lointaines et de la tension secteur (50hz)
2. Augmentation de la spécificité qui, pour un site bipolarisé, est estimée à 3mm (Kahane et al., 2006, Lachaux et al., 2003, Jerbi et al., 2009)

4.2 Données d'étude

Trois jeux de données intracrâniaux ont été exploré :

1. *Center-out* : étude de l'encodage et du décodage des actions et des intentions motrices lors de mouvements de main
2. *Occulo* : étude des intentions et décision de mouvements oculaires
3. *Emotions* : étude de l'encodage des émotions

4.2.1 Données *Center-out*

C'est le jeu de données qui a été le plus largement exploité. En effet, celui-ci a servis à étudier l'encodage (cf. III) et le décodage (cf. IV) des actions motrices chez l'homme.

Descriptif des données

Six sujets (six femmes), implantés au département de l'épilepsie de l'hôpital de Grenoble ont donné leur consentement écrit pour passer l'expérience, sous la supervision du personnel médical. Le tableau 4.2.1 résume les détails clinique des différents sujets.

	Dominance	Age	Genre	Zone épileptique
P1	D	19	F	Frontal (RH)
P2	D	23	F	Frontal (LH)
P3	D	18	F	Frontal (RH)
P4	D	18	F	Frontal (RH)
P5	D	31	F	Insula (RH)
P6	D	24	F	Frontal (LH)
Moyenne : 22.17 ± 4.6				

Figure 4.1 – *Détails cliniques des sujets ayant participé à la tâche Center-out*

Matériel d'acquisition

De 12 à 15 multi-électrodes ont été implantées dans différentes structures. Chaque multi-électrode possède entre 10 et 15 sites mesurant 0.8mm et séparés de 1.5mm. La localisation anatomique des électrodes s'est faite en utilisant le schéma d'implantation (exemple en annexe A.4) et l'atlas proportionnel de Talairach et Tournoux (Talairach and Tournoux, 1993). La visualisation de la pré-implantation s'est faite par $IRM - 3D$ et un $CT - scan$ a été utilisé pour la post-implantation. Enfin, un IRM a également servis pour visualiser les électrodes implantées dans la matière blanche. Les coordonnées Talairach ont été déduites du $CT - scan$ puis ont été transformées en MNI afin de pouvoir les superposer dans un cerveau standard (cf. figure ci-dessous).
Le système Micromed a été utilisé pour visionner l'acquisition de l'activité neuronale. Une électrode prise dans la matière blanche a été prise comme référence et un filtrage *passband* entre $[0.1, 200hz]$ a également été effectué *online*. La fréquence d'échantillonnage est de $1024hz$.

FIGURE 4.2 – *Implantation intracrâniale et couverture corticale de six sujets épileptiques ayant passés la tâche Center-out*

Descriptif de la tâche

La tâche est composée de trois phases :

1. Phase de repos : on demande au sujet de rester immobile pendant une durée de une seconde
2. Phase de préparation motrice (**CUE 1**) : une direction est imposée à l'écran (haut/bas/gauche/droite). On demande au sujet de se préparer pendant 1.5s à bouger la souris dans la direction imposée.
3. Phase d'exécution motrice (**CUE 2**) : le sujet exécute le mouvement en bougeant la souris du centre à l'extrémité de l'écran indiquée (environ 1.5s) puis de revient vers le centre.

FIGURE 4.3 – *Descriptif de la tâche Center-out*

Cette tâche à conduit à deux études essentielles :

— L'étude de l'encodage des intentions motrices (cf. III) : on étudiera le décodage du repos vs préparation, repos vs exécution et de la préparation vs exécution

— L'étude du décodage des directions de mouvement (cf. IV) que ce soit pendant la préparation ou pendant l'exécution motrice.

4.2.2 Autres données

— Occulo : Données occulo mais un seul sujet donc pas super cool. De plus, les résultats sont un peu vieux et mériteraient de se pencher dessus une fois pour toute.

— Emotions : Ce serait pas mal que j'ai quelques résultats sur les émotions, histoire de montrer aussi un peu la diversité des données utilisées.

Ouverture 5

Nos contributions portent sur : . . .

Le *premier chapitre* expose la problématique de la thèse.
Le *deuxième chapitre* présente en détail . . .

etc.
Cette thèse a fait l'objet de divers travaux écrits : . . .

Deuxième partie

Étude 1 : niveau de chance et évaluation statistique des résultats de classification par apprentissage supervisé

Introduction

L'utilisation des outils de *machine-learning*, dans le domaine des neurosciences, est de plus en plus fréquente et se sert de ces méthodes comme outils de validation. En effet, l'extraction d'attributs puis la classification de ceux-ci permet d'explorer ce qui, dans l'activité neuronale, reflète un changement d'état. Il y a donc un besoin d'évaluer la performance de décodage et sa validité statistique et donc quantifier à quel point ce décodage s'éloigne de ce que l'on pourrait atteindre par chance.
Lorsque l'on parle de chance dans le domaine de l'apprentissage machine, il faut distinguer la chance théorique de la chance empirique. Des problèmes à deux classes, à quatre classes ou à huit classes auront respectivement des seuils de chance de 50%, 25% et 12.5%. Il faut considérer ces seuils à titre indicatif, comme première estimation, puisque en réalité ceux-ci sont atteints pour une infinité d'échantillons. Pour comprendre ce problème de seuil de chance, prenons un exemple basé sur des pièces. On sait que le seuil de chance d'avoir soit pile soit face est de 50%. Si on lance quatre pièces en même temps ce seuil de 50% est exprimé lorsqu'il y a autant de piles que de faces, donc deux pièces chacun. Maintenant, si une pièce ne représente pas son groupe, les répartitions sont de 75% contre 25%. Avec 10 pièces, si une pièce ne représente pas son groupe on obtient 60% contre 40%. Pour 100 pièces on a 51% contre 49% etc. Plus le sac contient de pièces, plus le seuil de chance pratique va tendre vers le seuil de chance théorique. Si les spécialistes du *machine-learning* ont parfaitement conscience de ce problème, la communauté neuro-scientifique affiche parfois un manque de rigueur face à cette limitation.

Cette étude, qui se destine principalement les étudiants ou aux personnes débutant dans le domaine, a pour objectif de quantifier ce problème du nombre d'essais et de présenter les outils pour le mesurer. Pour cela, nous avons générer des signaux Gaussien aléatoires et nous avons mesuré l'impact du nombre d'essais sur le seuil de chance ainsi que sur les principaux outils du *machine-learning* (classifieur / cross-validation / nombre de répétitions). En effet, nous avons cherché à savoir si les classifieurs (Linear Discriminant Analysis , Naive Bayes , Support Vector Machine) étaient affectés de la même façon pour des petits échantillons, si certains types de cross-validation (k-Fold et Leave-One-Out) étaient plus robustes et si le nombre de répétitions permettaient de minimiser cet effet. De plus, nous rappelons que des outils analytiques (loi binomiale) ou empiriques (permutations) préviennent de ce problème du nombre d'échantillons en introduisant des valeurs statistiques. Enfin, nous avons répliqué cette étude sur des données MEG et iEEG et fournissons une toolbox Matlab pour reproduire ces résultats et contenant les outils statistiques présentés.

G Model
NSM-7116; No. of Pages 11
ARTICLE IN PRESS
Journal of Neuroscience Methods xxx (2015) xxx–xxx
Contents lists available at ScienceDirect
Journal of Neuroscience Methods
journal homepage: www.elsevier.com/locate/jneumeth

Computational Neuroscience

Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy

Etienne Combrisson[a,b], Karim Jerbi[a,c,*]

[a] DYCOG Lab, Lyon Neuroscience Research Center, INSERM U1028, UMR 5292, University Lyon I, Lyon, France
[b] Center of Research and Innovation in Sport, Mental Processes and Motor Performance, University of Lyon I, Lyon, France
[c] Psychology Department, University of Montreal, QC, Canada

ARTICLE INFO

Article history:
Received 28 July 2014
Received in revised form 6 January 2015
Accepted 7 January 2015
Available online xxx

Keywords:
k-Fold cross-validation
Small sample size
Classification
Multi-class decoding
Brain–computer-interfaces (BCIs)
Machine learning
Binomial cumulative distribution
Classification significance
Decoding accuracy
MEG
ECoG
Intracranial EEG

ABSTRACT

Machine learning techniques are increasingly used in neuroscience to classify brain signals. Decoding performance is reflected by how much the classification results depart from the rate achieved by purely random classification. In a 2-class or 4-class classification problem, the chance levels are thus 50% or 25% respectively. However, such thresholds hold for an infinite number of data samples but not for small data sets. While this limitation is widely recognized in the machine learning field, it is unfortunately sometimes still overlooked or ignored in the emerging field of brain signal classification. Incidentally, this field is often faced with the difficulty of low sample size. In this study we demonstrate how applying signal classification to Gaussian random signals can yield decoding accuracies of up to 70% or higher in two-class decoding with small sample sets. Most importantly, we provide a thorough quantification of the severity and the parameters affecting this limitation using simulations in which we manipulate sample size, class number, cross-validation parameters (*k*-fold, leave-one-out and repetition number) and classifier type (Linear-Discriminant Analysis, Naïve Bayesian and Support Vector Machine). In addition to raising a red flag of caution, we illustrate the use of analytical and empirical solutions (binomial formula and permutation tests) that tackle the problem by providing statistical significance levels (*p*-values) for the decoding accuracy, taking sample size into account. Finally, we illustrate the relevance of our simulations and statistical tests on real brain data by assessing noise-level classifications in Magnetoencephalography (MEG) and intracranial EEG (iEEG) baseline recordings.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Applying machine learning algorithms to brain signals in order to predict intentions or decode cognitive states has become an increasingly popular technique over the last decade. The surge in the use of machine learning methods in neuroscience has been largely fueled by the tremendous increase in brain–computer interface (BCI) and brain signal decoding research either using non-invasive recordings such as Electroencephalography (EEG) or Magnetoencephalography (MEG) (e.g. Aloise et al., 2012; Besserve et al., 2007; Jerbi et al., 2011; Krusienski and Wolpaw, 2009; Toppi et al., 2014; Waldert et al., 2008) or with intracranial EEG (e.g. Ball et al., 2009; Derix et al., 2012; Hamamé et al. (2012); Korczyn et al. (2013); Lachaux et al., 2007a,b; Leuthardt et al., 2004, 2006; Mehring et al., 2004; Pistohl et al., 2012; Schalk et al., 2008; Jerbi et al., 2007a,2009a,2013). Machine learning and signal classification techniques are powerful and complex tools that have to be used with caution. While most machine learning experts are well aware of the various caveats to watch out for, certain theoretical limitations of these methods can easily elude students and neuroscience researchers new to the field of machine learning and brain–computer interface research.

In supervised learning, samples of a subset of the data and knowledge of their corresponding class (label) are used to train a model to distinguish between two or more classes. The trained classifier is then tested on the remaining data samples (the hold-out samples). This procedure is generally repeated several times by varying the subsets used for training and those used for testing, a standard procedure known as cross-validation. The percent of over-all correct label (or class) prediction across the test samples of the multiple folds is known as the correct classification

* Corresponding author at: Psychology Department, University of Montreal, QC, Canada.
E-mail address: karim.jerbi@umontreal.ca (K. Jerbi).

http://dx.doi.org/10.1016/j.jneumeth.2015.01.010
0165-0270/© 2015 Elsevier B.V. All rights reserved.

Please cite this article in press as: Combrisson E, Jerbi K. Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. J Neurosci Methods (2015), http://dx.doi.org/10.1016/j.jneumeth.2015.01.010

rate (sometimes called decoding accuracy). Conversely, the mean of misclassified samples over the folds is a measure of classifier prediction error.

The performance of a classifier in neural decoding studies is often assessed by how close its correct classification rate is to the maximum of 100%, or alternatively, how strongly it departs from the *chance-level* rate achieved by a classifier that would randomly associate the samples to the various classes. For instance, in a two-class or four-class classification problem, the probabilistic chance level indicating totally random classification is 50% or 25% respectively. Yet, although such probabilistic chance-levels widely applied in brain signal classification studies, they can be problematic because they are strictly speaking only valid for infinite sample sizes. While it will not come to anyone as a surprise that no study to date was able to acquire infinite data, it is intriguing how rarely brain signal classification studies acknowledge this limitation or take it into account. For a two-class classification problem with small sample size, 60%, 70% or even higher decoding percentages can in theory arise by chance (see simulation results below). As a consequence, for finite samples, a decoding percentage can only be considered reliable if it substantially, or better still, *significantly* departs from the theoretical level in statistical terms. But how can we assess the significance of the departure of a decoder from the outcome of total random classification? For a given sample size and a given number of classes, what would be the statistically significant threshold of correct classification that one needs to exceed in order to consider the decoding *statistically significant*? Although these questions have been widely recognized and addressed in the machine learning field (e.g. Kohavi, 1995; Martin and Hirschberg, 1996a,b), it is unfortunately often overlooked in the emerging field of brain signal classification which, incidentally, is often faced with low sample sizes for which the problem is even more critical.

Not all the previous brain decoding reports suffer from the caveat of using theoretical chance-level as reference. However, numerous studies only apply statistical assessment when testing for significant differences between the performance of multiple classifiers, or when comparing decoding across experimental conditions, but unfortunately neglect to provide a statistical assessment of decoding that accounts for sample size (e.g. Felton et al., 2007; Haynes et al., 2007; Bode and Haynes, 2009; Kellis et al., 2010; Hosseini et al., 2011; Sitaram et al., 2011; Hill et al., 2006; Wang et al., 2010; Bleichner et al., 2014; Babiloni et al., 2000; Ahn et al., 2013; Morash et al., 2008; Neuper et al., 2005; Kayikcioglu and Aydemir, 2010; Momennejad and Haynes, 2012). A number of such studies use theoretical percent chance-levels (e.g. 50% in a 2-class classification) as a reference against which classifier decoding performance is assessed. By doing so, such studies fail to account for the effect of finite sample size. This may have little effect in the case of large sample size or when extremely high decoding results are obtained, however, the bias and erroneous impact of such omissions can be critical for smaller sample sizes or when the decoding accuracies are barely above the theoretical chance levels.

Note however, that the rigorous assessment of significant classification thresholds is not equally ignored across the various types of neuronal decoding studies; it seems that the omissions (or unfortunate tendency to rely on the theoretical chance levels) are more common in more recent sub-branches of the neuronal decoding field. This is the case for signal classification and BCI studies based on non-invasive (fMRI, EEG and MEG) brain recordings in humans, and possibly electrocorticographic macro-electrode recordings in patients, where the methods (including classifiers, features and statistics) are less well-established than in the field of neuronal spike decoding in primates for instance.

In this brief article, we address caveats related to interpreting brain classification performances with small sample sizes. The paper is written with the broad neuroscience readership in mind and is oriented, in particular, to students and researchers new to neural signal classification. First of all, we describe how applying signal classification to randomly generated signals can yield decoding accuracies (correct classification rates) that strongly depart from theoretical chance levels, with values up to 70% and higher with small sample sizes (instead of the expected theoretical 50% for 2-class decoding). Most importantly, we illustrate and quantify the phenomenon by using simulations in which we manipulate sample size, class number, cross-validation parameters and classifier type. In addition to raising a red flag of caution, we recommend practical alternatives to overcome the problem. We describe a straight-forward method to derive a statistically significant threshold that accounts for sample size and provides confidence intervals for the classification accuracy achieved by cross-validations. A reference table is also provided to allow readers to quickly look-up the percent correct classification thresholds that need to be exceeded in order to assert statistical significance of the findings for a range of possible sample sizes, classes and significance levels.

2. Materials and methods

2.1. Data simulation and classification

2.1.1. Generating normally distributed random data

In order to simulate a situation with classification results that approach the theoretical chance level, we generated 100 data sets of zero-mean Gaussian white noise. The normally distributed variables in each data set were generated in MATLAB (Mathworks Inc., MA, USA) via a pseudo-random number generator. Each one of the 100 data sets was randomly split into c subsets data (here we used $c = 2$- or 4-classes) and we then evaluated the classification performance obtained by applying different classification algorithms to these simulated datasets. Because the variables in each 'simulated class' were drawn from the exact same Gaussian random distribution data set, applying supervised machine learning algorithms should fail to distinguish between classes and should theoretically yield chance-level classification rates (50% for $c = 2$ and 25% for $c = 4$). To examine the effect of sample size on how close the empirical classifications are to the theoretically expected chance level we varied the total number of samples n from 24 to 500. In other words, in the 2-class simulation for instance, the number of samples in each class varied from 12 to 250. Note that the code we implemented for the generation of random data for classification purposes is provided online (see Appendix A).

2.1.2. Classification algorithms

We implemented three types of machine learning algorithms: linear discriminant analysis (LDA), naïve Bayes (NB) classifier and a support vector machine (SVM), the latter with two different kernels: a linear kernel and a radial basis function (RBF) kernel. These three methods, which are frequently used for neural signal classification in the context of brain–computer interface research are briefly described in the following.

Linear discriminant analysis: LDA (Fisher, 1936) is a straightforward and fast algorithm which assumes that the independent variables in each class are normally distributed with identical covariance (homoscedasticity assumption). For a two dimension problem, the LDA tries to find a hyperplane that maximizes the mean distance between the two classes while minimizing the interclass variance. A multiclass problem can be tackled as a multiple two-class problem by discriminating each class from the rest using multiple hyperplanes.

Naive Bayesian classifier: The NB model (e.g. Fukunaga, 1990) is a probabilistic classifier that assigns features to the class to which they have the highest probability of belonging. NB assumes that the

Please cite this article in press as: Combrisson E, Jerbi K. Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. J Neurosci Methods (2015), http://dx.doi.org/10.1016/j.jneumeth.2015.01.010

features in each class are normally distributed and independent. The name arises from the fact that it is based on applying Bayes' theorem with strong (naive) independence assumptions.

Support vector machine: SVM (Boser et al. 1992; Burges, 1998; Cortes and Vapnik, 1995; Vapnik, 1995) classifiers originate from statistical learning theory. An SVM searches for a hyperplane that maximize margins between the hyperplane and the closest features in the training set. For non-linearly separable classes, SVM uses a kernel function to project features in a higher dimensional space in order to reduce the nonlinear problem to a linear one, which is then separable by a hyperplane. The (Gaussian) Radial Basis Function (RBF) kernel is a popular choice. In this study, both linear and RBF kernels were used for SVM classification.

Details of the theoretical background of various classifiers can be found in standard statistics and machine learning textbooks and various reviews (e.g. Lotte et al., 2007; Wieland and Pittore, 2014). Here, we used MATLAB implementation for the LDA and NB and the libsvm library for multi-class SVM.

2.1.3. Repeated and stratified k-fold cross-validation

To compute the decoding accuracy achieved by each one of the classifiers on the random data, we used standard stratified k-fold cross-validation. For a given data set size, all available N samples are partitioned into k folds, where $(k-1)$ folds are used for training the classifier model (training set) and the remaining fold is used for validation (test set). This procedure is then repeated k times so that each fold is used once as test set. The stratified option ensures that each fold has approximately the same proportion of samples from each class as in the original dataset as a whole. The case $k = N$ (e.g. 200 folds in a data set of 200 samples) is called leave-one-out (LOO) cross-validation because one element is used to test the performance of a classifier trained on the rest of the data. Because k-fold cross-validation involves a random partition, the variance of the classifier can in theory be reduced by repeating the full cross validation procedure q times. Therefore, in addition to testing different classifier types, this study explores the effect of the following parameters: n (sample size, 20–500), k (number of cross-validation folds: 5, 10 and leave-one-out) and q (number of repetitions: 1, 5 and 20).

2.2. Statistical significance of classification using a binomial cumulative distribution

For a given number of classes c, the percent theoretical chance level of classification is given by $100/c$. For example, for a 4-class problem, the chance level is $100/4 = 25\%$. This threshold is based on the assumption of infinite sample size. In practice, the empirical chance level depends on the number of samples available. One way to address this limitation is to test for the statistical significance of the decoding accuracy. This can be done by assuming that the classification errors obey a binomial cumulative distribution, where for a total of n samples and c classes, the probability to predict the correct class at least z times by chance is given by:

$$P(z) = \sum_{i=z}^{n} \binom{n}{i} \times \left(\frac{1}{c}\right)^{i} \times \left(\frac{c-1}{c}\right)^{n-1}$$

Although neural signal classification studies predominantly evaluate decoding performance by how well the results depart from the theoretical chance level, several BCI studies have in addition, used the binomial cumulative distribution to derive statistical significance thresholds (e.g. Ang et al., 2010; Demandt et al., 2012; Pistohl et al., 2012; Waldert et al., 2007, 2008, 2012). In this study, we use the MATLAB (Mathworks Inc., MA, USA) function *binoinv* to compute the statistically significant threshold $St(\alpha) = binoinv(1-\alpha, n, 1/c) \times 100/n$, where α is the significance level given by $\alpha = z/n$ (i.e. the ratio of tolerated false positives z – i.e. number of observations correctly classified by chance with respect to all observations n). For instance, for a sample size of $n = 40$ and a 2-class classification problem ($c = 2$), computing the threshold for statistical significance of the decoding at $\alpha = 0.001$ using the above formulation yields 70.0%. In other words, at $n = 40$, any decoding percentage below 70% is not statistically significant (at $p < 0.001$), whereas if one relied on the theoretical threshold for two classes (i.e. 50%) a decoding accuracy of 67% might have been considered relevant. Table 1 provides the minimal thresholds as a function of selected sample sizes, class number and significance levels. Note that code for the calculation of these analytical significance levels is provided online (see Appendix A).

2.3. Statistical significance of classification using permutation tests

The statistical significance of decoding can also be assessed by non-parametric statistical methods, namely using permutation tests (Good, 2000; Nichols & Holmes, 2002). By randomly permuting the observations across classes and calculating classification accuracy at each permutation, it is possible to establish an empirical null distribution of classification accuracies on random observations. The tails of this distribution can then be used to determine significance boundaries for a given rate of tolerated false positives (i.e. correct classifications that occur by chance). For instance, if the original (without randomization) classification accuracy is higher than the 95 percentile of empirical performance distribution established by randomly permuting the data, then one can assert that the original classification is significant with $p < 0.05$. The advantage of this empirical approach is that it does not require particular assumption about statistical properties of the samples.

An intuitive illustration of this procedure would be as follows: one performs for example 99 random permutations of the labels (classes) in the data and computes the classification accuracy for each permutation. This provides an empirical distribution of 99 classification accuracy values. Now if the classification performance obtained with the original (unpermuted) data is higher than the maximum of the empirical distribution, one can conclude that it is significant with $\alpha = 0.01$.

Permutations test provide a useful empirical approach to deriving statistical significance of classifier performance (e.g. Golland and Fischl, 2003; Ojala and Garriga, 2010; Meyers and Kreiman, 2011). To demonstrate the utility to derive significance boundaries as a function of sample size and thus compare it to the use of the binomial formula. To this end, we used simulated random data with associated labels (as described in Section 2.1) and computed the classification performance (using LDA) for 10,000 permutations (randomly exchanging labels of the original observations). From this we derived the accuracy thresholds that correspond to the 99%, 99.9% and 99.99% percentile of the distribution (i.e. $p < 0.01$, $p < 0.001$, and $p < 0.0001$ respectively). This was done for each sample size value n (20–500), which allowed us to depict the evolution of the empirical significance boundaries as a function of sample size. Note that code for the calculation of permutation-based empirical significance levels is provided online (see Appendix A).

2.4. Classification of baseline data from real brain signals

Because real data does not necessarily have the same properties as those implemented in our random data simulations (zero-mean Gaussian white noise), we also calculated the correct classification rate (as a function of sample size) that is achieved when classifying real brain data that do not contain any true discrepancies. This was carried out for pre-stimulus or baseline recordings in MEG (4 subjects) and with intracranial EEG recordings (4 patients). The

Table 1
Look-up table for statistically significant classification performance. Minimal correct classification rate (%) to assert statistical significance (at a given p-value) as a function of sample size n and number of classes c. Threshold values are based on the binomial cumulative distribution function and are rounded to the first digit.

n	c											
	2-Classes				4-Classes				8-Classes			
	$p<0.05$	$p<0.01$	$p<10^{-3}$	$p<10^{-4}$	$p<0.05$	$p<0.01$	$p<10^{-3}$	$p<10^{-4}$	$p<0.05$	$p<0.01$	$p<10^{-3}$	$p<10^{-4}$
20	70.0%	75.0%	85.0%	90.0%	40.0%	50.0%	55.0%	65.0%	25.0%	30.0%	40.0%	45.0%
40	62.5%	67.5%	75.0%	77.5%	37.5%	42.5%	47.5%	52.5%	22.5%	25.0%	30.0%	35.0%
60	60.0%	65.0%	70.0%	73.3%	35.0%	38.3%	43.3%	46.7%	20.0%	23.3%	26.7%	30.0%
80	58.7%	62.5%	67.5%	70.0%	32.5%	36.2%	41.2%	43.7%	18.7%	21.2%	25.0%	27.5%
100	58.0%	62.0%	65.0%	68.0%	32.0%	35.0%	39.0%	42.0%	18.0%	21.0%	24.0%	26.0%
200	56.0%	58.0%	61.0%	63.0%	30.0%	32.5%	35.0%	37.0%	16.5%	18.0%	20.0%	22.0%
300	54.7%	56.7%	59.0%	60.7%	29.0%	31.0%	33.0%	34.7%	15.7%	17.0%	18.7%	20.0%
400	54.0%	55.7%	57.7%	59.2%	28.5%	30.0%	31.7%	33.2%	15.2%	16.5%	17.7%	19.0%
500	53.6%	55.2%	57.0%	58.2%	28.2%	29.6%	31.2%	32.4%	15.0%	16.0%	17.2%	18.2%

rationale here is that baseline (pre-stimulus) data is not expected to show any genuine discriminative brain patterns related to post-stimulus events, and as such, it is comparable to random background noise. Therefore, signal classification on these baseline periods should fail, and the accuracies that classifiers achieve can be taken as an empirical representation of chance-level decoding.

2.4.1. Illustrative data from MEG rest activity

We used illustrative data from 4 subjects scanned with a whole-head MEG system (151 sensors; VSM MedTech, BC, Canada) acquired at 1250 Hz sampling rate and with a band pass filter of 0–200 Hz. The participants provided written informed consent, and the experimental procedures were approved by the Institutional Review Board and by the National French Science Ethical Committee. The MEG data segments used for the purpose of the current analysis were extracted from the pre-stimulus baseline of a visuomotor MEG experiment (Jerbi et al., 2007b), and each trial was assigned one of 2 (or of 4) arbitrary labels for the 2-class (or 4-class) classification. Oscillatory alpha (8–12 Hz) power was computed using Hilbert transform and subsequently used as feature in an LDA-based classification procedure. We used 10-fold cross-validation and the whole procedure was repeated for increasing values of trial numbers (sample size n) ranging from 20 to 200 (in steps of 8).

2.4.2. Illustrative data from intracranial EEG baseline activity

We used illustrative data from 4 epilepsy patients stereotactically implanted with intracranial depth electrodes (0.8 mm diameter, 10–15 contact leads, DIXI Medical Instruments, Besançon, France). The intracerebral EEG (iEEG) recordings were conducted using a video-SEEG monitoring system (Micromed, Treviso, Italy), which allowed for the simultaneous recording from 128 depth-EEG electrode sites (More details of the routine SEEG acquisitions in Jerbi et al., 2009b). The data were bandpass filtered online from 0.1 to 200 Hz and sampled at 1024 Hz. The recordings were performed at the epilepsy department of the Grenoble University Hospital (headed by Dr. Philippe Kahane). All participants provided written informed consent, and the experimental procedures were approved by the Institutional Review Board and by the National French Science Ethical Committee.

The data segments used here were extracted from the pre-stimulus (baseline) of a standard motor task and each trial was associated with one of 2 (or of 4) labels for the 2-class (or 4-class) classification. The labels assigned to each pre-stimulus baseline trial were in fact the genuine post-stimulus events for the same trials (but no true discrimination can be expected prior to stimulus onset as the post-stim event could not be known or inferred during the pre-stimulus period). Broadband gamma (60–250 Hz) power was computed using Hilbert transform and subsequently used as feature in an LDA-based classification procedure. As for the MEG data, we used 10-fold cross-validation and the whole procedure was repeated for increasing values of trial numbers (sample size n) ranging from 20 to 200 (in steps of 8).

3. Results

3.1. Empirical evaluation of chance level decoding as a function of sample size

Fig. 1 shows the decoding accuracies obtained by conducting 10-fold cross validation on 100 randomly generated data sets. The decoding is depicted as a function of increasing sample size (from 24 to 500) and for the case of 2-class (left column) and 4-class (right column) classification. Although the theoretical chance levels for these configurations are 50% and 25% respectively, the results show how much the empirical decoding accuracies obtained with random data deviate from these probabilistic values.

The small sample size problem: as expected, the variance of the decoding accuracy across the 100 simulated random data sets is high, and the more so for small sample sizes. As illustrated in Fig. 1, while the decoding does converge toward the theoretical chance level as the sample size increases, the values achieved with small sample size ($n<100$) can be disturbingly high. For instance, the highlighted examples (solid black line) in panels (a) to (f) illustrate how decoding accuracies as high as 70% for 2-class classification (or 50% for 4-classes) can be observed even when conducting classification on subsets of randomly generated data with randomly associated labels.

The small sample issue is persistent and qualitatively similar across all classifiers used. The first three rows of Fig. 1 show the results obtained with LDA, NB and SVM (with an RBF kernel). Panels (g) and (h) of Fig. 1 show that cross-validation results in all three classifiers have comparable deviation across the 100 simulated data sets. The variance of cross-validation over the 100 random data sets is high for small sample sizes (<200 observations) and drops off with increasing sample size.

3.2. Tweaking cross-validation parameters does not solve the small sample problem

It might be tempting to think that changing the cross-validation parameters might be a way to get around the small sample problem illustrated here. To address this we evaluated the impact of varying (a) the number of cross-validation folds, and (b) the number of repetitions of the cross-validation, on the reported deviation of the cross-validation results (cf. Fig. 1g and h) across the 100 data sets and all sample sizes. The results in Fig. 2(a–c) show that applying 5- and 10-fold cross-validation to the random data yielded substantially the same results, and that leave-one-out (LOO) cross-validation actually provided worse results (i.e.

Fig. 1. Classifier decoding rates as a function of sample size when applied to random data sets using 10-fold cross-validation. (a) Two-class LDA classification rate (%) as a function of sample size (empirical results increasingly deviate from the 50% chance-level as the sample size gets smaller). The backline line shows the evolution of cross-validation results for one specific data set out of the 100 depicted in multiple colors. (b) Same as panel (a) but using 4-class classification, i.e. at each sample size *n*, the data is split into 4 virtual classes instead of two, (c and d) Same as (a and b) but for a Naïve Bayesian classifier. (e and f) Same as (a and b) but for an SVM classifier using an RBF kernel. (g) Evolution of cross-validation standard deviation across the 100 data sets for each of the three classifiers for 2-class decoding. (h) Same as panel (g) but for 4-class decoding.

Please cite this article in press as: Combrisson E, Jerbi K. Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. J Neurosci Methods (2015), http://dx.doi.org/10.1016/j.jneumeth.2015.01.010

Fig. 2. Effect of cross-validation parameters on the variability of 2-class decoding performance computed across 100 sets of random data. (a–c) Effect of the number of folds (k): drop in cross-validation variance as sample size n increases, shown for $k = 5$, $k = 10$ (default), and $k = n$ (i.e. leave-one-out) and for all three classifiers LDA (panel a), NB (panel b) and SVM (panel c). (d–f) Effect of cross-validation repetition number: drop in cross-validation variance as sample size n increases, shown for repetition values $q = 1$ (default), $q = 5$, and $q = 20$ and for all three classifiers LDA (panel d), NB (panel e) and SVM (panel f). Note that the strong deviation from 50% chance-level for small sample sizes is persistent across all panels, and appears to be worst for LOO cross-validation with LDA.

higher variance). Moreover, repeating the cross-validation procedure (whether 5 or 20 times) achieved a negligible reduction of variance (Fig. 2d and e). Overall, these observations indicate that neither changing the number of folds nor to the number of overall repetitions has an impact on the variance of decoding accuracy (i.e. cross-validation results) across the 100 sets of Gaussian white noise.

3.3. *Estimating statistical significance of decoding accuracy: binomial formula and permutation tests*

Panels (a) and (b) in Fig. 3 show the evolution of the minimal statistically significant decoding rate as a function of sample size (respectively for 2- and 4-classes) using the binomial cumulative distribution (described in Section 2.2). The plots depicted for three distinct significance levels (10^{-2}, 10^{-3} and 10^{-4}) all show that the minimal correct decoding rate that is required in order to assert significance, decreases as the number of samples increases. Given small sample sizes (e.g. below 100 observations), to be statistically significant, the decoding accuracy must be substantially higher than the probabilistic chance level. For example, for 40 observations, a 2-class decoding is statistically significant (at $p < 0.001$) only if it exceeds the threshold of 75%. Note that for sample sizes as high as 500 observations, statistical significance still requires correct decoding higher than 55% (at $p < 0.01$), i.e. at least 5% above the theoretical chance level. A more comprehensive overview of the statistical decoding thresholds (wider ranges of p-values and of class number) computed for selected sample sizes (20–500), is provided in Table 1.

Panels (c) and (d) in Fig. 3 depict not only the evolution of the decoding boundaries for 2-class and 4-class decoding, using the binomial formula but also using the permutation test approach (see Section 2.3). Interestingly, the boundaries (for each level of admitted false positives) using both methods are reasonably close. The boundaries obtained with permutations show a slight tendency to be more restrictive than the binomial formula. While this is a little more apparent for small values of n, the difference between the two methods rapidly vanishes as n increases.

3.4. *MEG and iEEG baseline data reveal erroneously high decoding results*

Fig. 4 depicts the results of the empirical estimation of *de facto* chance-level decoding in illustrative MEG and iEEG data segments taken during pre-stimulus baseline periods (where no decoding is theoretically expected). Similarly to our findings using random data simulations (Fig. 1 a–f), the baseline MEG and iEEG data trials also led to decoding rates that strongly departed from the theoretical chance levels of 50% for 2-class classification and 25% for 4-class classification. Also in line with the results of the simulated data, the effects observed here were again highest for small sample sizes and dropped off slowly with increasing n. Note that the results in Fig. 4 show consistent performances across the 4 subjects at each value of n (with MEG and with iEEG). Finally, the superimposed gray curves (which depict the significance boundary given by the binomial formula as a function of sample size) nicely follow the trend of the % correct classification rate, and also illustrate cases of tolerated false positives for a given alpha.

4. Discussion

The current study has two primary take-home messages. The first is emphasizing the importance of watching out for a potential caveat that may arise when using departure from the theoretical chance-level as evidence for meaningful decoding. By launching various classifiers on normally distributed random data (Gaussian white noise), we demonstrate and quantify to which extent small samples lead to decoding accuracies that overshoot the chance-level merely by chance. This observation follows from the fact that

Fig. 3. Estimation of the statistical significance thresholds for 2- and 4-class classification as a function of sample size (assuming prediction errors are binomially distributed). Panels (a) and (b) show the evolution of the minimal statistically significant decoding rate as a function of sample size (respectively for 2 and 4 classes) using the binomial cumulative distribution (see Section 2.2). The plots were derived for significance levels 10^{-2}, 10^{-3} and 10^{-4}. As an example: panel (a) indicates that given a total of 100 data samples, a 2-class decoding result can only be considered statistically significant (at $p < 0.001$) if it exceeds 65%. This minimal value drops to 59% for 300 samples, but rises up to 75% if only 40 data points are available (See Table 1). Panels (c) and (d) show the same statistically significant decoding rate as a function of sample size (respectively for 2 and 4 classes) but now using both the binomial cumulative distribution (continuous lines) and the data-driven permutation-based approach (dashed lines) applied to the simulated random data (see Section 2.3 for details).

small samples are a bad approximation of true randomness and that as a result, the level $100/c$ (where c is the number of classes) is a purely theoretical chance-level that only holds for infinite sample sizes and that is particularly violated for small sample sizes. This basic fact is often overlooked in the neuronal decoding literature, where it is sometimes tempting to interpret for instance a 65% decoding accuracy in a 2-class classification as reflecting true neuronal decoding, without taking sample size into account. We have shown here that such levels of classification can be achieved with small samples of randomly generated data. This issue is not problematic for huge data samples, however, in data obtained from brain signal recordings in humans (such EEG or MEG), sample size can often be small. The effect of small samples on the reliability of probabilistic thresholds is therefore of particular importance in neural decoding and brain–computer interface studies. This effect is possibly even more critical when attempting to decode neuronal signals acquired using intracranial recordings (electrocorticography or stereoactic-EEG) and in clinical BCI applications where even less data samples might be available.

Furthermore, our exploration of the effects of classifier type (LDA, NB and SVM), cross-validation partition (number of folds) and cross-validation repetition number (up to 20), indicates that none of these parameters has a noticeable impact on the variance of the classification when applied to random data. The small sample size problem cannot be circumvented by tweaking these parameters and even for larger sample sizes of white noise any reduction in classification variance remains negligible. Note that the explored parameters and classifier comparisons performed here only address the variance and bias of the techniques when applied to normally distributed random data, reviews and comparisons of classifiers can be found elsewhere (e.g. Lotte et al., 2007).

Fig. 4. Experimental assessment of chance-level classification accuracy in baseline (pre-stimulus) MEG and intracranial EEG data. (a) Two-class LDA classification rate (%) of MEG baseline data (alpha power features) as a function of sample size (illustrative data in 4 participants *MEG S1–S4*). The gray lines show the evolution of statistical significance boundaries computed with the binomial formula. Points lying above the gray lines thus represent false positives (type I errors) (b) Same as panel (a) but using 4-class classification, (c and d) Same as (a and b) but using baseline data (gamma power features) from intracranial EEG recordings (illustrative data in 4 epilepsy patients *iEEG S1–S4*).

Ten-fold cross-validation, which we used here as default, has been shown to be a reasonable choice providing low variance (Kohavi, 1995; Martin and Hirschberg, 1996a). Nevertheless, we also explored 5-fold and LOO cross-validation, alongside repetition number (Fig. 2). We found that none of these parameters could help reduce the cross-validation variance for low sample sizes. What is more, leave-one-out cross-validation showed even higher variability (in particular when using LDA), which is in agreement with previous reports suggesting that, despite its low bias, its high variance leads to unreliable estimates (Efron, 1983). Note that estimating the variance of cross-validation results across its k folds is generally problematic. Naive estimators that do not take into account error correlations due to the overlap between training and test sets (across the cross-validation folds) can severely underestimate variance (Bengio and Grandvalet, 2004). The cross-validation variances reported here were computed across the 100 independent data sets of Gaussian white noise.

The second take-home message from our study is an important reminder that one way to overcome this limitation is to seek statistically significant thresholds on decoding accuracy, rather than relying solely on the theoretical chance-level to claim successful decoding. This has been demonstrated here using a sample-size dependent threshold computation derived from the binomial cumulative distribution function. The underlying assumption that the number of errors is binomially distributed is commonly used in statistical learning (Kohavi, 1995; Breiman et al., 1984) but the statistical bounds it provides are unfortunately rarely exploited in brain signal decoding studies (e.g. Quiroga and Panzeri, 2009; Müller-Putz et al., 2008; Ang et al., 2010; Arvaneh et al., 2013; Demandt et al., 2012; Galan et al., 2014; Lampe et al., 2014; Pistohl

et al., 2012; Waldert et al., 2008, 2007, 2012). Kohavi (1995) provides a proof that *k*-fold cross-validation is binomial if the classifier induction method is stable under cross-validation. Note also that the validity of the assumption that prediction errors are binomially distributed has also been demonstrated for the specific case of 10-fold cross-validation with small samples (Martin and Hirschberg, 1996b). The latter study also emphasizes that the textbook formula based on the normal approximation to the binomial is not a good approximation to the confidence interval of an error rate estimate for small samples.

In addition to the binomial formula, we have also demonstrated the use of permutation tests as an alternative method to derive statistical significance boundaries for classifier performance as a function of sample size (Fig. 3c and d). Permutation tests provide a reliable and data-driven approach to the problem and has been proposed and used in numerous previous studies (e.g. Golland and Fischl, 2003; Ojala and Garriga, 2010; Meyers and Kreiman, 2011). Our analysis shows how, via multiple random shuffling of the data (or class labels), permutation tests can provide an estimate of sample-size dependent chance-level decoding accuracy. These empirical chance levels need to be exceeded in order to assert significance of a classification for a given rate of tolerated false positives. When applied to random noise signals, we found that the significance boundaries derived using permutations are reasonably close to those obtained using the binomial formula. Deciding which of the two approaches is more convenient when applied to real brain signals will likely depend on the data at hand. Permutation tests do not make any assumptions about the distribution of the data and provide a data-driven approach; however they also come with the burden of high computational cost, which dramatically increases with sample size, and with the level of statistical significance required.

Meyers and Kreiman (2011) note that deriving significance thresholds via the binomial formula as discussed here and elsewhere (e.g. Quiroga and Panzeri, 2009) comes with theoretical limitations that one should keep in mind, in particular, when combined with cross-validation; its application to mean performance over all folds violates the assumption of data point independence and leads to *p*-values that are too small. From a practical perspective, the impact of this theoretical limitation is likely to depend on the data at hand and on the selected cross-validation parameters. Simulations show that cross-validation parameters (number of cross-valdiation folds and repetitions) have an impact on the cumulative distribution function of classification accuracies (e.g. Noirhomme et al., 2014). As a result, cross-validation parameters, alongside classifier type and feature space, collectively lead to deviations from a binomial cumulative distribution. These deviations can be significant for small sample sizes (e.g. $N < 100$), which would advocate against using the binomial formula for statistical assessments under such circumstances (Noirhomme et al., 2014). In contrast, permutation tests being inherently data-driven, do take cross-validation parameters into account. As far as the Gaussian white noise simulated in the current study is concerned, permutation tests and the binomial formula appear to provide reasonably similar significance boundaries. Comparing the output of the binomial formula and (the more time consuming) permutation test, on at least a portion of the data, could be a pragmatic way to decide on whether the former provides a suitable and fast *approximation* of the latter.

Moreover, our analysis of decoding accuracy using real brain signals (with random labeling) is in line with our simulation results. This is a reassuring finding, as the latter were based on zero-mean Gaussian white noise while the former were based on power features (alpha and gamma-bands) derived from real brain data. The baseline-period MEG and iEEG data suggest that the binomial formula provides a reasonable estimation of chance-level decoding in these data sets. As a general rule, whenever possible, it is highly recommended to use baseline data as a recording in which no task-dependent encoding occurs and thus within chance-level decoding is expected. Comparisons with pre-stimulus (baseline) decoding performances should be used as an additional sanity check whenever such data is available (Meyers and Kreiman, 2011).

An alternative framework that can be applied to measure and compare classifier performance, is the use of receiver operating characteristic (ROC) analysis and in particular the area under the ROC curve (AUC) (Ling et al., 2003; Huang and Ling, 2005; Bradley, 1997). It has also been shown that calculating the probability density function (pdf) for each point on a ROC curve for any given sample size can be used to produce confidence intervals for ROC curves that are valid for small sample sizes (Tilbury et al., 2000). Adaptations of this method might be particularly suited to assessing classifier performance in BCI research (Hamadicharef, 2010). Other solutions that have been proposed to tackle the small sample size problem include frameworks that combine cross-validation with bootstrapping (e.g. Fu et al., 2005) and the use of class-dependent PCA in conjunction with linear discriminant feature extraction (Das and Nenadic, 2009). It is noteworthy that a few authors have even suggested that classification studies should be based primarily on effect size estimation with confidence intervals, rather than on significance tests and *p*-values (Berrar and Lozano, 2013).

In summary, the notion of statistical significance for decoding rates (or prediction error) and the small sample size problem have been tackled in the field of statistical learning for a long time (e.g. Raudys and Jain, 1991). However, these notions have not been sufficiently acknowledged in the relatively recent surge in application of machine learning methods in neuroscience. In the worse cases, this can unfortunately lead to erroneous interpretation of decoding results. Beyond its importance for brain–computer interface research specifically, signal classification is also increasingly used in neuroscience with the broader aim of elucidating the functional role of specific neuronal features (i.e. unraveling neuronal encoding by investigating single-trial neuronal decoding). Incidentally, this is where researchers are likely to be tempted to consider low (but above chance-level) decoding accuracies (e.g. 68% in a two-class classification) as being relevant. The use of confidence intervals and robust estimation of statistical significance is of particular importance in such studies, and even more so in cases with low trial numbers (e.g. below 150 observations). Machine learning and cross-validation accuracy in multi-class decoding may therefore not be thought of as a less-strict approach that can circumvent traditional rigorous statistical comparisons of data from multiple experimental conditions. Finally, whether signal classification is used in a BCI context *stricto sensu* or within a framework to conduct basic neuroscience analysis, we highly recommend systematically reporting the decoding accuracy as well as its statistical significance. We hope that the simulation results, statistical approaches and practical recommendations discussed here will be helpful in illustrating the problem and providing ways of tackling it.

Acknowledgements

Etienne Combrisson is currently supported by a Ph.D. Scholarship awarded by the Ecole Doctorale Inter-Disciplinaire Sciences-Santé (EDISS), Lyon, France. This work was partly performed within the framework of the LABEX CORTEX (ANR-11-LABX-0042) of Université de Lyon, within the program ANR-11-IDEX-0007. This research was undertaken, in part, thanks to funding from the Canada Research Chairs program. The authors are grateful for the collaboration with the research and clinical staff of the Magnetoencephalography (MEG) center at the Pitié-Salpêtrière Hospital

in Paris and the University Hospital in Grenoble (Dr. Philippe Kahane).

Appendix A.

A. Software availability: The MATLAB scripts and functions that were developed and used in this study have been made available online for the community. The provided code can be used to generate, label and classify random data. It also provides routines to compute and plot, as a function of sample size, (a) analytical chance levels via the binomial formula as well as (b) empirical chance levels via permutation tests. We hope that this set of tools will help students and researchers replicate and extend our analyses. The code can be downloaded from Mathwork's File Exchange platform at the following URL: http://www.mathworks.fr/matlabcentral/fileexchange/48274-random-data-classification

References

Ahn M, Ahn S, Hong JH, Cho H, Kim K, Kim BS, et al. Gamma band activity associated with BCI performance: simultaneous MEG/EEG study. Front Hum Neurosci 2013:7, Available from: http://journal.frontiersin.org/Journal/10.3389/fnhum.2013.00848/full.
Aloise F, Schettini F, Aricò P, Salinari S, Babiloni F, Cincotti F. A comparison of classification techniques for a gaze-independent P300-based brain–computer interface. J Neural Eng 2012;9(4):045012.
Ang KK, Guan C, Sui Geok Chua K, Ang BT, Kuah C, Wang C, et al. Clinical study of neurorehabilitation in stroke using EEG-based motor imagery brain-computer interface with robotic feedback. IEEE 2010:5549–52 [cited 2014 Jul 4]. Available from: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5626782.
Arvaneh M, Guan C, Ang KK, Quek C. EEG data space adaptation to reduce intersession nonstationarity in brain–computer interface. Neural Comput 2013;25(May (8)):2146–71.
Babiloni F, Cincotti F, Lazzarini L, Millán J, Mouriño J, Varsta M, et al. Linear classification of low-resolution EEG patterns produced by imagined hand movements. IEEE Trans Rehabil Eng 2000;8(June (2)):186–8.
Ball T, Schulze-Bonhage A, Aertsen A, Mehring C. Differential representation of arm movement direction in relation to cortical anatomy and function. J Neural Eng 2009;6(1):016006.
Bengio Y, Grandvalet Y. No unbiased estimator of the variance of *k*-fold cross-validation. J Mach Learn Res 2004;5:1089–105.
Berrar D, Lozano JA. Significance tests or confidence intervals: which are preferable for the comparison of classifiers? J Exp Theor Artif Intell 2013;25(June (2)):189–206.
Besserve M, Jerbi K, Laurent F, Baillet S, Martinerie J, Garnero L, et al. Classification methods for ongoing EEG and MEG signals. Biol Res 2007;40(4):415–37.
Bleichner MG, Jansma JM, Sellmeijer J, Raemaekers M, Ramsey NF. Give me a sign: decoding complex coordinated hand movements using high-field fMRI. Brain Topogr 2014;27(March (2)):248–57.
Bode S, Haynes J-D. Decoding sequential stages of task preparation in the human brain. Neuroimage 2009;45(April (2)):606–13.
Boser BE, Guyon IM, Vapnik VN. A training algorithm for optimal margin classifiers. ACM 1992:144–52 [cité 25.07.14].
Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognit 1997;30(July (7)):1145–59.
Breiman L, Friedman JH, Olshen R, Stone CJ. Classification and regression trees; 1984.
Burges CJ. A tutorial on support vector machines for pattern recognition. Data Min Knowl Discov 1998;2(2):121–67.
Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20(September (3)):273–97.
Das K, Nenadic Z. An efficient discriminant-based solution for small sample size problem. Pattern Recognit 2009;42(May (5)):857–66.
Demandt E, Mehring C, Vogt K, Schulze-Bonhage A, Aertsen A, Ball T. Reaching movement onset- and end-related characteristics of eeg spectral power modulations. Front Neurosci 2012:6, http://www.frontiersin.org/Journal/10.3389/fnins.2012.00065/full.
Derix J, Iljina O, Schulze-Bonhage A, Aertsen A, Ball T. "Doctor" or "darling"? Decoding the communication partner from ECoG of the anterior temporal lobe during non-experimental, real-life social interaction. Front Hum Neurosci 2012:6, http://www.frontiersin.org/Journal/10.3389/fnhum.2012.00251/full.
Efron B. Estimating the error rate of a prediction rule: improvement on cross-validation. J Am Stat Assoc 1983;78(June (382)):316–31.
Felton EA, Wilson JA, Williams JC, Garell PC. Electrocorticographically controlled brain–computer interfaces using motor and sensory imagery in patients with temporary subdural electrode implants. Report of four cases. J Neurosurg 2007;106(March (3)):495–500.
Fisher RA. The use of multiple measurements in taxonomic problems. Ann Eugen 1936;7(September (2)):179–88.
Fu WJ, Carroll RJ, Wang S. Estimating misclassification error with small samples via bootstrap cross-validation. Bioinformatics 2005;21(May (9)):1979–86.
Fukunaga K. Introduction to statistical pattern recognition. 2nd ed. Boston: Academic Press; 1990.
Galan F, Baker MR, Alter K, Baker SN. Missing kinaesthesia challenges precise naturalistic cortical prosthetic control, May. Report no.: 004861; 2014, http://biorxiv.org/lookup/doi/10.1101/004861.
Golland P, Fischl B. Permutation tests for classification: towards statistical significance in image-based studies. In: Proc. IPMI: international conference on information processing and medical imaging, LNCS, vol. 2732; 2003. p. 330–41.
Good PI. Permutation tests: a practical guide to resampling methods for testing hypotheses. 2nd ed. New York: Springer; 2000.
Hamadicharef B. AUC confidence bounds for performance evaluation of Brain-Computer Interface. In: IEEE 3rd International (Volume:5) Conference on Biomedical Engineering and Informatics (BMEI); 2010. p. 1988–91, Available from: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5639671.
Hamamé CM, Vidal JR, Ossandón T, Jerbi K, Dalal SS, Minotti L, et al. Reading the mind's eye: online detection of visuo-spatial working memory and visual imagery in the inferior temporal lobe. NeuroImage 2012;59(January (1)):872–9.
Haynes J-D, Sakai K, Rees G, Gilbert S, Frith C, Passingham RE. Reading hidden intentions in the human brain. Curr Biol 2007;17(February (4)):323–8.
Hill NJ, Lal TN, Schröder M, Hinterberger T, Wilhelm B, Nijboer F, et al. Classifying EEG and ECoG signals without subject training for fast BCI implementation: comparison of nonparalyzed and completely paralyzed subjects. IEEE Trans Neural Syst Rehabil Eng 2006;14(June (2)):183–6.
Hosseini SMH, Mano Y, Rostami M, Takahashi M, Sugiura M, Kawashima R. Decoding what one likes or dislikes from single-trial fNIRS measurements. NeuroReport 2011;22(April (6)):269–73.
Huang J, Ling CX. Using AUC and accuracy in evaluating learning algorithms. IEEE Trans Knowl Data Eng 2005;17(3):299–310.
Jerbi K, Bertrand O, Schoendorff B, Hoffmann D, Minotti L, Kahane P, et al. Online detection of gamma oscillations in ongoing intracerebral recordings: From functional mapping to brain computer interfaces. In: Noninvasive Funct Source Imaging Brain Heart Int Conf Funct Biomed Imaging 2007 NFSI-ICFBI 2007 Jt Meet 6th Int Symp On. IEEE; 2007a. p. 330–3, http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4387767.
Jerbi K, Lachaux J-P, Karim N, Pantazis D, Leahy RM, Garnero L, et al. Coherent neural representation of hand speed in humans revealed by MEG imaging. Proc Natl Acad Sci 2007b;104(18):7676–81.
Jerbi K, Freyermuth S, Minotti L, Kahane P, Berthoz A, Lachaux J. Watching brain TV and playing brain ball International review of neurobiology. In: Brain machine interfaces for space applications: enhancing astronaut capabilities. San Diego: Elsevier Academic Press; 2009a. p. 159–68 [chapter 12]. http://linkinghub.elsevier.com/retrieve/pii/S0074774209860121.
Jerbi K, Ossandón T, Hamamé CM, Senova S, Dalal SS, Jung J, et al. Task-related gamma-band dynamics from an intracerebral perspective: review and implications for surface EEG and MEG. Hum Brain Mapp 2009b;30(June (6)):1758–71.
Jerbi K, Vidal JR, Mattout J, Maby E, Lecaignard F, Ossandon T, et al. Inferring hand movement kinematics from MEG, EEG and intracranial EEG: from brain-machine interfaces to motor rehabilitation. IRBM 2011;32(February (1)):8–18.
Jerbi K, Combrisson E, Dalal SS, Vidal JR, Hamame CM, Bertrand O, et al. Decoding cognitive states and motor intentions from intracranial EEG: how promising is high-frequency brain activity for brain–machine interfaces? Epilepsy Behav 2013;28(2):283–302.
Kayikcioglu T, Aydemir O. A polynomial fitting and k-NN based approach for improving classification of motor imagery BCI data. Pattern Recognit Lett 2010;31(August (11)):1207–15.
Kellis S, Miller K, Thomson K, Brown R, House P, Greger B. Decoding spoken words using local field potentials recorded from the cortical surface. J Neural Eng 2010;7(October (5)):056007.
Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection; 1995. p. 1137–45, http://frostiebek.free.fr/docs/Machine%20Learning/validation-1.pdf.
Korczyn AD, Schachter SC, Brodie MJ, Dalal SS, Engel J, Guekht A, et al. Epilepsy, cognition, and neuropsychiatry (Epilepsy, Brain, and Mind, part 2). Epilepsy Behav 2013;28(2):283–302.
Krusienski DJ, Wolpaw JR. Brain-computer interface research at the wadsworth center developments in noninvasive communication and control. Int Rev Neurobiol 2009;86:147–57.
Lachaux J-P, Jerbi K, Bertrand O, Minotti L, Hoffmann D, Schoendorff B, et al. A Blueprint for Real-Time Functional Mapping via Human Intracranial Recordings. PLoS ONE 2007a;2(October (10)):e1094.
Lachaux J-P, Jerbi K, Bertrand O, Minotti L, Hoffmann D, Schoendorff B, et al. BrainTV: a novel approach for online mapping of human brain functions. Biol Res 2007b;40(January (4)):401–13.
Lampe T, Fiederer LDJ, Voelker M, Knorr A, Riedmiller M, Ball T. A brain–computer interface for high-level remote control of an autonomous, reinforcement-learning-based robotic system for reaching and grasping. In: Proceedings of the 19th international conference on intelligent user interfaces. New York, NY, USA: ACM; 2014. p. 83–8, http://dx.doi.org/10.1145/2557500.2557533.
Leuthardt EC, Schalk G, Wolpaw JR, Ojemann JG, Moran DW. A brain–computer interface using electrocorticographic signals in humans. J Neural Eng 2004;1(June (2)):63.
Leuthardt EC, Miller KJ, Schalk G, Rao RPN, Ojemann JG. Electrocorticography-based brain computer interface–the Seattle experience. IEEE Trans Neural Syst Rehabil Eng 2006;14(June (2)):194–8.

Ling CX, Huang J, Zhang H. AUC: a statistically consistent and more discriminating measure than accuracy; 2003. p. 519–24, http://arion.csd.uwo.ca/faculty/ling/papers/ijcai03.pdf.
Lotte F, Congedo M, Lécuyer A, Lamarche F, Arnaldi B. A review of classification algorithms for EEG-based brain–computer interfaces. J Neural Eng [Internet] 2007 [cited 2012 Oct 3];4. Available from: http://hal.archives-ouvertes.fr/docs/00/13/49/50/PDF/article.pdf.
Martin JK, Hirschberg DS. Small sample statistics for classification error rates I: error rate measurements. Irvine: Information and Computer Science, University of California; 1996a, http://www.ics.uci.edu/~dan/pubs/TR96-21.pdf.
Martin JK, Hirschberg DS. Small sample statistics for classification error rates II: confidence intervals and significance tests [Internet]. Information and Computer Science. Irvine: University of California; 1996b, Disponible sur: http://www.ics.uci.edu/~dan/pubs/TR96-22.pdf.
Mehring C, Nawrot MP, de Oliveira SC, Vaadia E, Schulze-Bonhage A, Aertsen A, et al. Comparing information about arm movement direction in single channels of local and epicortical field potentials from monkey and human motor cortex. J Physiol – Paris 2004;98(July (4–6)):498–506.
Meyers EM, Kreiman G. Tutorial on pattern classification in cell recordings. In: Kriegeskorte N, Kreiman G, editors. Understanding visual population codes. Boston: MIT Press; 2011.
Momennejad I, Haynes J-D. Human anterior prefrontal cortex encodes the "what" and "when" of future intentions. Neuroimage 2012;61(May (1)):139–48.
Morash V, Bai O, Furlani S, Lin P, Hallett M. Classifying EEG signals preceding right hand, left hand, tongue, and right foot movements and motor imageries. Clin Neurophysiol 2008;119(November (11)):2570–8.
Müller-Putz GR, Scherer R, Brunner C, Leeb R, Pfurtscheller G. Better than random? A closer look on BCI results. Int J Bioelectromagn 2008;10(1):52–5.
Neuper C, Scherer R, Reiner M, Pfurtscheller G. Imagery of motor actions: differential effects of kinesthetic and visual-motor mode of imagery in single-trial EEG. Brain Res Cogn Brain Res 2005;25(December (3)):668–77.
Nichols TE, Holmes AP. Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum Brain Mapp 2002;15(1):1–25.
Noirhomme Q, Lesenfants D, Gomez F, Soddu A, Schrouff J, Garraux G, et al. Biased binomial assessment of cross-validated estimation of classification accuracies illustrated in diagnosis predictions. NeuroImage: Clin 2014;4:687–94.
Ojala M, Garriga GC. Permutation tests for studying classifier performance. J Mach Learn Res 2010;11(June):1833–63.
Pistohl T, Schulze-Bonhage A, Aertsen A, Mehring C, Ball T. Decoding natural grasp types from human ECoG. NeuroImage 2012;59(January (1)):248–60.
Quiroga RQ, Panzeri S. Extracting information from neuronal populations: information theory and decoding approaches. Nat Rev Neurosci 2009;10:173.
Raudys SJ, Jain AK. Small sample size effects in statistical pattern recognition: recommendations for practitioners. IEEE Trans Pattern Anal Mach Intell 1991;13(March (3)):252–64.
Schalk G, Miller KJ, Anderson NR, Wilson JA, Smyth MD, Ojemann JG, et al. Two-dimensional movement control using electrocorticographic signals in humans. J Neural Eng 2008;5(March (1)):75–84.
Sitaram R, Lee S, Ruiz S, Rana M, Veit R, Birbaumer N. Real-time support vector classification and feedback of multiple emotional brain states. NeuroImage 2011;56(May (2)):753–65.
Tilbury JB, Van Eetvelt WJ, Garibaldi JM, Curnsw JSH, Ifeachor EC. Receiver operating characteristic analysis for intelligent medical systems-a new approach for finding confidence intervals. IEEE Trans Biomed Eng 2000;47(7):952–63.
Toppi J, Risetti M, Quitadamo LR, Petti M, Bianchi L, Salinari S, et al. Investigating the effects of a sensorimotor rhythm-based BCI training on the cortical activity elicited by mental imagery. J Neural Eng 2014;11(June (3)):035010.
Vapnik V. The nature of statistical learning theory. New York: Springer Science & Business Media; 1995.
Waldert S, Braun C, Preissl H, Birbaumer N, Aertsen A, Mehring C. Decoding performance for hand movements: EEG vs. MEG. IEEE 2007:5346–8.
Waldert S, Preissl H, Demandt E, Braun C, Birbaumer N, Aertsen A, et al. Hand movement direction decoded from MEG and EEG. J Neurosci 2008;28(January (4)):1000–8.
Waldert S, Tüshaus L, Kaller CP, Aertsen A, Mehring C. fNIRS exhibits weak tuning to hand movement direction. PLoS ONE 2012;7(November (11)):e49266.
Wang W, Sudre GP, Xu Y, Kass RE, Collinger JL, Degenhart AD, et al. Decoding and cortical source localization for intended movement direction with MEG. J Neurophysiol 2010;104(November (5)):2451–61.
Wieland M, Pittore M. Performance Evaluation of Machine Learning Algorithms for Urban Pattern Recognition from Multi-spectral Satellite Images. Remote Sens 2014;6(March (4)):2912–39.

Please cite this article in press as: Combrisson E, Jerbi K. Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. J Neurosci Methods (2015), http://dx.doi.org/10.1016/j.jneumeth.2015.01.010

Troisième partie

Étude 2 : encodage de l'intention et de l'exécution motrice

Introduction

blabla Intro

Blabla étude

Title

From intentions to actions: Neural oscillations encode motor processes through phase, amplitude and phase-amplitude coupling

Authors

Etienne Combrisson[1,2,3], Marcela Perrone-Bertolotti[4], Juan LP Soto[5], Golnoush Alamian[1], Philippe Kahane[4], J-P Lachaux[3], Aymeric Guillot[2], Karim Jerbi[1]

Affiliations

[1] Psychology Department, University of Montreal, QC, Canada
[2] Center of Research and Innovation in Sport, Mental Processes and Motor Performance, University Claude Bernard Lyon I, University of Lyon, Villeurbanne, France
[3] Lyon Neuroscience Research Center, Brain Dynamics and Cognition, INSERM U1028, UMR 5292, Lyon University
[4] Neurology Dept and GIN U836 INSERM-UJF-CEA, Grenoble University Hospital, Grenoble, France
[5] Telecommunications and Control Engineering Department, University of Sao Paulo, Sao Paulo, Brazil

Corresponding author

Etienne Combrisson
Address:
Psychology Department
University of Montreal
Pavillon Marie-Victorin
90, avenue Vincent d'Indy, Quebec, Canada
E-mail address: e.combrisson@gmail.com

List of potential referees (For cover letter):

Andrea Brovelli
Julien Voisin
Nikolai Axmacher
Michel Le van Quyen
Lauri Parkkonnen
Nathan Crone
Andrea Greggoriu
Joachim Gross
Paolo Belardinelli
Hyojin Park
Felix Darvas
Tonio Ball

ABSTRACT

Goal-directed motor behavior is associated with changes in patterns of rhythmic neuronal activity across widely distributed brain areas. In particular, movement initiation and execution is mediated by patterns of synchronization and desynchronization that occur concurrently across distinct frequency bands and in multiple motor cortical areas. To date, motor-related local oscillatory modulations are predominantly examined by quantifying increases or suppressions in spectral power. However, beyond signal power, further spectral properties such as phase and phase-amplitude coupling (PAC) have also been shown to carry information with regards to the oscillatory dynamics underlying motor processes. The distinct functional roles of phase, amplitude and PAC across planning and execution of goal-directed motor behavior remain largely elusive. Here, we address this question with unprecedented resolution thanks to rare multi-site intracerebral EEG recordings in human subjects performing a delayed motor task. To compare the roles of phase, amplitude and PAC in motor planning and execution, task-related modulations of electrophysiological brain signals were monitored at movement execution and during the delay period where motor intention is present but execution is withheld. In particular, we used a machine-learning framework to identify the key contributions of the various neuronal responses. We found a high degree of overlap between the brain network patterns observed during planning and execution. Prominent amplitude increases in the delta (2-4 Hz) and high gamma (60-200 Hz) bands were observed during planning and execution. In contrast, motor alpha (8-13 Hz) and beta (13-30 Hz) power were suppressed during execution, but enhanced during the delay period. Interestingly, single-trial classification revealed that low-frequency phase information was the most discriminant feature in dissociating action from intention. Additionally, despite providing weaker decoding, PAC features led to statistically significant classification of motor states, particularly in anterior cingulate cortex and premotor brain areas. These results advance our understanding of the distinct and partly overlapping involvement of phase, amplitude and PAC in the neuronal mechanisms underlying motor intentions and executions.

Keywords

Neural oscillations, motor intention, motor planning, phase, phase-amplitude coupling, intracranial EEG

1 Introduction

The simplest motor act, such as stretching out your arm to grasp a cup of coffee, is mediated by a rich and complex chain of neuronal processes. What in essence may seem as the realization of a straightforward motor command is in fact carried out by a cascade of events ranging from action selection and planning to motor execution and monitoring. The neural mechanisms that mediate the transformation of intentions into actions have been the subject of a thriving body of research for decades (Ariani et al., 2015; Brovelli et al., 2005; Desmurget and Sirigu, 2009; Jeannerod, 1994; Kalaska, 2009; Lau, 2004; Paus, 2001; Schwartz, 2016; Snyder et al., 1997). However, because the neuronal processes at play can be observed at various spatial scales and with different recording techniques, parallel streams of research have given rise to a rich but fragmented understanding of the local and large-scale integrative electrophysiological mechanisms involved in motor control.

Both human and non-human primate research provides solid evidence that goal-directed motor behavior is associated with changes in patterns of rhythmic neuronal activity across widely distributed brain areas (Schnitzler and Gross, 2005). Movement initiation and execution is mediated by patterns of synchronization and desynchronization that occur concurrently across distinct frequency bands and in multiple motor cortical areas (Cheyne et al., 2008; Jurkiewicz et al., 2006; Pfurtscheller et al., 2003; Saleh et al., 2010)

To date, motor-related local oscillatory modulations are by and large examined by quantifying increases or suppressions in spectral power (Cheyne et al., 2008; Jurkiewicz et al., 2006; Pfurtscheller et al., 2003; Saleh et al., 2010). However, beyond band-limited oscillatory power, other spectral properties, namely phase and phase-amplitude coupling (PAC), are also thought to play a key role in neuronal encoding and information processing. The involvement of phase information in neuronal encoding has been extensively investigated in numerous perceptual modalities and higher-order cognitive tasks (Drewes and VanRullen, 2011; Dugue et al., 2011; Jensen et al., 2014; Klimesch et al., 2008, 2007; Montemurro et al., 2008; Palva and Palva, 2007; Sauseng and Klimesch, 2008; Sherman et al., 2016; VanRullen et al., 2011). In comparison, the role of phase and phase-based measures in the neuronal mechanisms that mediate motor processes is still under-investigated. Interestingly, a few studies provide evidence for the involvement of low-frequency phase and amplitude in the neuronal encoding of movement parameters (Hammer et al., 2016, 2013; Jerbi et al., 2011, 2007; Milekovic et al., 2012; Miller et al., 2012; Waldert et al., 2009, 2008). Nevertheless, the spatial, temporal and spectral dynamics of putative phase coding across the chain of processes starting from goal encoding, to motor planning and motor command execution, are still largely unresolved.

Furthermore, recent years have witnessed a surge in the interest in the putative mechanistic function of PAC (Cohen et al., 2008; Hemptinne et al., 2013; Lee and Jeong, 2013; Newman et al., 2013; Voytek, 2010, Bahramisharif et al., 2013) and numerous measures of PAC have been proposed (Canolty, 2006; Nakhnikian et al., 2016; Tort et al., 2010; Voytek et al., 2013, Ozkurt, 2012). Conceptually, PAC may provide a flexible framework for information processing by means of cross-frequency synchronization (Canolty and Knight, n.d.; Hyafil et al., 2015). However, despite important advances (Hemptinne et al., 2013; Özkurt and Schnitzler, 2011; Soto and Jerbi, 2012; Yanagisawa et al., 2012), the precise role of PAC in mediating motor planning and execution is still an open question. More specifically, the distinct functional role and mechanistic links between phase, amplitude and PAC estimates during motor behavior remain largely elusive.

Here, we compare the involvement of all three measures with high spatial, spectral and temporal resolution via access to rare multi-site intracerebral depth electrode recordings in humans performing a delayed motor task. In particular, we monitored modulations of neural activity at movement execution but also during the delay time window where motor intention is present but execution is withheld. In addition, we used a machine-learning approach to identify the key contributions and differences between the various neuronal responses. We found a high degree of overlap between power modulations patterns in the planning and execution periods, with prominent increases in the delta (2-4 Hz) and high gamma (60-200 Hz) bands. Alpha (8-13 Hz) and beta (13-30 Hz) power as well as very low-frequency phase parameters (<1,5 Hz) were the most discriminant features in dissociating action from intention. Despite providing weaker decoding, PAC features led to statistically significant classification of motor states, particularly in premotor brain areas. These results advance our understanding of the distinct and partly overlapping involvement of phase, amplitude and PAC in the neuronal mechanisms underlying motor intentions and executions.

2 Material and methods

2.1 Participants

Six patients with intractable epilepsy participated in this study (6 females, mean age 22.17±4.6). The patients were stereotactically implanted with multi-lead EEG depth electrodes at the Epilepsy Department of the Grenoble Neurological Hospital (Grenoble, France). In collaboration with the medical staff and based on visual inspection, electrodes presenting pathological waveforms were discarded from the present study. All participants provided written informed consent, and the experimental procedures were approved by the Institutional Review Board and by the National French Science Ethical Committee. Patient-specific clinical details are provided in Table 1.

2.2 Electrode implantation and Stereotactic EEG recordings

Each patient was implanted with stereotactic electroencephalography (SEEG) electrodes. Each one of them has a diameter of 0.8 mm and, depending on the implanted structure, composed with 10 to 15 contacts of 2 mm wide and 1.5 mm apart (DIXI Medical Instrument). Intracranial EEG signals were recorded from a total of 748 intracerebral sites across all patients (126 sites in each participant, except for one patient who had 118 recording sites). At the time of acquisition, a white matter electrode was used as reference and data were bandpass filtered from 0.1 to 200 Hz and sampled at 1024 Hz. Electrode locations were determined using the stereotactic implantation scheme and the Talairach and Tournoux proportional atlas (Talairach and Tournoux, 1993). The electrodes were localized in each individual subject in Talairach coordinates (based on post-implantation CT) and then a transformation to standard MNI coordinate system was performed according to standard routines and previously reported procedures (Bastin et al., 2016; Jerbi et al., 2010, 2009; Ossandon et al., 2011).

Figure 1. Delayed center-out task definition and Implantation visualization for the 6 subjects on a 3-D standard (MNI) brain *A*, Depth electrodes representation on a top, right and frontal view. We used one color pear subject *B*, Design of the delayed motor task where the direction of the movement is instructed at Cue1, and the actual movement in carried out at Cue 2 ('Go signal'). The timeline consists of three periods: Rest, a delay period (motor planning/intention) and movement execution.

2.3 Delayed center-out motor task

After a rest period of 1000ms, the participants were visually cued to to prepare a movement towards a visually presented target in one of four possible directions: up, down, left or right (*Planning phase*). Next, after a 1500 ms delay period, a Go signal prompted the subjects to move the cursor towards the target (*Execution Phase*). The Go signal consisted of a central cue changing from white to black. Figure 1c shows the task design.

	Handedness	Age	Gender	Epilepsy
P1	D	19	F	Frontal lobe (RH)
P2	D	23	F	Frontal lobe (LH)
P3	D	18	F	Frontal lobe (RH)
P4	D	18	F	Frontal lobe (RH)
P5	D	31	F	Insular (RH)
P6	D	24	F	Frontal lobe (LH)

Table 1. Patient data: age, gender, and broad description of epilepsy type as determined by the clinical staff of the Grenoble Neurological Hospital, Grenoble, France (Recording sites with epileptogenic activity were excluded from the analyses).

2.4 Data preprocessing

SEEG data preprocessing was conducted according to our routine procedures (Bastin et al., 2016; Jerbi et al., 2009). These include signal bipolarization where each electrode site is re-referenced to its direct neighbor. Bipolar re-referencing can increase sensitivity and reduce artefacts by canceling out distant signals that are picked up by adjacent electrode contacts (e.g. mains power). The spatial resolution of bipolar SEEG with the electrodes used here is approximately 3 mm (Jerbi et al., 2009; Kahane et al., 2006; Lachaux et al., 2003). Next, using visual inspection and time-frequency explorations of the signal we exclude electrodes containing pathological epileptic activity. In addition, electrodes located close to the extra-ocular eye muscles were systematically excluded to avoid contaminating our analyses with eye-movement activity. The pre-processing led to a total of 583 bipolar derivations across all participants.

2.5 Spectral analyses

We investigated phase, power and phase-amplitude coupling in several standard frequency bands defined as follow: very low frequency component (VLFC) [0.1-1.5Hz], delta (δ) [2-4Hz], theta (θ) [5-7Hz], alpha (α) [8-13Hz], beta (β) [13-30Hz], low gamma (low γ) [30-60] and broadband gamma (high γ) [60-200Hz]. The power features were computed in six bands (δ, θ, α, β, low γ and high γ), while the phase features were extracted for 4 bands (VLFC, δ, θ and α) and the phase-amplitude coupling was extracted using three combinations (δ, θ and α for phase and high γ for amplitude). In total 13 features were extracted for each SEEG bipolar derivation.

2.5.1 Spectral power estimation

Band-specific power modulations were computed using the Hilbert transform: we first filtered the data segments corresponding to each trial in the required band using a two-way least-squares FIR filter in order to avoid phase-shifting. We then computed the Hilbert transform of the filtered signal and calculated power by taking the square of the amplitude component (envelope). For the specific case of high-gamma, we split the 60-200Hz range into multiple non-overlapping10 Hz bands. As

performed in our previous studies, the broadband gamma power is then obtained by taking the mean of all of the successive 10Hz-wide normalized bands (Jerbi et al., 2009; Ossandon et al., 2011, Bastin et al., 2016). Comparable power estimations were obtained when using a wavelet approach (Tallon-Baudry et al., 1996) (Morlet wavelets) but we deliberately chose to use the Hilbert-based power computation to have a homogeneous methodological framework across all three features, since the Hilbert method was used to assess phase and PAC features (see below).
Statistical evaluation of tasked-based power modulations: Mapping power modulations on the standard MNI brain, is achieved using significant normalized power. The baseline period was defined as a 500 ms time window during pre-stimulus rest (from -750 ms to -250ms). The power in each frequency band was normalized by computing the relative change compared to this baseline at the same frequency (i.e. subtracting then dividing by the mean of the baseline).
Significant power modulations were obtained by standard permutation tests where power values (across time and frequency) during the task are randomly permuted with the corresponding value during baseline. A total of 1000 permutations were performed yielding a null distribution for the relative power which provides a minimal p value of 0.001.

2.5.2 Phase estimation

To extract phase features in a given frequency band, we first apply a bandpass filter to the bipolarized signals. Next, we extract the angle of the complex time-series based on the Hilbert transform. This gives the instantaneous phase for all time points. For the sake of decoding we used a phase value every 250 ms. This was done to have the same temporal resolution across all features (Power features are computed in 500 ms windows with a 50% overlap).
*Statistical evaluation of tasked-based phase modulations: w*e used Rayleigh's test to compute significant phase modulations (Babiloni et al., 2002; Lakatos, 2005; Tallon-Baudry et al., 1996) using the circular statistics toolbox (Berens and others, 2009).

2.5.3 Phase-Amplitude Coupling (PAC) estimation

Different methods for the estimation of PAC are used in the literature (Jensen and Colgin, 2007; Canolty and Knight, 2010; Tort et al., 2010; Soto and Jerbi, 2012; Aru et al., 2015). In order to choose which method to apply here, we first simulated synthetic PAC signals (e.g. Tort et al., 2010) and tested the behavior of several methods, namely Mean Vector Length (MVL) (Canolty, 2006), the Height-Ratio (HR) (Lakatos, 2005), Kullback-Leiber divergence (Tort et al., 2010), normalized direct PAC (ndPAC) (Ozkurt, 2012). We found that the tested PAC methods generally provided very comparable results. The slight differences arose from applying different normalization/surrogate data procedures. Based on this, we chose to use the MVL method combined with a normalization method where the surrogate values were calculated by randomly swapping phase and amplitude across trials (cf. Tort et al., 2010). In short, the PAC estimation procedure applied here can be summarized as follows: First, the low-frequency phase and high-frequency amplitude signals are obtained by filtering and Hilbert transformation in the frequencies of interest (Canolty et al. 2006). Being frequency-dependent, the optimal filter orders were adapted separately for the phase of the slower oscillations (3 cycles used) and for the amplitude of the faster oscillations (6 cycles), as proposed in previous studies (e.g. Bahramisharif et al., 2013). Next, surrogate data are generated by randomly swapping the phase trial data and amplitude trial data (i.e. randomly associating the high frequency amplitude envelope of a trial with the low-frequency phase time course of another trial). This was repeated 1000 times. Then, the normalized PAC value is then obtained by normalizing the MVL by the surrogate data yielding a z-score.

To visualize the emergence of PAC we used a combination of methods: (a) phase-alignment of single-trial TF maps (Canolty, 2006; Hemptinne et al., 2013), (b) a comodulogram (e.g. Foster and Parvizi, 2012; Pittman-Polletta et al., 2014) and (c) Event Related Phase-Amplitude Coupling

(Voytek et al., 2013). When computing the comodulogram, we used the entire planning and execution periods (1500 ms for each) in order to maximize the per-trial number of oscillation cycles for PAC estimation.

Statistical evaluation of tasked-based PAC modulations: The statistical assessment of PAC was achieved by comparing the true PAC values to the null distribution of PAC values computed using surrogate data, i.e. random shuffling across trials of phase and amplitude signals (Tort et al. 2010). A real PAC value that is higher than the 999th highest PAC value obtained with surrogate data, is considered significant at $p<0.001$.

Figure 2. Oscillatory feature extractions from intracranial EEG signals shown for an illustrative electrode in human premotor cortex (BA6). (A) Event-related potential (ERP) across the experiment timeline and anatomical location of the electrode. (B) Trial-by-trial raw iEEG signals. (C) Phase-locking factor indicating stimulus phase-locking (D) Single-trial phase shown here for the VLFC (<1.5 Hz) frequency range. (E) Time-Frequency representation. (F) Trial-by-trial broadband gamma envelopes (Hilbert amplitude, 60-200 Hz). (G) Phase-amplitude coupling maps during planning and during execution revealing prominent alpha-gamma coupling present during planning but absent during execution. (H) Event-related PAC depicts for the same electrode the time course of alpha-Gamma PAC.

2.6 Signal classification

We set out to explore the feasibility of using multi-site human LFP data (583 bipolar electrode sites) to perform three types of motor-state classifications: (a) during execution vs rest, (b) intention vs rest, and (c) execution vs intention. We compared the performance of several classification algorithms (Linear Discriminant Analysis (LDA), Naïve Bayes (NB), k-th Nearest Neighbor (KNN), Support Vector Machine (SVM) with linear and Radial Basis Function kernels and Random Forest). For single features classification, LDA, NB and SVM all provided similar results. We chose to use the LDA approach for its speed, which is a particularly important advantage here given that classifier performance was evaluated using computationally demanding permutation tests.

2.7 Decoding accuracy and statistical evaluation of decoding performance

Classification performance was evaluated using standard stratified 10-fold cross-validation. Here, the data set is first pseudo-randomly split into 10 equally sized observations: 9 segments are used for training the classifier and the last one as test set. This procedure is repeated 10 times so that every observation in the data is used exactly once for testing and at least once for training, but never at the same time. This strict separation of training and testing is critical to ensure the test data are naïve and do not violate basic classification principles (Lemm et al., 2011). The use of stratification seeks to ensure that the relative proportion of labels (or classes) in the whole data set is reasonably preserved within each of the segments after the split. The above procedure was repeated 10 times to reduce the effect of the random generation of the folds, yielding a 10 times 10-fold cross-validation framework. The performance of the achieved decoding is then calculated using the decoding accuracy (DA) metric, which is computed as the mean correct classification across all folds. Although the use of theoretical chance-levels (e.g. DA=50% for binary classification) can provide some useful indication on classifier performance in the presence of a high number of observations, the use of statistics is mandatory in order to assess the significance of classification performance (Combrisson and Jerbi, 2015). To this end, we used a permutation testing framework where the cross-validation and DA calculation are recomputed after randomly shuffling the labels of the classes. For each site and for each type of feature (i.e. power, phase and phase-amplitude coupling) 1000 permutations were generated, which allowed for statistical assessments with p values as lows as 0.001 (Combrisson and Jerbi, 2015; Ojala and Garriga, 2010, Golland and Fischl, 2003; Meyers and Kreiman, 2012).

2.8 Data mapping to a 3-D standard cortical representation

To facilitate the interpretation of the results, all significant task-based feature modulations and decoding results were remapped from the intracranial electrode sites onto a standard cortical representation. To achieve this, all electrode coordinates are first transformed from individual Talairach space to standard MNI space, and custom Matlab code was written to project the data from SEEG sites onto the cortical surface. In practice, data from the iEEG electrodes were assigned to the vertices on the MNI cortical mesh that fell within a fixed radial distance (here 10 mm) from each electrode. This cortical representation technique is methodologically consistent with methods used in previous iEEG studies (Bastin et al., 2016; Jerbi et al., 2009; Ossandon et al., 2012). In addition to generating brain-wide visualization of all significant features and decoding performances, this method can also be used to display the cortical coverage provided by all electrodes in this study (dark grey areas in Figure 2A).

3 Results

The frequency domain analysis of the over 500 intracranial recording sites across all participants revealed that the delayed motor task is mediated by complex patterns of spectral modulations across widely distributed brain areas. Figure 2 illustrates task-related modulations of amplitude, phase and phase-amplitude coupling measured via a recording site in premotor cortex (BA6) in one participant. The ERP and single-trial raw iEEG recordings for this electrode are shown in panels 2A and 2B respectively. Panels 2C and 2D depict the phase-locking factor and a trial-by-trial phase representation across time. The illustrative time-frequency map (Fig 2E) depicts a typical power modulation pattern where movement execution is associated with a strong increase in broadband gamma power and simultaneous power suppression in the alpha, beta and low gamma frequency bands. The delay period (between Cue1 and Cue2) at this site showed a different pattern with power increases at slow frequencies in the delta to alpha range. Because of the high SNR of the intracranial EEG recordings used, it is possible to estimate broadband gamma (60-200 Hz) power on a trial-by-trial basis (Fig 2F). The PAC maps in Figure 2G illustrate the changes in PAC values for the same recording site across motor planning and execution. A peak in PAC between alpha phase and high gamma amplitude is observed during the pre-movement delay period but vanishes during execution, where the maximum PAC is seen between delta phase and low-gamma amplitude. The predominance of the alpha-gamma coupling in the delay period can also confirmed when the TF maps are computed by first realigning the single-trial TF maps to alpha phase in the vicinity of the cue (Figure not shown). Finally, time-resolved event-related PAC estimations (Fig 2H) illustrate the existence at this location of an alpha-gamma coupling specific to the delay period. These oscillatory feature estimations were primarily presented to illustrate the wide variety and high signal-to-noise ratio (SNR) of the features explored in this study, using averaging over trials and, most importantly, in trial-by-trial mode. In the following we discuss the global results obtained using the data from all participants. First we describe the task-based oscillatory modulations and then we examine the results of the data-mining approach.

3.1 Task-based Spectral power modulations

The cortical mapping of power modulations during the delay period and motor execution reveal distinct patterns of increases and decreases across the various frequency bands (Figure 3). As one would expect from previous invasive reports in humans (Crone, 1998a,b; Crone et al., 2006; Babiloni et al., 2016; Bundy et al., 2016; Rektor et al., 2006), our results confirm motor execution is associated with prominent suppressions of alpha/mu and beta band power and with increases in high gamma power, all occurring primarily in motor and premotor cortices. The power modulations shown in figure 3, also indicate significant task-based modulations beyond these areas, extending to parietal, prefrontal and cingulate areas. When comparing the brain-wide significant power changes for both planning and execution, we first see the large degree of overlap between the statistically significantly active areas in both conditions. Second, two distinct patterns seem to emerge when we consider the direction of the effects across frequency bands: The alpha, beta and low-gamma bands show a reversal of the effect from a significant increase during planning to a significant suppression during execution. By contrast, the remaining bands (delta, theta and high gamma) all show a consistent significant power increase both during the planning and execution periods, although the effect is stronger at execution. It is noteworthy that the high gamma increase was prominent over multiple frontal and prefrontal brain areas bilaterally, with strong peaks in motor and premotor cortices during execution. Most of these areas also showed a significant increase of power above baseline levels during the delay period, potentially related to movement goal-encoding and motor planning processes.

Figure 3. Task-related modulations of spectral power during planning and execution of upper limb movements. The color bars represent percent relative changes (%) of power during planning (250-750ms) and during execution (2000-2500ms) with respect to baseline power during pre-stimulus rest (-750 to -250ms). The power modulations are shown for delta (2-4Hz), theta (5-7Hz), alpha (8-13Hz), beta (13-30Hz), low gamma (30-60Hz) and broadband gamma (60-200Hz). All modulations shown are statistically significant (permutation tests, $p<0.05$, FDR-corrected). Dark grey areas represent cortical regions for which electrode coverage was available but where the modulations did not reach statistical significance.

3.2 Task-based Phase modulations

The patterns of statistically significant changes in phase across all frequency bands show a predominant effect in the lower frequencies (Figure 4). In particular, the VLFC range (i.e. <1.5 Hz) shows consistent phase values in motor cortex with a reversal in sign between the selected instants in planning and execution windows. Interestingly, in addition to being sparser, consistent phase effects in the theta and alpha range were found in prefrontal areas. The significant delta phase angle over primary motor cortex during execution is in line with previous reports of delta-range coupling between motor cortex and movement parameters (Jerbi et al., 2007). Dark grey areas show the regions for which data were available but which did not exceed the significance threshold (Rayleigh test).

Figure 4. Task-related modulations of instantaneous phase during planning and execution of upper limb movements. The color bars represent mean phase computed at 500 ms (planning) and at 2250 ms (execution), which corresponds to the centers of the windows used for power (Fig 3). Phase modulations were computed for very low frequency component (VLFC) (<1.5 Hz), delta (2-4Hz), theta (5-7Hz) and alpha (8-13Hz) bands. All modulations shown are statistically significant (Rayleigh's test based on the circular statistics toolbox (Berens et al. (2009), $p<0.001$). Dark grey areas represent cortical regions for which electrode coverage was available but where the modulations did not reach statistical significance.

3.3 Task-based PAC modulations

Statistically significant levels of phase-amplitude coupling were found in multiple brain areas and with very similar distributions at execution and during the delay period in the absence of movement (Figure 5). In fact, our results suggest higher PAC during the goal-encoding and planning phase than during movement execution. The most prominent PAC effects in primary motor cortex were obtained for alpha-gamma interactions, but were only present during the planning period. This M1 alpha-gamma PAC virtually disappears at the time of execution. Interestingly, it is replaced by M1 delta-gamma PAC during motor execution. In addition, as shown in the right panel of Figure 5, depth recordings in the medial wall also reveal statistically significant PAC in ACC (cingulate motor cortex, CMA BA32) and in medial premotor areas (SMA, BA6) for all three slow frequency phase ranges explored (delta, theta and alpha). These PAC modulations are likely to reflect the involvement of SMA and cingulate motor areas in action selection, planning, execution as well as inhibition.

Figure 5. Task-related modulations of phase-amplitude coupling during planning and execution of upper limb movements (Left panel: top view, Right panel: medial view). The color bars represent mean PAC modulation index (MI) which quantifies the co-modulation of broadband gamma amplitude (60-200Hz) with delta (2-4Hz), theta (5-7Hz) and alpha (8-13Hz) bands. All modulations shown are statistically significant ($p<0.001$, surrogate data) and are estimated from the entire planning and execution time windows. Dark grey areas represent cortical regions for which electrode coverage was available but where the modulations did not reach statistical significance.

3.4 Single-trial decoding of motor processes using machine learning

The application of a supervised learning framework allowed us to determine which features, among those discussed above, provide single-trial decoding of motor states (binary classification applied pairwise to rest, planning and execution states). This was achieved using a cross-validation approach in which a classifier is repeatedly trained on a subset of the data and then tested on previously unseen single epoch observations (test set). This allows us to quantify the decoding strength of each feature as the percent correct classification rate.
The results in Fig 6 show the best decoding accuracies (DA) achieved with (on the x-axes) each feature type (power, phase and PAC) across all frequency bands, and (on the y-axes) all Brodmann areas for which recordings were available. The histograms above each panel depict the highest decoding accuracies obtained with each feature and indicate in which Brodmann area it was observed. The histograms on the right side of each panel indicate how many significant features were located in each Brodmann area. It is noteworthy that all features (amplitude, phase and PAC) computed across all frequencies and most of the probed brain areas all contain – to variable degrees- discriminant information on motor states (statistically significant decoding, $p<0.001$, permutation tests).

3.4.1 Classification similarities between Exec vs Rest and Intention vs Rest

The highest percentages of correct prediction were obtained when classifying Rest vs. Execution (as high as 99,4% on single-trial classifications). We found a high degree of similarity between the decoding patterns for Execution vs Rest (Fig 6A) and Intention vs Rest (Fig 6B). In principle, most of the features that provided significant decoding of execution vs Rest also provided significant decoding of Intention vs Rest. The features that yield the highest discrimination in both cases primarily involve spectral power in alpha (or mu), beta, low-gamma, and broadband gamma bands, in addition to phase in the very low frequency range (<1.5 Hz). The most prominent brain areas involved in both types of decoding are Brodmann areas 4, 6, 8, 9, 32 and 40. These were the regions that contained the highest number of significantly decoding sites (gray histograms), and these were also the Brodmann areas that yielded the highest levels of decoding accuracy. The highest DA performances reached 99,4% (Beta power in primary motor cortex) when decoding Execution vs Rest and 91,3% (VLFC phase in inferior parietal cortex, BA40) in the case of Intention vs Rest.

3.4.2 Classification differences between Exec vs Rest and Intention vs Rest

Although the decoding matrices in Figs 6A and 6B show a high degree of similarity, there are noteworthy differences. Brodmann area 40 plays a more prominent role in distinguishing intention from rest than actual movement execution from rest. This is in line with the role of inferior parietal lobe in motor planning (Caspers et al., 2008; Mattingley et al., 1998; Rushworth et al., 2001) Interestingly the involvement of BA40 was manifest in the phase and PAC features more than with power, although power in BA40 also provided statistically significant decoding (Fig 6B). In addition, power-based classification results show that while primary motor (BA4) and premotor areas (BA6) were prominent in predicting execution, the most noticeable intention-decoding power features were recorded from BA8, 9, 13 and 32 (Top histogram in Panel 6B). This is consistent with the previously reported involvement of BA8 and BA9 in working memory, executive planning and behavioral inhibition. Additionally, BA8 is also involved in visuomotor processing, visuospatial attention and motor imagery. The important decoding levels in BA13 could be related to the role of insular cortex, among other things, in motor planning (Lacourse et al., 2005; Stephan et al., 1995). Remarkably, the best PAC-mediated decoding of Intention vs Rest was obtained with delta-gamma coupling in the dorsal anterior cingulate cortex (dACC, BA32, DA=70%). This prominent PAC in ACC is consistent with the role of this area in motor inhibition, visuo spatial attention, motor

planning and imagery (Cheyne et al., 2012; Jahanshahi et al., 1995; Paus, 2001). To the best of our knowledge, this is the first report of a motor-related delta-gamma PAC modulation in ACC.

3.4.3 Classification of motor execution versus motor intention

The decoding matrix for Execution vs Intention (Fig 6C) shows a markedly different pattern. Because of the similarity between the neuronal responses in both conditions (cf. previous sections), the differences are subtle and, as a result, the decoding performances here are lower than those reported for the Execution vs Rest or Intention vs Rest classifications. In addition, fewer brain areas provide statistically significant decoding. Interestingly, among all features, the one that provided the best discrimination between the intention period and the execution period is the phase of the very low frequency range (<1.5 Hz), which yielded a decoding peak of 88,9% in BA 8. Interestingly, the second best decoding feature, after the low-frequency phase, was broadband gamma with a peak of decoding of (DA=83,5%) in posterior Insula (BA13). Moreover, although they did not provide the highest classifications performance, the premotor cortex sites (BA6) led to statistically significant decoding of intention vs execution with every single feature tested (peaking at 82,3% with VLFC phase). This is in line with the established involvement of premotor cortex in motor execution, planning and imagery, as well as visuomotor and visuospatial attention (Ball et al., 1999; Gallivan et al., 2013; Hanakawa et al., 2008; Miller et al., 2010). Finally, Figure 6C also shows that among all PAC features used, theta-gamma coupling provided the lowest decoding, while delta-gamma and alpha-gamma provided the best decoding, predominantly in primary motor (BA4) and premotor (BA6) cortices. This is most likely explained by the significant decoding accuracy (in these areas) of delta-gamma and alpha-gamma observed respectively in execution-rest (Fig 6A) and intention-rest (Fig 6B).

A
Decoding Execution vs Rest (p<0.001)
Top decoding for each type of feature and associated brodmann area
DA (%)
100
75
50
BA 6
BA 6
BA 4
BA 4
BA 6
BA 8
BA 6
BA 40
BA 47
BA 40
BA 4
BA 9
BA 6
0
672
Brodmann Area
Number of significant features inside each brodmann area
δ θ α β Low γ High γ VLFC δ θ α δ & γ θ & γ α & γ
50 Decoding accuracy (%) 100
B
Decoding Intention vs Rest (p<0.001)
Top decoding for each type of feature and associated brodmann area
DA (%)
BA 8
BA 13
BA 13
BA 32
BA 9
BA 13
BA 40
BA 40
BA 9
BA 8
BA 32
BA 4
BA 6
0
675
Brodmann Area
Number of significant features inside each brodmann area
δ θ α β Low γ High γ VLFC δ θ α δ & γ θ & γ α & γ
50 Decoding accuracy (%) 100
C
Decoding Execution vs Intention (p<0.001)
Top decoding for each type of feature and associated brodmann area
DA (%)
BA 13
BA 6
BA 40
BA 4
BA 6
BA 13
BA 8
BA 6
BA 9
BA 9
BA 4
BA 6
BA 6
0
210
Brodmann Area
Number of significant features inside each brodmann area
δ θ α β Low γ High γ VLFC δ θ α δ & γ θ & γ α & γ
50 Decoding accuracy (%) 100

Figure 6

Single-trial pairwise classification of motor states (pre-stimulus rest, intention and execution) using power, phase and PAC features. (**A**) Execution vs rest, (**B**) Intention vs rest (**C**) Execution vs intention. Each of the three decoding matrices depict the best percent decoding accuracies (DA) achieved with each feature type (power, phase and PAC) across all frequency bands (x-axes), and across all Brodmann areas for which recordings were available (y-axes). Yellow indicates 100% DA, while deep blue indicates 50% DA. Only statistically significant decoding accuracies are reported ($p<0.001$, permutation test). The bar plots above each panel show the highest decoding accuracies obtained with each feature (red: power, blue: phase and green: PAC) and indicate in which Brodmann area it was observed. The gray bar plots on the right side of each panel indicate how many significant features were located in each Brodmann area.

4 Discussion

4.1 Summary of findings

Previous reports have reported complex patterns of overlap and segregation between the networks of brain areas involved in action representation, planning and execution (e.g. Jeannerod, 1994; Stephan et al., 1995; Hanakawa et al., 2008; Guillot et al., 2012;) So far, electrophysiological explorations of these patterns in humans have focused on modulations of rhythmic activity, primarily measured through spectral power modulations. Using rare direct recordings in humans performing a delayed motor task, we here provide an intracranial investigation of the similarities and discrepancies between activations across areas involved in the build-up to action and those that actually mediate movement execution. To the best of our knowledge, this study provides the first account of the involved brain dynamics through systematic investigation of the roles of phase, amplitude and phase-amplitude coupling, measured in widely distributed brain areas (extending beyond primary motor areas) and across a wide range of frequencies (up to 200 Hz). Furthermore, in addition to statistical comparisons, an important addition of this study is the use of supervised learning and single-trial classification framework as a way to assess the distinct or overlapping information content of these features.

4.2 Planning and execution are associated with prominent phase, amplitude and PAC modulations

Our analyses of task-based modulations revealed spatially distributed patterns of statistically significant changes in power, phase and PAC, both during the delay period and following the execution cue. We found a large overlap between the patterns of significant modulations observed in the planning processes and in movement execution. In task-based power changes significant increases in the delta, theta and broadband gamma bands were present during the planning phase and further enhanced at the time of execution. By contrast, although largely present in the same areas, alpha, beta and low-gamma power showed a rather consistent pattern of inversion from increases following the preparation cue to decreases following the execution cue. These similarities and differences between activation patterns in intention and execution also echo previously reported similarities and differences when comparing motor imagery versus motor execution. In particular, action representation and motor inhibition are likely to be common to the delay period of the task used here and to motor imagery tasks (Stephan et al., 1995). Phase and PAC also displayed similar patterns of responses for action preparation and execution, although they also pinpointed to clear discrepancies in motor areas, such as prominent alpha-gamma coupling during planning, but delta-gamma coupling during execution. Although not spatially exhaustive, the reported pre- and peri-movement modulations of power, phase and PAC was found over large parts of parietal, frontal and prefrontal areas, including medial areas such as the dACC. Although, the reported power modulations are largely consistent with previous iEEG reports (Crone, 1998; Crone et al., 2006, 1998; Pfurtscheller and Lopes da Silva, 1999), the wide-spread significant modulations of phase and PAC effects, during the delay-period and movement execution, are to our knowledge novel findings that extend the current understanding of their role in mediating motor behavior (cf.. Yanagisawa et al., 2012).

4.3 Decoding patterns reveal functional overlap and discrepancies across phase, amplitude and PAC

The machine learning framework allowed a second level of investigation of the role of the explored features through the assessment of their decoding accuracy. Most importantly, the classification strategy extends the standard statistical analyses by switching from comparisons of means to evaluating the predictive power of data computed from single-trials; once trained, the classifier is individually applied to each single-trial of the test set. In other words, the mean decoding performances reflect the ability of each feature to discriminate between the classes (rest, intention, execution) on single samples of data. Taken together, the decoding analyses provide a rich multi-dimensional exploration of the functional involvement of the phase, amplitude and phase-amplitude in motor behavior. The main findings can be summarized as follows: First of all, they reveal a prominent role for slow frequency phase primarily in inferior parietal areas (up to DA>90%) and phase-amplitude coupling in dACC (up to DA>70%) during the delay-period preceding motor execution. Second, in contrast to classification in the planning phase, the highest classification accuracies for movement execution were achieved by power features rather than phase, and the best accuracies were found predominantly in sites in motor and premotor areas (BA4, 6 and 8). While during planning, the best power-based decoding levels were observed in prefrontal regions (BA8, 9, 13 and 32). Thirdly, we show that when it comes to directly distinguishing the neuronal correlates of execution from those mediating intention the highest classifications (among all explored regions) were achieved either with the phase of the very low frequency range (<1.5 Hz) in BA8 (DA>88%) or broadband gamma (60-200 Hz) power posterior Insula (DA>83%).

4.4 Relationship with previous electrophysiological findings in humans

Overall, the present analysis of invasive human data provides strong evidence for a prominent role of phase and phase-amplitude relationships in motor goal encoding and execution. These findings extend previous invasive electrophysiological studies in human that have investigated the correlates of motor intention and/or execution. For the large part, such studies are primarily based on the estimation of cortical power modulations in multiple frequency components of the local-field potential (LFP). Nevertheless, there is some evidence by a few studies that suggests a critical role for low-frequency phase and amplitude information in mediating movement parameters (Hammer et al., 2016, 2013; Jerbi et al., 2011, 2007; Milekovic et al., 2012; Miller et al., 2012; Waldert et al., 2009, 2008; (Kajihara et al., 2015; Liu et al., 2011) . Using ECoG recordings in epielpsy patients, (Hammer et al., 2013) directly addressed the role of phase information in decoding movement kinematics (i.e. position, velocity, and acceleration). By separately exploring kinematics decoding with spectral amplitudes and phase of the low frequency component (LFC), the authors came to the conclusion that the ECoG LFC phase is indeed much more informative than amplitudes (Hammer et al., 2013). Interestingly, a recent study by the same group has demonstrated that the LFC signal is more adapted to speed then to velocity decoding (Hammer et al., 2016). These findings are globally in line with reports of low fequency range coherence (<4 Hz) between limb speed and the activity of the primary motor cortex during continuous movements (Jerbi et al., 2007). Moreover, movement direction classification has also been achieved with significant success using the LFC component both with invasive and non-invasive brain recordings (Jerbi et al., 2011; Waldert et al., 2009, 2008).

The results of the current study differ from the above reports in several ways. First of all, while most studies typically focus on movement execution, the delayed-motor task used here allowed us to examine the role of phase information in the motor encoding and planning phase preceding actual movement execution. Secondly, beyond phase or amplitude, we performed a systematic exploration of phase-amplitude coupling effects and found that it yielded significant modulations and motor

state decoding. Thirdly, the intracerebral recordings used here provide a different spatial sampling compared to most previous electrocorticographic studies. In particular, the recording sites used here cover widely distributed brain areas beyond primary and secondary motor areas (such as parietal, prefrontal and insular cortices) and include some medial brain areas (e.g. dACC). Our rare exploration confirms and extends the growing body of evidence for the role of phase and phase-related measures in encoding motor processes along the path from intention to execution.

4.5 A key role for phase coding in goal-directed motor behavior

The overwhelming body of current research on the oscillatory brain dynamics that mediate motor behavior, continues to explore local activations in terms of amplitude modulations, while phase information is predominantly employed in the context of inter-regional connectivity assessments (da Silva, 2006; Lachaux et al., 1999; Le Van Quyen et al., 2001; Roach and Mathalon, 2008). Our findings support the view that phase carries critical information that is often overlooked and which, as shown here, can actually provide more task-specific information than spectral amplitude. One may thus ask what is the underlying physiological phenomenon that may explain the relevance of phase? One potential explanation is that phase modulations may be closely related to neuronal firing patterns. Evidence for this hypothesis has been found in non-human primates, for instance in auditory and visual cortices (Ng et al., 2013; Montemurro et al., 2008). In auditory cortex, stimulus selective firing patterns were found to imprint on the phase rather than the amplitude of theta oscillations both in LFPs and EEG data (Ng et al. 2013). By applying a stimulus decoding technique to intracortical LFPs and single cell recordings in macaque auditory cortex, the authors found that the stimuli which were successfully discriminated by firing rates were also discriminated by phase patterns but not by oscillation amplitude. In visual cortex, information theoretical approaches have also shown that spikes and low frequency (1-4 Hz) LFP phase provide, when combined, more information on visual stimuli than that achieved with spikes alone (Latham and Lengyel, 2008; Montemurro et al., 2008). In the hippocampus, the timing of action potentials relative to LFP theta phase has been shown to provide information about position with higher precision than what could be inferred from the firing rate alone (Dragoi and Buzsáki, 2006; O'Keefe and Recce, 1993).

The low-frequency phase decoding results obtained in the current study could in theory be explained by (and be consistent with) the hypothesis that the precise relationship between slow LFP phase and neuronal firing reported in other modalities (as discussed above) also operates in human motor brain areas during limb movement preparation and execution. This needs to be tested with microelectrode data allowing for simultaneous access to LFP and neuronal firing in humans. This was not possible with the SEEG data used for the purpose of this study. Furthermore, the significant modulations and decoding results reported here using phase-amplitude coupling, could suggest a specific mechanism by which phase, broadband gamma amplitude and neuronal firing are precisely lined up. Whittingstall and Logothetis (2009) observed such a phenomenon in Macaque visual cortex, where multi-unit firing responses were found to be strongest only when increases in EEG gamma power occurred during a specific phase of the delta (2-4 Hz) wave. Obviously, the tempting speculation that this may also be at play in areas where we found significant PAC decoding cannot be confirmed without simultaneous monitoring of neuronal firing.

Although the notion of phase coding *per se* is of course not new, the results found here extend an emerging body of literature that indicates that phase coding may play a more important role in motor behavior than previously admitted. In particular, we found that phase-based motor state decoding is (i) widely distributed, extending beyond primary motor areas, (ii) that it occurs both during action planning and execution, and (iii) that phase-amplitude coupling could be an additional phase-based key feature involved in the neuronal coding of goal-directed behavior.

4.6 Limitations and open questions

The results reported here have a number of limitations. As in all previous intracranial EEG studies, only limited spatial sampling of the involved networks is possible. The over 500 intracerebral sites used here only provide partial spatial coverage of the brain. This said the patients were selected primarily based on whether they had electrodes implanted in frontal, prefrontal or parietal areas, thus providing a reasonable coverage of the targeted networks. Moreover, as in all previous human invasive recordings, the participants suffer from drug-resistant epilepsy, which could limit the generalizability of the findings to healthy subjects. To address this in the best possible way, our standard procedure (e.g. Jerbi et al. 2009) is to exclude electrodes that display pathological activity (such as epileptic spikes) and to focus on task-related changes and multi-trial analyses thereby reducing the impact of neuronal activations that are not related to the task and which occur randomly across the recording session. Therefore, our findings would definitely benefit from future replication in non-invasive recordings with healthy controls. Furthermore, it is important to acknowledge that our task design did not allow for a fine-grained disambiguation of the various distinct processes that are expected to be at play in the delay period. Clearly, the delay period will encompass a wide range of processes such as stimulus encoding, visuomotor transformations, motor imagery, action selection, motor preparation as well as working memory and maintenance processes. To limit the bias towards visual stimulus encoding when analyzing the planning/intention period, we deliberately centered our phase and amplitude analyses on 500ms after target onset. However, more generally, this limitation will have to be addressed with a dedicated experimental design in future investigations. Moreover, the physiological interpretations made here also bare their share of limitations. The way the features explored here (amplitude, phase and PAC) explicitly relate to the notion of information processing in the human brain is still poorly understood. Rather than identifying the precise functional relevance of each feature, our results emphasize their task-specific modulations and their ability to successfully predict on singe-trial data the state of the system (i.e. rest, motor planning or motor execution).

Despite the above limitations, we feel that the rare access to intracerebral depth EEG recordings in human subjects, in particular with simultaneous monitoring of lateral and medial wall structures, provides a privileged access to the neural dynamics mediating human cognition with very high spatial, temporal and spectral precision. In the long run, such data can help bridge the gap with neuroimaging studies and with electrophysiological recordings in non-human primates.

5 Conclusion

The findings reported here provide novel experimental evidence for the role of oscillatory amplitude and phase properties in motor planning and execution. In particular, the evidence for phase and PAC-based coding provide compelling support for the key role of phase in encoding motor representations and mediating movement execution, across widely distributed brain areas. We hope that these results will pave the way for a better understanding and for novel hypotheses with respect to the role of phase, amplitude and the coordination between the two, in goal-directed motor behavior in humans.

Acknowledgments

Etienne Combrisson was supported in part by a PhD Scholarship awarded by the Ecole Doctorale Inter-Disciplinaire Sciences-Santé (EDISS), Lyon, France, and by PhD funding from the Natural Sciences and Engineering Research Council of Canada (NSERC). This work was partly performed within the framework of the LABEX CORTEX (ANR-11-LABX-0042) of Université de Lyon, within the program ANR-11-IDEX-0007 and thanks to funding from the Canada Research Chairs program and a Discovery Grant (RGPIN-2015-04854) awarded by the Natural Sciences and Engineering Research Council of Canada to K.J. The authors are grateful for the collaboration with patients and clinical staff at the Epilepsy Department of the Grenoble University Hospital.

Conflict of interest

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

Ariani, G., Wurm, M.F., Lingnau, A., 2015. Decoding Internally and Externally Driven Movement Plans. J. Neurosci. Off. J. Soc. Neurosci. 35, 14160–14171. doi:10.1523/JNEUROSCI.0596-15.2015

Aru, J., Aru, J., Priesemann, V., Wibral, M., Lana, L., Pipa, G., Singer, W., Vicente, R., 2015. Untangling cross-frequency coupling in neuroscience. Curr. Opin. Neurobiol., SI: Brain rhythms and dynamic coordination 31, 51–61. doi:10.1016/j.conb.2014.08.002

Babiloni, C., Del Percio, C., Vecchio, F., Sebastiano, F., Di Gennaro, G., Quarato, P.P., Morace, R., Pavone, L., Soricelli, A., Noce, G., Esposito, V., Rossini, P.M., Gallese, V., Mirabella, G., 2016. Alpha, beta and gamma electrocorticographic rhythms in somatosensory, motor, premotor and prefrontal cortical areas differ in movement execution and observation in humans. Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol. 127, 641–654. doi:10.1016/j.clinph.2015.04.068

Babiloni, C., Babiloni, F., Carducci, F., Cincotti, F., Rosciarelli, F., Arendt-Nielsen, L., Chen, A.C.N., Rossini, P.M., 2002. Human brain oscillatory activity phase-locked to painful electrical stimulations: A multi-channel EEG study. Hum. Brain Mapp. 15, 112–123. doi:10.1002/hbm.10013

Bahramisharif, A., van Gerven, M.A.J., Aarnoutse, E.J., Mercier, M.R., Schwartz, T.H., Foxe, J.J., Ramsey, N.F., Jensen, O., 2013. Propagating Neocortical Gamma Bursts Are Coordinated by Traveling Alpha Waves. J. Neurosci. 33, 18849–18854. doi:10.1523/JNEUROSCI.2455-13.2013

Ball, T., Schreiber, A., Feige, B., Wagner, M., Lücking, C.H., Kristeva-Feige, R., 1999. The role of higher-order motor areas in voluntary movement as revealed by high-resolution EEG and fMRI. Neuroimage 10, 682–694.

Bastin, J., Deman, P., David, O., Gueguen, M., Benis, D., Minotti, L., Hoffman, D., Combrisson, E., Kujala, J., Perrone-Bertolotti, M., Kahane, P., Lachaux, J.-P., Jerbi, K., 2016. Direct Recordings from Human Anterior Insula Reveal its Leading Role within the Error-Monitoring Network. Cereb. Cortex bhv352. doi:10.1093/cercor/bhv352

Berens, P., others, 2009. CircStat a MATLAB toolbox for circular statistics. J Stat Softw 31, 1–21.

Brovelli, A., Lachaux, J.-P., Kahane, P., Boussaoud, D., 2005. High gamma frequency oscillatory activity dissociates attention from intention in the human premotor cortex. NeuroImage 28, 154–164. doi:10.1016/j.neuroimage.2005.05.045

Bundy, D.T., Pahwa, M., Szrama, N., Leuthardt, E.C., 2016. Decoding three-dimensional reaching movements using electrocorticographic signals in humans. J. Neural Eng. 13, 026021. doi:10.1088/1741-2560/13/2/026021

Canolty, R.T., 2006. High Gamma Power Is Phase-Locked to Theta. science 1128115, 313.

Canolty, R. T., & Knight, R. T. (2010). The functional role of cross-frequency coupling. *Trends in Cognitive Sciences*, *14*(11), 506–515. http://doi.org/10.1016/j.tics.2010.09.001

Cheyne, D.O., Ferrari, P., Cheyne, J.A., 2012. Intended actions and unexpected outcomes: automatic and controlled processing in a rapid motor task. Front. Hum. Neurosci. 6, 237. doi:10.3389/fnhum.2012.00237

Cheyne, D., Bells, S., Ferrari, P., Gaetz, W., Bostan, A.C., 2008. Self-paced movements induce high-frequency gamma oscillations in primary motor cortex. NeuroImage 42, 332–342. doi:10.1016/j.neuroimage.2008.04.178

Cohen, M.X., Elger, C.E., Fell, J., 2008. Oscillatory activity and phase–amplitude coupling in the human medial frontal cortex during decision making. J. Cogn. Neurosci. 21, 390–402.

Combrisson, E., Jerbi, K., 2015. Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. J. Neurosci. Methods 250, 126–136. doi:10.1016/j.jneumeth.2015.01.010

Crone, N., 1998. Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. I. Alpha and beta event- related desynchronization. Brain 121, 2271–2299. doi:10.1093/brain/121.12.2271

Crone, N.E., Miglioretti, D.L., Gordon, B., Lesser, R.P., 1998. Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. II. Event-related synchronization in the gamma band. Brain 121, 2301–2315.

Crone, N.E., Sinai, A., Korzeniewska, A., 2006. High-frequency gamma oscillations and human brain mapping with electrocorticography, in: Progress in Brain Research. Elsevier, pp. 275–295.

da Silva, F.H.L., 2006. Event-related neural activities: what about phase? Prog. Brain Res. 159, 3–17.

Desmurget, M., Sirigu, A., 2009. A parietal-premotor network for movement intention and motor awareness. Trends Cogn. Sci. 13, 411–419. doi:10.1016/j.tics.2009.08.001

Dragoi, G., Buzsáki, G., 2006. Temporal Encoding of Place Sequences by Hippocampal Cell Assemblies. Neuron 50, 145–157. doi:10.1016/j.neuron.2006.02.023

Drewes, J., VanRullen, R., 2011. This Is the Rhythm of Your Eyes: The Phase of Ongoing Electroencephalogram Oscillations Modulates Saccadic Reaction Time. J. Neurosci. 31, 4698–4708. doi:10.1523/JNEUROSCI.4795-10.2011

Dugue, L., Marque, P., VanRullen, R., 2011. The Phase of Ongoing Oscillations Mediates the Causal Relation between Brain Excitation and Visual Perception. J. Neurosci. 31, 11889–11893. doi:10.1523/JNEUROSCI.1161-11.2011

Foster, B.L., Parvizi, J., 2012. Resting oscillations and cross-frequency coupling in the human posteromedial cortex. NeuroImage 60, 384–391. doi:10.1016/j.neuroimage.2011.12.019

Gallivan, J.P., McLean, D.A., Flanagan, J.R., Culham, J.C., 2013. Where one hand meets the other: limb-specific and action-dependent movement plans decoded from preparatory signals in single human frontoparietal brain areas. J. Neurosci. 33, 1991–2008.

Golland, P., Fischl, B., 2003. Permutation tests for classification: towards statistical significance in image-based studies. Inf. Process. Med. Imaging Proc. Conf. 18, 330–341.

Guillot, A., Di Rienzo, F., MacIntyre, T., Moran, A., Collet, C., 2012. Imagining is not doing but involves specific motor commands: a review of experimental data related to motor inhibition. Front. Hum. Neurosci. 6, 247.

Hammer, J., Fischer, J., Ruescher, J., Schulze-Bonhage, A., Aertsen, A., Ball, T., 2013. The role of ECoG magnitude and phase in decoding position, velocity, and acceleration during continuous motor behavior. Front. Neurosci. 7. doi:10.3389/fnins.2013.00200

Hammer, J., Pistohl, T., Fischer, J., Kršek, P., Tomášek, M., Marusič, P., Schulze-Bonhage, A., Aertsen, A., Ball, T., 2016. Predominance of Movement Speed Over Direction in Neuronal Population Signals of Motor Cortex: Intracranial EEG Data and A Simple Explanatory Model. Cereb. Cortex 26, 2863–2881. doi:10.1093/cercor/bhw033

Hanakawa, T., Dimyan, M.A., Hallett, M., 2008. Motor Planning, Imagery, and Execution in the Distributed Motor Network: A Time-Course Study with Functional MRI. Cereb. Cortex 18, 2775–2788. doi:10.1093/cercor/bhn036

Hemptinne, C. de, Ryapolova-Webb, E.S., Air, E.L., Garcia, P.A., Miller, K.J., Ojemann, J.G., Ostrem, J.L., Galifianakis, N.B., Starr, P.A., 2013. Exaggerated phase–amplitude coupling in the primary motor cortex in Parkinson disease. Proc. Natl. Acad. Sci. doi:10.1073/pnas.1214546110

Hyafil, A., Giraud, A.-L., Fontolan, L., Gutkin, B., 2015. Neural Cross-Frequency Coupling: Connecting Architectures, Mechanisms, and Functions. Trends Neurosci. 38, 725–740. doi:10.1016/j.tins.2015.09.001

Jahanshahi, M., Jenkins, I.H., Brown, R.G., Marsden, C.D., Passingham, R.E., Brooks, D.J., 1995. Self-initiated versus externally triggered movements. I. An investigation using measurement of regional cerebral blood flow with PET and movement-related potentials in normal and Parkinson's disease subjects. Brain J. Neurol. 118 (Pt 4), 913–933.

Jeannerod, M., 1994. The representing brain: Neural correlates of motor intention and imagery. Behav. Brain Sci. 17, 187–202.

Jensen, O., Colgin, L.L., 2007. Cross-frequency coupling between neuronal oscillations. Trends Cogn. Sci. 11, 267–269. doi:10.1016/j.tics.2007.05.003

Jensen, O., Gips, B., Bergmann, T.O., Bonnefond, M., 2014. Temporal coding organized by coupled alpha and gamma oscillations prioritize visual processing. Trends Neurosci. 37, 357–369. doi:10.1016/j.tins.2014.04.001

Jerbi, K., Lachaux, J.-P., Karim, N., Pantazis, D., Leahy, R.M., Garnero, L., Baillet, S., others, 2007. Coherent neural representation of hand speed in humans revealed by MEG imaging. Proc. Natl. Acad. Sci. 104, 7676–7681.

Jerbi, K., Ossandón, T., Hamamé, C.M., Senova, S., Dalal, S.S., Jung, J., Minotti, L., Bertrand, O., Berthoz, A., Kahane, P., Lachaux, J.-P., 2009. Task-related gamma-band dynamics from an intracerebral perspective: Review and implications for surface EEG and MEG. Hum. Brain Mapp. 30, 1758–1771. doi:10.1002/hbm.20750

Jerbi, K., Vidal, J.R., Mattout, J., Maby, E., Lecaignard, F., Ossandon, T., Hamamé, C.M., Dalal, S.S., Bouet, R., Lachaux, J.-P., Leahy, R.M., Baillet, S., Garnero, L., Delpuech, C., Bertrand, O., 2011. Inferring hand movement kinematics from MEG, EEG and intracranial EEG: From brain-machine interfaces to motor rehabilitation. IRBM 32, 8–18. doi:10.1016/j.irbm.2010.12.004

Jerbi, K., Vidal, J.R., Ossandon, T., Dalal, S.S., Jung, J., Hoffmann, D., Minotti, L., Bertrand, O., Kahane, P., Lachaux, J.-P., 2010. Exploring the electrophysiological correlates of the default-mode network with intracerebral EEG. Front. Syst. Neurosci. 4, 27. doi:10.3389/fnsys.2010.00027

Jurkiewicz, M.T., Gaetz, W.C., Bostan, A.C., Cheyne, D., 2006. Post-movement beta rebound is generated in motor cortex: Evidence from neuromagnetic recordings. NeuroImage 32, 1281–1289. doi:10.1016/j.neuroimage.2006.06.005

Kahane, P., Landré, E., Minotti, L., Francione, S., Ryvlin, P., 2006. The Bancaud and Talairach view on the epileptogenic zone: a working hypothesis. Epileptic Disord. Int. Epilepsy J. Videotape 8, S16–26.

Kajihara, T., Anwar, M.N., Kawasaki, M., Mizuno, Y., Nakazawa, K., Kitajo, K., 2015. Neural dynamics in motor preparation: From phase-mediated global computation to amplitude-mediated local computation. NeuroImage 118, 445–455. doi:10.1016/j.neuroimage.2015.05.032

Kalaska, J.F., 2009. From Intention to Action: Motor Cortex and the Control of Reaching Movements, in: Sternad, D. (Ed.), Progress in Motor Control. Springer US, Boston, MA, pp. 139–178.

Klimesch, W., Freunberger, R., Sauseng, P., Gruber, W., 2008. A short review of slow phase synchronization and memory: Evidence for control processes in different memory systems? Brain Res., Brain Oscillations in Cognition and Cognitive Disorders 1235, 31–44. doi:10.1016/j.brainres.2008.06.049

Klimesch, W., Sauseng, P., Hanslmayr, S., 2007. EEG alpha oscillations: The inhibition–timing hypothesis. Brain Res. Rev. 53, 63–88. doi:10.1016/j.brainresrev.2006.06.003

Lachaux, J.-P., Rodriguez, E., Martinerie, J., Varela, F.J., others, 1999. Measuring phase synchrony in brain signals. Hum. Brain Mapp. 8, 194–208.

Lachaux, J.P., Rudrauf, D., Kahane, P., 2003. Intracranial EEG and human brain mapping. J. Physiol.-Paris 97, 613–628. doi:10.1016/j.jphysparis.2004.01.018

Lacourse, M.G., Orr, E.L.R., Cramer, S.C., Cohen, M.J., 2005. Brain activation during execution and motor imagery of novel and skilled sequential hand movements. NeuroImage 27, 505–519. doi:10.1016/j.neuroimage.2005.04.025

Lakatos, P., 2005. An Oscillatory Hierarchy Controlling Neuronal Excitability and Stimulus Processing in the Auditory Cortex. J. Neurophysiol. 94, 1904–1911. doi:10.1152/jn.00263.2005

Latham, P.E., Lengyel, M., 2008. Phase Coding: Spikes Get a Boost from Local Fields. Curr. Biol. 18, R349–R351. doi:10.1016/j.cub.2008.02.062

Lau, H.C., 2004. Attention to Intention. Science 303, 1208–1210. doi:10.1126/science.1090973

Le Van Quyen, M., Foucher, J., Lachaux, J.-P., Rodriguez, E., Lutz, A., Martinerie, J., Varela, F.J., 2001. Comparison of Hilbert transform and wavelet methods for the analysis of neuronal synchrony. J. Neurosci. Methods 111, 83–98.

Lee, J., Jeong, J., 2013. Correlation of risk-taking propensity with cross-frequency phase–amplitude coupling in the resting EEG. Clin. Neurophysiol. doi:10.1016/j.clinph.2013.05.007

Lemm, S., Blankertz, B., Dickhaus, T., Müller, K.-R., 2011. Introduction to machine learning for brain imaging. NeuroImage 56, 387–399. doi:10.1016/j.neuroimage.2010.11.004

Liu, J., Perdoni, C., He, B., 2011. Hand movement decoding by phase-locking low frequency EEG signals. Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf. 2011, 6335–6338. doi:10.1109/IEMBS.2011.6091564

Meyers, E., Kreiman, G., 2012. Tutorial on pattern classification in cell recording. Vis. Popul. Codes 517–538.

Milekovic, T., Fischer, J., Pistohl, T., Ruescher, J., Schulze-Bonhage, A., Aertsen, A., Rickert, J., Ball, T., Mehring, C., 2012. An online brain–machine interface using decoding of movement direction from the human electrocorticogram. J. Neural Eng. 9, 046003. doi:10.1088/1741-2560/9/4/046003

Miller, K.J., Schalk, G., Fetz, E.E., den Nijs, M., Ojemann, J.G., Rao, R.P., 2010. Cortical activity during motor execution, motor imagery, and imagery-based online feedback. Proc. Natl. Acad. Sci. 107, 4430–4435.

Miller, K.J., Hermes, D., Honey, C.J., Hebb, A.O., Ramsey, N.F., Knight, R.T., Ojemann, J.G., Fetz, E.E., 2012. Human Motor Cortical Activity Is Selectively Phase-Entrained on Underlying Rhythms. PLoS Comput. Biol. 8, e1002655. doi:10.1371/journal.pcbi.1002655

Montemurro, M.A., Rasch, M.J., Murayama, Y., Logothetis, N.K., Panzeri, S., 2008. Phase-of-Firing Coding of Natural Visual Stimuli in Primary Visual Cortex. Curr. Biol. 18, 375–380. doi:10.1016/j.cub.2008.02.023

Nakhnikian, A., Ito, S., Dwiel, L.L., Grasse, L.M., Rebec, G.V., Lauridsen, L.N., Beggs, J.M., 2016. A novel cross-frequency coupling detection method using the generalized Morse wavelets. J. Neurosci. Methods 269, 61–73. doi:10.1016/j.jneumeth.2016.04.019

Newman, E.L., Gillet, S.N., Climer, J.R., Hasselmo, M.E., 2013. Cholinergic Blockade Reduces Theta-Gamma Phase Amplitude Coupling and Speed Modulation of Theta Frequency Consistent with Behavioral Effects on Encoding. J. Neurosci. 33, 19635–19646. doi:10.1523/JNEUROSCI.2586-13.2013

Ng, B.S.W., Logothetis, N.K., Kayser, C., 2013. EEG Phase Patterns Reflect the Selectivity of Neural Firing. Cereb. Cortex 23, 389–398. doi:10.1093/cercor/bhs031

O'Keefe, J., Recce, M.L., 1993. Phase relationship between hippocampal place units and the EEG theta rhythm. Hippocampus 3, 317–330.

Ojala, M., Garriga, G.C., 2010. Permutation tests for studying classifier performance. J. Mach. Learn. Res. 11, 1833–1863.

Ossandon, T., Jerbi, K., Vidal, J.R., Bayle, D.J., Henaff, M.-A., Jung, J., Minotti, L., Bertrand, O., Kahane, P., Lachaux, J.-P., 2011. Transient Suppression of Broadband Gamma Power in the Default-Mode Network Is Correlated with Task Complexity and Subject Performance. J. Neurosci. 31, 14521–14530. doi:10.1523/JNEUROSCI.2483-11.2011

Ossandon, T., Vidal, J.R., Ciumas, C., Jerbi, K., Hamame, C.M., Dalal, S.S., Bertrand, O., Minotti, L., Kahane, P., Lachaux, J.-P., 2012. Efficient "Pop-Out" Visual Search Elicits Sustained Broadband Gamma Activity in the Dorsal Attention Network. J. Neurosci. 32, 3414–3421. doi:10.1523/JNEUROSCI.6048-11.2012

Ozkurt, T.E., 2012. Statistically Reliable and Fast Direct Estimation of Phase-Amplitude Cross-Frequency Coupling. Biomed. Eng. IEEE Trans. On 59, 1943–1950.

Özkurt, T.E., Schnitzler, A., 2011. A critical note on the definition of phase–amplitude cross-

frequency coupling. J. Neurosci. Methods 201, 438–443. doi:10.1016/j.jneumeth.2011.08.014
Palva, S., Palva, J.M., 2007. New vistas for α-frequency band oscillations. Trends Neurosci. 30, 150–158. doi:10.1016/j.tins.2007.02.001
Paus, T., 2001. Primate anterior cingulate cortex: where motor control, drive and cognition interface. Nat. Rev. Neurosci. 2, 417–424. doi:10.1038/35077500
Pfurtscheller, G., Lopes da Silva, F.H., 1999. Event-related EEG/MEG synchronization and desynchronization: basic principles. Clin. Neurophysiol. 110, 1842–1857.
Pfurtscheller, G., Graimann, B., Huggins, J.E., Levine, S.P., Schuh, L.A., 2003. Spatiotemporal patterns of beta desynchronization and gamma synchronization in corticographic data during self-paced movement. Clin. Neurophysiol. 114, 1226–1236. doi:10.1016/S1388-2457(03)00067-1
Pittman-Polletta, B., Hsieh, W.-H., Kaur, S., Lo, M.-T., Hu, K., 2014. Detecting phase-amplitude coupling with high frequency resolution using adaptive decompositions. J. Neurosci. Methods 226, 15–32. doi:10.1016/j.jneumeth.2014.01.006
Rektor, I., Sochůrková, D., Bocková, M., 2006. Intracerebral ERD/ERS in voluntary movement and in cognitive visuomotor task. Prog. Brain Res. 159, 311–330. doi:10.1016/S0079-6123(06)59021-1
Roach, B.J., Mathalon, D.H., 2008. Event-related EEG time-frequency analysis: an overview of measures and an analysis of early gamma band phase locking in schizophrenia. Schizophr. Bull. 34, 907–926.
Saleh, M., Reimer, J., Penn, R., Ojakangas, C.L., Hatsopoulos, N.G., 2010. Fast and Slow Oscillations in Human Primary Motor Cortex Predict Oncoming Behaviorally Relevant Cues. Neuron 65, 461–471. doi:10.1016/j.neuron.2010.02.001
Sauseng, P., Klimesch, W., 2008. What does phase information of oscillatory brain activity tell us about cognitive processes? Neurosci. Biobehav. Rev. 32, 1001–1013. doi:10.1016/j.neubiorev.2008.03.014
Schnitzler, A., Gross, J., 2005. Normal and pathological oscillatory communication in the brain. Nat. Rev. Neurosci. 6, 285–296. doi:10.1038/nrn1650
Schwartz, A.B., 2016. Movement: How the Brain Communicates with the World. Cell 164, 1122–1135. doi:10.1016/j.cell.2016.02.038
Sherman, M.T., Kanai, R., Seth, A.K., VanRullen, R., 2016. Rhythmic Influence of Top–Down Perceptual Priors in the Phase of Prestimulus Occipital Alpha Oscillations. J. Cogn. Neurosci.
Snyder, L.H., Batista, A.P., Andersen, R.A., 1997. Coding of intention in the posterior parietal cortex. Nature 386, 167–170. doi:10.1038/386167a0
Soto, J.L., Jerbi, K., 2012. Investigation of cross-frequency phase-amplitude coupling in visuomotor networks using magnetoencephalography, in: Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE. IEEE, pp. 1550–1553.
Stephan, K.M., Fink, G.R., Passingham, R.E., Silbersweig, D., Ceballos-Baumann, A.O., Frith, C.D., Frackowiak, R.S., 1995. Functional anatomy of the mental representation of upper extremity movements in healthy subjects. J. Neurophysiol. 73, 373–386.
Talairach, J., Tournoux, P., 1993. Referentially oriented cerebral MRI anatomy: an atlas of stereotaxic anatomical correlations for gray and white matter. Thieme.
Tallon-Baudry, C., Bertrand, O., Delpuech, C., Pernier, J., 1996. Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in human. J. Neurosci. 16, 4240–4249.
Tort, A.B.L., Komorowski, R., Eichenbaum, H., Kopell, N., 2010. Measuring Phase-Amplitude Coupling Between Neuronal Oscillations of Different Frequencies. J. Neurophysiol. 104, 1195–1210. doi:10.1152/jn.00106.2010
VanRullen, R., Busch, N., Drewes, J., Dubois, J., 2011. Ongoing EEG Phase as a Trial-by-Trial Predictor of Perceptual and Attentional Variability. Front. Psychol. 2, 60.
Voytek, B., 2010. Shifts in gamma phase–amplitude coupling frequency from theta to alpha over

posterior cortex during visual tasks. Front. Hum. Neurosci. 4. doi:10.3389/fnhum.2010.00191

Voytek, B., D'Esposito, M., Crone, N., Knight, R.T., 2013. A method for event-related phase/amplitude coupling. NeuroImage 64, 416–424. doi:10.1016/j.neuroimage.2012.09.023

Waldert, S., Pistohl, T., Braun, C., Ball, T., Aertsen, A., Mehring, C., 2009. A review on directional information in neural signals for brain-machine interfaces. J. Physiol.-Paris 103, 244–254. doi:10.1016/j.jphysparis.2009.08.007

Waldert, S., Preissl, H., Demandt, E., Braun, C., Birbaumer, N., Aertsen, A., Mehring, C., 2008. Hand Movement Direction Decoded from MEG and EEG. J. Neurosci. 28, 1000–1008. doi:10.1523/JNEUROSCI.5171-07.2008

Whittingstall, K., Logothetis, N.K., 2009. Frequency-Band Coupling in Surface EEG Reflects Spiking Activity in Monkey Visual Cortex. Neuron 64, 281–289. doi:10.1016/j.neuron.2009.08.016

Yanagisawa, T., Yamashita, O., Hirata, M., Kishima, H., Saitoh, Y., Goto, T., Yoshimine, T., Kamitani, Y., 2012. Regulation of Motor Representation by Phase–Amplitude Coupling in the Sensorimotor Cortex. J. Neurosci. 32, 15467–15475. doi:10.1523/JNEUROSCI.2929-12.2012

Quatrième partie

Étude 3 : décodage des directions de mouvement pendant et avant l'exécution de mouvement de membres supérieurs

Introduction

blabla Intro

Blabla étude

Cinquième partie

Étude 4 : optimisation des paramètres de la bande gamma

Introduction

blabla Intro

Blabla étude

Sixième partie

Étude 5 : décodage des émotions

Introduction

blabla Intro

Blabla étude

Discussion générale

Enfin : la conclusion générale!!!
Au cours de ce mémoire, nous avons développé un modèle ...

1. **Modélisation**
2. **Inférence statistique**

Perspectives

Dans la continuité directe de notre travail de thèse, nous pouvons ...

A Annexes

Mettre juste la page de rÃ©fÃ©rences de brainpipe

A.1 Comparatif de methodes PAC (Tort et al., 2010)

TABLE 1. *Summary of characteristics of the phase-amplitude coupling measures studied*

Phase-Amplitude Coupling Measure	Tolerance to Noise	Amplitude Independent	Sensitivity to Multimodality	Sensitivity to Modulation Width
Modulation index	Good	Yes	Good	Good
Heights ratio	Good	Yes	No discrimination	No
Mean vector length	Good	No	Restricted	Reasonable
Amplitude PSD	Low	No	Restricted	Good
Phase-locking value	Low	No*	Restricted	Low
Correlation measure	Low	No*	Restricted	Low
GLM measure	Low	No*	Restricted	Low
Coherence value	Low	No*	Restricted	Low

* Under the presence of noise.

FIGURE A.1 – *Comparatif de methodes PAC (Tort et al., 2010)*

A.2 Pipeline standard de classification

Figure A.2 – *Pipeline standard de classification*

A.3 Comparatif de classifieurs (Pedregosa et al., 2011)

Figure A.3 – *Comparatif de classifieurs (Pedregosa et al., 2011)*

A.4 Exemple de schéma d'implantation

Figure A.4 – *Exemple de schéma d'implantation*

Bibliographie

Bahramisharif, A., van Gerven, M. A. J., Aarnoutse, E. J., Mercier, M. R., Schwartz, T. H., Foxe, J. J., Ramsey, N. F., and Jensen, O. (2013). Propagating Neocortical Gamma Bursts Are Coordinated by Traveling Alpha Waves. *Journal of Neuroscience*, 33(48) :18849–18854.

Bastin, J., Deman, P., David, O., Gueguen, M., Benis, D., Minotti, L., Hoffman, D., Combrisson, E., Kujala, J., Perrone-Bertolotti, M., Kahane, P., Lachaux, J.-P., and Jerbi, K. (2016). Direct Recordings from Human Anterior Insula Reveal its Leading Role within the Error-Monitoring Network. *Cerebral Cortex*, page bhv352.

Bekaert, M. H., Botte-Lecocq, C., Cabestaing, F., Rakotomamonjy, A., et al. (2009). Les interfaces Cerveau-Machine pour la palliation du handicap moteur sévère. *Sciences et Technologies pour le Handicap*, 3(1) :95–121.

Berens, P. and others (2009). CircStat a MATLAB toolbox for circular statistics. *J Stat Softw*, 31(10) :1–21.

Bertrand, O., Bohorquez, J., and Pernier, J. (1994). Time-frequency digital filtering based on an invertible wavelet transform : an application to evoked potentials. *IEEE Transactions on Biomedical Engineering*, 41(1) :77–88.

Besserve, M., Jerbi, K., Laurent, F., Baillet, S., Martinerie, J., Garnero, L., et al. (2007). Classification methods for ongoing EEG and MEG signals. *Biol Res*, 40(4) :415–437.

Boser, B. E., Guyon, I. M., and Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152. ACM.

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7) :1145–1159.

Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2) :121–167.

Canolty, R. T., Edwards, E., Dalal, S. S., Soltani, M., Nagarajan, S. S., Kirsch, H. E., Berger, M. S., Barbaro, N. M., and Knight, R. T. (2006). High Gamma Power Is Phase-Locked to Theta Oscillations in Human Neocortex. *Science*, 313(5793) :1626–1628.

Canolty, R. T. and Knight, R. T. (2010). The functional role of cross-frequency coupling.

Chai, H. and Domeniconi, C. (2004). An evaluation of gene selection methods for multi-class microarray data classification. In *Proceedings of the Second European Workshop on Data Mining and Text Mining in Bioinformatics*, pages 3–10.

Cohen, M. X. (2008). Assessing transient cross-frequency coupling in EEG data. *Journal of Neuroscience Methods*, 168(2) :494–499.

Combrisson, E. and Jerbi, K. (2015). Exceeding chance level by chance : The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *Journal of Neuroscience Methods*, 250 :126–136.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3) :273–297.

Das, S. (2001). Filters, wrappers and a boosting-based hybrid for feature selection. In *ICML*, volume 1, pages 74–81. Citeseer.

de Hemptinne, C., Ryapolova-Webb, E. S., Air, E. L., Garcia, P. A., Miller, K. J., Ojemann, J. G., Ostrem, J. L., Galifianakis, N. B., and Starr, P. A. (2013). Exaggerated phase–amplitude coupling in the primary motor cortex in Parkinson disease. *Proceedings of the National Academy of Sciences*.

Demandt, E., Mehring, C., Vogt, K., Schulze-Bonhage, A., Aertsen, A., and Ball, T. (2012). Reaching Movement Onset- and End-Related Characteristics of EEG Spectral Power Modulations. *Frontiers in Neuroscience*, 6.

Ding, C. and Peng, H. (2005). Minimum Redundancy Feature Selection from Microarray Gene Expression Data.

Duda, R. O., Hart, P. E., and Stork, D. G. (2001). Pattern classification. 2nd. *Edition. New York*.

Dvorak, D. and Fenton, A. A. (2014). Toward a proper estimation of phase–amplitude coupling in neural oscillations. *Journal of Neuroscience Methods*, 225 :42–56.

Efron, B. and Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.

Engel, A. K., Moll, C. K. E., Fried, I., and Ojemann, G. A. (2005). Invasive recordings from the human brain : clinical insights and beyond. *Nature Reviews Neuroscience*, 6(1) :35–47.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of eugenics*, 7(2) :179–188.

Fix, E. and Hodges Jr, J. L. (1951). Discriminatory analysis-nonparametric discrimination : consistency properties. Technical report, DTIC Document.

Fukunaga, K. (1990). Introduction to statistical pattern classification.

Graimann, B., Allison, B., and Pfurtscheller, G. (2009). Brain–Computer Interfaces : A Gentle Introduction. In Graimann, B., Pfurtscheller, G., and Allison, B., editors, *Brain-Computer Interfaces*, pages 1–27. Springer Berlin Heidelberg, Berlin, Heidelberg.

Guyon, I. and Elisseeff, A. (2003). An introduction to variable and feature selection. *The Journal of Machine Learning Research*, 3 :1157–1182.

Hanakawa, T., Dimyan, M. A., and Hallett, M. (2008). Motor Planning, Imagery, and Execution in the Distributed Motor Network : A Time-Course Study with Functional MRI. *Cerebral Cortex*, 18(12) :2775–2788.

Hochberg, L. R., Bacher, D., Jarosiewicz, B., Masse, N. Y., Simeral, J. D., Vogel, J., Haddadin, S., Liu, J., Cash, S. S., van der Smagt, P., and Donoghue, J. P. (2012). Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature*, 485(7398) :372–375.

Huang, J. and Ling, C. X. (2005). Using AUC and accuracy in evaluating learning algorithms. *Knowledge and Data Engineering, IEEE Transactions on*, 17(3) :299–310.

Hyafil, A., Giraud, A.-L., Fontolan, L., and Gutkin, B. (2015). Neural Cross-Frequency Coupling : Connecting Architectures, Mechanisms, and Functions. *Trends in Neurosciences*, 38(11) :725–740.

Jerbi, K., Ossandón, T., Hamamé, C. M., Senova, S., Dalal, S. S., Jung, J., Minotti, L., Bertrand, O., Berthoz, A., Kahane, P., and Lachaux, J.-P. (2009). Task-related gamma-band dynamics from an intracerebral perspective : Review and implications for surface EEG and MEG. *Human Brain Mapping*, 30(6) :1758–1771.

Jervis, B. W., Nichols, M. J., Johnson, T. E., Allen, E., and Hudson, N. R. (1983). A fundamental investigation of the composition of auditory evoked potentials. *IEEE transactions on bio-medical engineering*, 30(1) :43–50.

Kahane, P., Landré, E., Minotti, L., Francione, S., and Ryvlin, P. (2006). The Bancaud and Talairach view on the epileptogenic zone : a working hypothesis. *Epileptic disorders : international epilepsy journal with videotape*, 8 :S16–26.

King, J.-R. and Dehaene, S. (2014). Characterizing the dynamics of mental representations : the temporal generalization method. *Trends in Cognitive Sciences*, 18(4) :203–210.

Kohavi, R. and others (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *IJCAI*, volume 14, pages 1137–1145.

Lachaux, J., Rudrauf, D., and Kahane, P. (2003). Intracranial EEG and human brain mapping. *Journal of Physiology-Paris*, 97(4-6) :613–628.

Lachaux, J.-P., Rodriguez, E., Martinerie, J., Varela, F. J., and others (1999). Measuring phase synchrony in brain signals. *Human brain mapping*, 8(4) :194–208.

Lajnef, T., Chaibi, S., Ruby, P., Aguera, P.-E., Eichenlaub, J.-B., Samet, M., Kachouri, A., and Jerbi, K. (2015). Learning machines and sleeping brains : Automatic sleep stage classification using decision-tree multi-class support vector machines. *Journal of Neuroscience Methods*, 250 :94–105.

Lakatos, P. (2005). An Oscillatory Hierarchy Controlling Neuronal Excitability and Stimulus Processing in the Auditory Cortex. *Journal of Neurophysiology*, 94(3) :1904–1911.

Latinne, P., Debeir, O., and Decaestecker, C. (2001). Limiting the number of trees in random forests. In *International Workshop on Multiple Classifier Systems*, pages 178–187. Springer.

Ling, C. X., Huang, J., and Zhang, H. (2003). AUC a statistically consistent and more discriminating measure than accuracy. In *IJCAI*, volume 3, pages 519–524.

Liu, J., Ranka, S., and Kahveci, T. (2008). Classification and feature selection algorithms for multi-class CGH data. *Bioinformatics*, 24(13) :i86–i95.

Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., Arnaldi, B., et al. (2007). A review of classification algorithms for EEG-based brain–computer interfaces. *Journal of neural engineering*, 4.

Mellinger, J., Schalk, G., Braun, C., Preissl, H., Rosenstiel, W., Birbaumer, N., and Kübler, A. (2007). An MEG-based brain-computer interface (BCI). *Neuroimage*, 36(3) :581.

Nakhnikian, A., Ito, S., Dwiel, L., Grasse, L., Rebec, G., Lauridsen, L., and Beggs, J. (2016). A novel cross-frequency coupling detection method using the generalized Morse wavelets. *Journal of Neuroscience Methods*, 269 :61–73.

Ojala, M. and Garriga, G. C. (2010). Permutation tests for studying classifier performance. *The Journal of Machine Learning Research*, 11 :1833–1863.

Ossandon, T., Jerbi, K., Vidal, J. R., Bayle, D. J., Henaff, M.-A., Jung, J., Minotti, L., Bertrand, O., Kahane, P., and Lachaux, J.-P. (2011). Transient Suppression of Broadband Gamma Power in the Default-Mode Network Is Correlated with Task Complexity and Subject Performance. *Journal of Neuroscience*, 31(41) :14521–14530.

Ozkurt, T. E. (2012). Statistically Reliable and Fast Direct Estimation of Phase-Amplitude Cross-Frequency Coupling. *Biomedical Engineering, IEEE Transactions on*, 59(7) :1943–1950.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn : Machine Learning in Python. *Journal of Machine Learning Research*, 12 :2825–2830.

Penny, W., Duzel, E., Miller, K., and Ojemann, J. (2008). Testing for nested oscillation. *Journal of Neuroscience Methods*, 174(1) :50–61.

Pfurtscheller, G., Muller-Putz, G., Scherer, R., and Neuper, C. (2008). Rehabilitation with Brain-Computer Interface Systems. *Computer*, 41(10) :58–65.

Rickert, J. (2005). Encoding of Movement Direction in Different Frequency Ranges of Motor Cortical Local Field Potentials. *Journal of Neuroscience*, 25(39) :8815–8824.

Talairach, J. and Tournoux, P. (1993). *Referentially oriented cerebral MRI anatomy : an atlas of stereotaxic anatomical correlations for gray and white matter*. Thieme.

Tallon-Baudry, C., Bertrand, O., Delpuech, C., and Pernier, J. (1997). Oscillatory γ-band (30–70 Hz) activity induced by a visual search task in humans. *The Journal of neuroscience*, 17(2) :722–734.

Tort, A. B. L., Komorowski, R., Eichenbaum, H., and Kopell, N. (2010). Measuring Phase-Amplitude Coupling Between Neuronal Oscillations of Different Frequencies. *Journal of Neurophysiology*, 104(2) :1195–1210.

Van Langhenhove, A., Bekaert, M. H., Cabestaing, F., N'Guyen, J. P., et al. (2008). Interfaces cerveau-ordinateur et rééducation fonctionnelle : étude de cas chez un patient hémiparésique. *Sciences et Technologies pour le Handicap*, 2(1) :41–54.

Vidaurre, C., Krämer, N., Blankertz, B., and Schlögl, A. (2009). Time domain parameters as a feature for EEG-based brain–computer interfaces. *Neural Networks*, 22(9) :1313–1319.

Vladimir, V. N. and Vapnik, V. (1995). The nature of statistical learning theory.

Voytek, B., D'Esposito, M., Crone, N., and Knight, R. T. (2013). A method for event-related phase/amplitude coupling. *NeuroImage*, 64 :416–424.

Waldert, S., Pistohl, T., Braun, C., Ball, T., Aertsen, A., and Mehring, C. (2009). A review on directional information in neural signals for brain-machine interfaces. *Journal of Physiology-Paris*, 103(3-5) :244–254.

Waldert, S., Preissl, H., Demandt, E., Braun, C., Birbaumer, N., Aertsen, A., and Mehring, C. (2008). Hand Movement Direction Decoded from MEG and EEG. *Journal of Neuroscience*, 28(4) :1000–1008.

Watrous, A. J., Deuker, L., Fell, J., and Axmacher, N. (2015). Phase-amplitude coupling supports phase coding in human ECoG. *eLife*, 4 :e07886.

Weinberger, K. Q., Blitzer, J., and Saul, L. K. (2005). Distance metric learning for large margin nearest neighbor classification. In *Advances in neural information processing systems*, pages 1473–1480.

Wieland, M. and Pittore, M. (2014). Performance Evaluation of Machine Learning Algorithms for Urban Pattern Recognition from Multi-spectral Satellite Images. *Remote Sensing*, 6(4) :2912–2939.

Worrell, G., Jerbi, K., Kobayashi, K., Lina, J., Zelmann, R., and Le Van Quyen, M. (2012). Recording and analysis techniques for high-frequency oscillations. *Progress in neurobiology*, 98(3) :265–278.

Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Yu, P. S., Zhou, Z.-H., Steinbach, M., Hand, D. J., and Steinberg, D. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems*, 14(1) :1–37.

Yanagisawa, T., Hirata, M., Saitoh, Y., Kato, A., Shibuya, D., Kamitani, Y., and Yoshimine, T. (2009). Neural decoding using gyral and intrasulcal electrocorticograms. *NeuroImage*, 45(4) :1099–1106.

Yanagisawa, T., Yamashita, O., Hirata, M., Kishima, H., Saitoh, Y., Goto, T., Yoshimine, T., and Kamitani, Y. (2012). Regulation of Motor Representation by Phase–Amplitude Coupling in the Sensorimotor Cortex. *The Journal of Neuroscience*, 32(44) :15467–15475.

Yu, L. and Liu, H. (2004). Redundancy based feature selection for microarray data. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 737–742. ACM.

Titre Décodage des intentions et des représentations motrices chez l'homme : analyse multi-échelle et application aux interfaces cerveau-machine

Résumé Le résumé en français (≈ 1000 caractères)

Mots-clés Les mots-clés en français

Title Le titre en anglais

Abstract Le résumé en anglais (≈ 1000 caractères)

Keywords Les mots-clés en anglais