ECoG signal processing for Brain Computer Interface with multiple degrees of freedom for clinical application

Marie-Caroline Schaeffer

▶ **To cite this version:**

Marie-Caroline Schaeffer. ECoG signal processing for Brain Computer Interface with multiple degrees of freedom for clinical application. Medical Physics [physics.med-ph]. Université Grenoble Alpes, 2017. English. NNT : 2017GREAS026 . tel-01763451

HAL Id: tel-01763451
https://theses.hal.science/tel-01763451

Submitted on 11 Apr 2018

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

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

Communauté
UNIVERSITÉ Grenoble Alpes

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTE UNIVERSITE GRENOBLE ALPES

Spécialité : MBS – Modèles, méthodes et algorithmes en biologie, santé et environnement

Arrêté ministériel : 25 mai 2016

Présentée par

Marie-Caroline SCHAEFFER

Thèse dirigée par **Tetiana AKSENOVA (EDISCE), CEA**

préparée au sein du **Laboratoire CLINATEC**
dans **l'École Doctorale Ingénierie pour la Santé, la Cognition et l'Environnement**

Traitement du signal ECoG pour Interface Cerveau Machine à grand nombre de degrés de liberté pour application clinique

Thèse soutenue publiquement le **6 Juin 2017**,
devant le jury composé de :

Prof. Alim-Louis BENABID
Président
Dr. Laurent BOUGRAIN
Membre
Prof. François CABESTAING
Rapporteur
Dr. Marco CONGEDO
Membre
Prof. Saïd MOUSSAOUI
Rapporteur
Prof. Laura SACERDOTE
Membre

Acknowledgements

I feel very grateful to the members of my committee, Prof. Alim-Louis BENABID, Dr. Laurent BOUGRAIN, Dr. Marco CONGEDO and Prof. Laura SACERDOTE. My special thanks go to Prof. François CABESTAING and Prof. Saïd MOUSSAOUI for having accepted to act as reporters of this doctoral work.

My deepest gratitude goes to Dr. Tetiana AKSENOVA who has supervised this doctoral work. I wouldn't have been able to complete this work without her constant guidance, insight, availability, patience and kindness. It has been a great privilege to work and progress under her supervision.

I would also like to thank Guillaume CHARVET who co-supervised this doctoral work, and Prof. Alim-Louis BENABID and Corinne MESTAIS who head CLINATEC's BCI project. Their suggestions have always been helpful and illuminating.

It is a great pleasure to me to extend my thanks to all the associates of CLINATEC with whom I worked, had insightful discussions with or with whom I simply enjoyed meals or coffee breaks. Their presence has helped me along more than once.

Finally, I would like to thank my family and friends for their unwavering support throughout these last three years.

Preface

The present doctoral work, which has been supervised by Dr. T. AKSENOVA and co-supervised by G. CHARVET, has been prepared at CLINATEC, an Edmond J. SAFRA Biomedical research centre located on the MINATEC Campus at CEA/LETI in Grenoble. CLINATEC associates medical research and technological innovation to translate new solutions to patients, with the purpose of accelerating the development and clinical validation of innovative medical devices. This doctoral work has been completed within the framework of CLINATEC's motor Brain-Computer-Interface project, which is supervised by Prof. A.-L. BENABID, C. MESTAIS and G. CHARVET. The goal of this project is to bring the proof of concept that it is feasible for a tetraplegic subject to control complex effectors, for example a 4-limb exoskeleton, thanks to the monitoring and decoding of his brain activity. CLINATEC's "BCI and Tetraplegia" 5-year clinical trial has recently been authorized by the French regulatory agencies. Signal processing challenges specific to the clinical deployment of CLINATEC's Brain-Computer-Interface system, namely asynchronous mono-limb decoding, asynchronous sequential multi-limb decoding and decoding accuracy during active states, have been addressed in the present doctoral thesis.

Parts of the presented work have been published in journals or conference proceedings, or exposed during oral presentations:

Journal

Schaeffer, M. C., & Aksenova, T. (2017). Switching Markov Decoders for Asynchronous Trajectory Reconstruction from ECoG signals in Monkeys for BCI Applications. Journal of Physiology-Paris. doi: 10.1016/j.jphysparis.2017.03.002; Published online: 10 March 2017.

International conference proceedings

Schaeffer, M. C., & Aksenova, T. (2016, September). Hybrid Trajectory Decoding from ECoG Signals for Asynchronous BCIs. In International Conference on Artificial Neural Networks, Barcelona, Spain (pp. 288-296). Springer International Publishing.

Patent pending

Schaeffer, M. C., & Aksenova, T. (June 2016). Interface neuronale directe à décodage continu utilisant un mélange markovien d'experts.

Oral communications

M.C. Schaeffer, T. Aksenova. Markovian Mixture of Experts for 3D wrist movement decoding from ECoG recordings. GDR 2904 Multielectrode systems for Neuroscience: 6th annual meeting, 5-8 Jan 2016, Grenoble-Autrans. (speaker)

M.C. Schaeffer, E. Labyt, V. Rohu, N. Tarrin, I. Vergara, S. Cokgungor, A. Eliseyev, G. Charvet, C. Mestais, A.L. Benabid, T. Aksenova. Hand movement decoding from magnetoencephalographic signals for BCI applications. Société de Neurophysiologie Clinique de Langue Française (SNCLF) 2015, 1-3 Juillet 2015, Grenoble.

M.C. Schaeffer, S. Cokgungor, N. Tarrin, V. Rohu, E. Labyt, C. Mestais, A.L. Benabid, T. Aksenova. Decoding of 3D hand movements from Magnetoencephalography recordings of brain neuronal activity: preliminary results. BioHealth Computing Workshop, 1-3 Septembre 2014, Grenoble. (speaker)

Résumé

Introduction Les Interfaces Cerveau-Machine (ICM) sont des systèmes qui permettent à des patients souffrant d'un handicap moteur sévère d'interagir avec leur environnement en utilisant leur activité cérébrale pour contrôler des effecteurs extérieurs. Plusieurs étapes sont généralement nécessaires pour convertir l'activité cérébrale du patient en commandes permettant de contrôler un effecteur extérieur. Des caractéristiques spécifiques aux intentions de l'utilisateur sont tout d'abord extraites de son activité cérébrale, préalablement acquise et digitalisée. Un décodeur est appliqué sur ces caractéristiques cérébrales, et permet de les convertir en estimations des intentions du sujet. Après une étape optionnelle de post-traitement susceptible d'améliorer la qualité des estimations, les intentions décodées du sujet sont converties en commandes utilisées pour contrôler le ou les effecteur(s) de l'ICM, par exemple des orthèses ou des prothèses de membres inférieurs ou supérieurs dans le cas d'ICM motrices.

L'objectif de l'ICM motrice de CLINATEC est de permettre à des patients tétraplégiques de recouvrer de façon chronique une indépendance motrice en modulant leur activité électrocorticographique (ECoG) pour contrôler des effecteurs complexes tels qu'un exosquelette 4 membres. Des challenges spécifiques au déploiement clinique d'ICMs multi-effecteurs ont été étudiés dans la présente thèse de doctorat.

Un problème majeur des ICMs cliniques est la capacité à proposer aux utilisateurs un contrôle asynchrone sur l'effecteur. Contrairement aux ICMs synchrones qui sont périodiquement contrôlables par les utilisateurs et requièrent donc la présence d'un opérateur pour activer et désactiver le système, les ICMs asynchrones sont disponibles en continu. Lorsqu'une ICM fonctionne en mode asynchrone, il est particulièrement souhaitable de limiter des activations erronées de l'effecteur pendant les périodes dites de Non-Contrôle (NC). Un deuxième challenge résulte de la possible présence d'un effecteur multi-membres, qui rend nécessaire la généralisation du contrôle asynchrone mono-membre au cas multi-membres. Une stratégie d'activation séquentielle a été considérée dans la présente thèse, avec le but d'améliorer la robustesse du système et de faciliter le contrôle cérébral. Au-delà de la limitation des activations erronées du système, il est nécessaire d'éviter des mouvements parallèles résiduels d'effecteurs momentanément non-contrôlés. Finalement, la capacité des utilisateurs à exécuter des mouvements contrôlés par leur activité cérébrale est compromise quand les estimations de leurs intentions de mouvements ne sont pas suffisamment proches de leurs véritables intentions. La précision du décodage pendant les périodes dites de Contrôle Intentionnel (Intentional Control, IC) est donc essentielle.

Décodeurs pour ICM motrices précises et asynchrones Différents outils algorithmiques sont utilisés dans la communauté des ICMs pour assurer une conversion pertinente des signaux cérébraux des sujets en estimations de leurs intentions, tout particulièrement des modèles permettant d'estimer la valeur de variables discrètes (classifieurs) ou continues (modèles de régression ou filtres bayésiens) à partir des

signaux cérébraux. Deux stratégies de décodage sont principalement exploitées par les ICMs motrices. La première consiste à demander au sujet d'effectuer des tâches mentales (tout particulièrement des imaginations motrices ou des tâches cognitives, par exemple des calculs mentaux), qui sont généralement chacune associées à une direction possible de l'effecteur. La seconde vise à directement extraire des signaux cérébraux les paramètres cinématiques caractérisant la trajectoire désirée de l'effecteur. Les décodeurs correspondants sont ici désignés sous le terme de décodeurs cinématiques. Il a été suggéré que les décodeurs cinématiques permettent d'obtenir des ICMs motrices précises et intuitives. Les décodeurs cinématiques exploitent des décodeurs continus, principalement des modèles linéaires statiques ou des filtres bayésiens comme le filtre de Kalman.

Ces décodeurs cinématiques classiques échouent généralement à fournir des estimations neutres (i.e., associées à une vitesse nulle) pendant les états NC. De ce fait, des stratégies ont été spécifiquement élaborées pour gérer les états NC, c'est-à-dire pour limiter des activations erronées de l'ICM pendant les états NC. Deux approches principales ont émergé pour assurer un décodage asynchrone mono-membre performant, à savoir la gestion des états NC au niveau du décodeur cinématique ou d'un potentiel opérateur de post-traitement. Quand la gestion des états NC est faite au niveau du décodeur, elle est généralement réalisée via l'introduction de non-linéarités dans le décodeur. L'utilisation de modèles linéaires par morceaux a particulièrement été rapportée. Ces modèles par morceaux se basent sur une variable latente discrète pour introduire des non-linéarités dans un décodeur générique, par exemple un modèle de regression ou un filtre bayésien. Dans le cas de mixtures de régressions, des décodeurs discrets sont utilisés pour estimer les valeurs prises par la variable latente à partir des signaux cérébraux.

Un unique modèle est en revanche entraîné et/ou appliqué sur les signaux cérébraux NC et IC quand la gestion des états NC est faite au niveau du post-traitement. De nouveau, la valeur d'une variable latente discrète est généralement utilisée pour sélectionner l'opérateur de post-traitement à appliquer sur les estimations de mouvements.

Le décodage asynchrone multi-membres a été rarement étudié dans les études ICMs. L'utilisation d'un modèle par morceaux a néanmoins été rapportée pour le décodage de mouvements séquentiels multi-doigts.

Finalement, les tentatives pour améliorer la précision du décodage pendant les périodes IC se concentrent généralement sur l'utilisation de modèles non-linéaires, en particulier des décodeurs non-linéaires classiques (par exemple, des réseaux de neurones) ou linéaires par morceaux. Des méthodes de post-traitement sont également utilisées pour améliorer certaines caractéristiques des estimations des paramètres cinématiques d'intérêt, par exemple le bruit haute fréquence (lissage) ou la précision spatiale.

Méthodes Des modifications de la dépendence entre signaux cérébraux et paramètres cinématiques ont été constatées dans plusieurs études. Ces modifications ont par

exemple été observées quand des sujets se trouvaient en états NC ou IC, lorsqu'ils effectuaient des movements unimanuels ou bimanuels, ou encore au cours de différentes phases d'un mouvement du bras vers des cibles. Il a par conséquent été proposé dans la présente thèse d'utiliser un modèle par morceaux, plus précisemment un modèle markovien linéaire par morceaux (Markov Switching Linear Model, MSLM), pour répondre au challenge d'un contrôle asynchrone mono- and multi-membre précis. Ce modèle présente trois points clefs.

Le MSLM commute entre plusieurs modèles, à savoir un modèle NC et un ou plusieurs modèles ICs. Contrairement aux approches basées sur le post-traitement qui exploitent un unique modèle continu, le MSLM prend donc en compte de potentielles modifications d'une dépendance linéaire entre signaux cérébraux et paramètres cinématiques. Les études de la littérature mentionnées plus haut suggèrent que la commutation entre modèles est susceptible d'être utile à la fois pour intégrer la gestion des états NC (un modèle NC, un modèle IC) et de mouvements multi-membres séquentiels (un modèle NC, un modèle IC par membre) et pour améliorer la précision des estimations des paramètres cinématiques pendant les états IC (un modèle NC, un modèle IC par phase de mouvement). La pertinence de chaque modèle continu du MSLM, i.e. la valeur d'une variable latente discrète indiquant quel modèle continu est approprié, est déduite à chaque instant des données cérébrales. Elle permet de déduire une règle probabilistique utilisée pour pondérer les modèles continus du MSLM.

Le MSLM a été développé dans le cadre des modèles de régression statiques, et plus précisément comme une extension des Mélanges d'Experts (ME). Un modèle linéaire discriminant entre les caractéristiques cérébrales et les paramètres cinématiques est conditionné par l'état courant d'une variable latente discrète. La valeur de cette variable latente est directement déduite des signaux cérébraux et de leur distribution durant chaque état possible. Cette propriété distingue le MSLM des mixtures de filtres bayésiens dont l'utilisation a également été considérée dans la littérature. Ces filtres sont basés sur des modèles génératifs, et la valeur de la variable latente est déduite de l'adéquation entre chaque filtre et l'historique des signaux cérébraux. De surcroît, le MSLM est capable de traiter des données de grande dimension sans qu'il soit nécessaire de réaliser une étape préliminaire de réduction de dimension, laquelle est généralement nécessaire pour des modèles génératifs.

A l'opposé des mixtures de modèles de régression précédemment utilisées dans des études ICM, le MSLM réalise une détection d'état dynamique pour limiter les fausses activations de l'effecteur. Plus précisément, il est supposé que la séquence d'états latents (par exemple, NC et IC) est générée par une chaîne de Markov d'ordre 1. Cette hypothèse est susceptible de réduire le nombre d'états mal détectés, mais surtout d'améliorer les caractéristiques temporelles des fausses activations, c'est-à-dire de privilégier des fausses activations plus longues mais également plus rares. Un Modèle de Markov Caché (Hidden Markov Model, HMM) est utilisé pour estimer la valeur de l'état discret. Le modèle des Mixtures d'Experts a été modifié de façon à gérer des détections d'états dynamiques plutôt que statiques.

Des procédures d'apprentissage supervisé et non supervisé ont été présentées

dans la présente thèse. La procédure d'apprentissage supervisé est pertinente quand la valeur prise par la variable latente discrète est connue dans le jeu de données d'entraînement, ce qui est par exemple le cas pour les états NC et IC. L'apprentissage non-supervisé permet d'exploiter une variable latente discrète sans connaissance a priori sur sa valeur dans le jeu de données d'entraînement. Il peut en particulier être utilisé pour associer différents modèles IC à chacune des phases d'un mouvement complexe. Cette procédure d'apprentissage non-supervisé est basée sur l'algorithme espérance-maximisation (Expectation-Maximization, EM), et généralise à la fois l'algorithme EM utilisé pour entraîner des HMMs et celui qui permet d'identifier les paramètres de MEs.

Validation Deux jeux de données ont été utilisés pour explorer la capacité du MSLM à réaliser du décodage ECoG asynchrone mono- et multi-membres précis. Ces deux jeux de données sont libres d'accès. Le premier jeu de données est composé de données précliniques acquises pendant que des Primates Non-Humains réalisaient des mouvements asynchrones avec un bras (mono-membre) pour atteindre des cibles. Le second jeu de données rassemble des données cliniques enregistrées pendant que des patients exécutaient des mouvements de doigts séquentiels (flexions et extensions).

Deux décodeurs correspondant à des stratégies précédemment proposées pour intégrer la gestion des états NC dans des décodeurs cinématiques ont été choisis pour évaluer la performance comparative du MSLM. Le premier décodeur alternatif a été proposé pour mesurer l'intérêt d'utiliser une mixture de modèles. Il s'agit d'un modèle linéaire générique (filtre de Wiener avec post-processing markovien, MpWF) dont les estimations sont post-traitées pour limiter des activations erronées durant les périodes NC. Le second modèle a été choisi pour vérifier que le cadre discriminant choisi pour développer le MSLM est plus performant que le cadre génératif. Un filtre de Kalman commutatif (Switching Kalman Filter, SKF) a été choisi pour mener cette comparaison entre modélisations discriminante et générative. Les SKFs commutent entre plusieurs filtres de Kalman dont la pertinence est estimée à partir de leur adéquation avec l'historique des signaux cérébraux.

Un ensemble d'indicateurs de performance a été formé pour évaluer la capacité du MSLM, du MpWF et du SKF à réaliser du décodage asynchrone précis. La qualité de la gestion des états NC est généralement mesurée au moyen d'indicateurs basés sur la matrice de confusion. Il a ici été proposé d'additionellement considérer des indicateurs de performance dits "par blocs", à savoir le nombre de blocs de fausses activations et leur durée moyenne. Ces indicateurs par blocs prennent en compte la dynamique des fausses activations, et sont susceptibles de davantage refléter la qualité de la gestion des états de NC telle que perçue par l'utilisateur pendant des sessions de contrôle neuronal asynchrone en boucle fermée. Les indicateurs de performance associés à la reconstruction de paramètres cinématiques pendant les périodes IC portent généralement sur leur précision spatiale et temporelle (Coefficient de Corrélation de Pearson, distance euclidienne ou de Manhattan). De légères désynchronisations entre les trajectoires observées et estimées, c'est-à-dire de petites

erreurs temporelles, sont susceptibles d'avoir un impact important sur ces indicateurs de performance. De ce fait, il est ici proposé de calculer ces indicateurs à la fois entre les trajectoires mesurées et estimées originales (désynchronisées) et entre les trajectoires synchronisées. Les indicateurs calculés sur les trajectoires synchronisées se concentrent sur l'erreur spatiale de l'estimation de trajectoire. Une approche a été mise au point pour calculer des trajectoires synchronisées dans l'échelle de temps originelle de la trajectoire mesurée. Finalement, une procédure d'analyse statistique a été choisie pour mesurer la significativité des différences de performance constatées entre les trois décodeurs considérés.

Implémentation et résultats Des décodeurs MSLM, MpWF et SKF ont été implémentés pour réaliser des reconstructions de trajectoires asynchrones mono-membre et multi-membres sur les jeux de données précliniques et cliniques, respectivement. Des représentations temps-fréquence-espace de grande dimension ont tout d'abord été extraites des signaux cérébraux. Les paramètres de décodeurs avec 2 états (un état NC et un état IC) ont été identifiés au moyen de procédures d'entraînement supervisé pour les données mono-membre précliniques. Dans le cas de sessions associées à des mouvements complexes, un MSLM avec deux états IC a également été entraîné via l'algorithme d'apprentissage non-supervisé proposé dans la présente thèse. Des décodeurs avec 6 états (un état NC et cinq états IC) ont été calibrés en utilisant la procédure d'apprentissage supervisé pour les données multi-membres (multi-doigts) cliniques. Des études préliminaires ont été réalisées pour optimiser les composants du MSLM, c'est-à-dire ses modèles continus et le décodeur discret chargé d'estimer la valeur de la variable latente permettant de combiner les estimations des modèles continus. La méthode de réduction de dimension utilisée pour rendre le SKF compatible avec du décodage temps réel a également été optimisée.

Le décodeur MSLM a permis d'obtenir une gestion des états NC de meilleure qualité que le MpWF et SKF à la fois pour le décodage asynchrone mono-membre et multi-membres. Une amélioration de la précision des estimations des paramètres cinématiques a été constatée pour les données cliniques multi-membres (multi-doigts). La précision de la reconstruction des mouvements complexes observés dans certaines sessions d'acquisition pré-cliniques a été améliorée quand plusieurs modèles IC ont été identifiés.

Discussion La pertinence du modèle proposé pour réaliser un décodage asynchrone précis de mouvements mono-membre et multi-membres a été validée sur des données précliniques et cliniques. Plusieurs pistes d'amélioration pourront être explorées dans le futur, notamment l'utilisation d'un décodeur discret hiérarchique, le développement d'une procédure d'apprentissage non-supervisé variationelle pour une optimisation automatique du nombres de modèles du MSLM et celui d'une méthode d'apprentissage récursif pour une adaptation en ligne des paramètres du MSLM. Une étape de validation essentielle va également être menée à bien lors de l'essai clinique de l'ICM motrice de CLINATEC "BCI et tétraplégie", à savoir

l'évaluation de la performance du MSLM lors de sessions de contrôle en boucle fermée d'un exosquelette par un sujet tétraplégique.

Abstract

Brain-Computer Interfaces (BCI) are systems that allow severely motor-impaired patients to use their brain activity to control external devices, and thereby to interact with their environment. Several processing procedures are usually carried out to translate the user's neuronal activity into commands for effector control. Features specifically related to the user's intentions are first extracted from measures of the user's brain activity. A decoder is then applied to estimate the user's intention from these brain features. After being optionally enhanced by post-processing algorithms, this estimate of the user's intention is converted into commands used to drive the BCI effector(s), e.g. upper- or lower-limb orthoses or prostheses in the particular case of motor BCIs.

The goal of CLINATEC's motor BCI project is to bring the proof of concept that it is feasible for a tetraplegic subject to control complex effectors, for example a 4-limb exoskeleton, thanks to the monitoring and decoding of his electrocorticographic (ECoG) brain activity. Challenges specific to the clinical deployment of CLINATEC's BCI system are addressed in the present doctoral thesis. A major issue for BCI clinical applications is the ability to provide users with accurate asynchronous control over the effector. Unlike synchronous BCIs which are periodically controllable by users, asynchronous BCI decoders are continuously available. In asynchronous settings, the limitation of spurious effector activations during No-Control (NC) periods is particularly desirable. A second challenge arises from the multi-limb effector embedded into CLINATEC's BCI system, namely the extension of accurate asynchronous decoding to multi-limb control. A sequential upper-/lower-limb activation strategy has been chosen for CLINATEC's BCI system with the aim of improving the system's stability and of facilitating neural control. The corresponding decoding objective lies in avoiding parallel, residual movements of the momentarily non-controlled limbs. Finally, the ability of users to execute brain-controlled movements is compromised when the estimates of the user's intention are not sufficiently correlated with his true intentions. Decoding accuracy during Intentional Control (IC) periods is thus crucial.

Kinematic decoders, i.e. decoders which extract continuously-valued kinematic parameters from the neural signals, are generally expected to be precise and intuitive for motor BCIs. Recent studies have suggested they may be applicable for ECoG decoding. In the present doctoral thesis, a switching decoder, namely a Markov Switching Linear Model (MSLM), has been proposed for the task of accurate, asynchronous sequential multi-limb kinematic ECoG decoding. The MSLM strategy to support both NC/IC periods and multi-limb control consists of switching between NC and (possibly multiple) IC linear models to take into account the specific behaviours expected during NC and IC states. The relevance of each model, i.e. of each state, is estimated from the neural data and is used to combine the available continuous models. Switches between IC models are considered to address the issue of sequential multi-limb effector control (one IC model per limb) as well as to improve decoding accuracy during IC periods (one IC model per movement phase).

The MSLM has been developed in the frame of static regression models, and more precisely as an extension of Mixture of Experts (ME) models. A linear model between neural features and kinematic parameters is conditioned on the current state. A probabilistic rule is deduced from the likelihood of each possible state, and is used to weight the available regression models. In contrast with the previously reported BCI switching regression models, the MSLM performs dynamic state detection to limit spurious effector activations. More specifically, the sequence of NC and (possibly multiple) IC states is assumed to be generated by a first-order Markov chain. A Hidden Markov Model-based (HMM) discrete decoder is used for state estimation. Both supervised and unsupervised training procedures are presented in the present dissertation. The proposed unsupervised training procedure is based on the Expectation-Maximization (EM) algorithm, and extends both HMM- and ME- well known EM training algorithms. Unsupervised training permits to exploit internal states without precise a priori knowledge on their value in the training data set. In particular, it can be used to associate different IC models with different phases of complex movements.

The performance of the MSLM decoder was assessed for two decoding tasks, namely asynchronous wrist trajectory reconstruction (publicly available preclinical data set) and multi-limb (multi-finger) trajectory reconstruction (publicly available clinical data set). The MSLM was compared to a Wiener Filter with Markovian post-processing (MpWF) decoder and to a Switching Kalman Filter (SKF). These two decoders were chosen so as to represent two alternative strategies previously exploited for the considered tasks. The supervised MSLM decoder was found to outperform both the SKF and the post-processed MpWF decoder in terms of state detection accuracy and/or continuous decoding during IC states, thus reducing the number of spurious activations during asynchronous mono- and multi-limb decoding and/or improving decoding accuracy during IC periods. Evaluation of the MSLM decoder will be performed in CLINATEC's coming clinical trial.

Abbreviations and notations

Abbreviations

ACC	Accuracy
ANN	Artificial Neural Network
AP	Action Potential
AR	Auto-Regressive
AUC	Area Under the Curve
CCWT	Complex-Continuous Wavelet Transform
CV	Cross-Validation
BCI	Brain Computer Interface
CRF	Conditional Random Field
CSP	Common Spatial Pattern
CTPR	Correct True Positive Rate
DBN	Dynamic Bayesian Network
DoF	Degree of Freedom
DTW	Dynamic Time Warping
ECG	Electrocardiographic, Electrocardiography, Electrocardiogram
ECoG	Electrocorticographic, Electrocorticography
EEG	Electroencephalographic, Electroencephalography
EM	Expectation-Maximization
EOG	Electrooculographic, Electrooculography, Electrooculogram
ERP	Event-Related Potential
FN	False Negative
FP	False Positive
FPR	False Positive Rate
FES	Functional Electrical Stimulation
fMRI	Functional Magnetic Resonance Imaging
fNIRS	Functional Near-Infrared Spectroscopy
GAM	Generalized Additive Model
GLM	Generalized Linear Model
HMM	Hidden Markov Model
IC	Intentional Control
ICA	Independent Component Analysis
IOHMM	Input-Output Hidden Markov Model
KF	Kalman Filter
kNN	k-Nearest Neighbours
LASSO	Least Absolute Shrinkage and Selection Operator
LDA	Linear Discriminant Analysis
LFP	Local Field Potential
LGF	Laplace-Gaussian Filtering
LMP	Local Motor Potential
LR	Logistic Regression

M1	Primary Motor Cortex
MAP	Maximum A Posteriori
ME	Mixture of Experts
MEA	Microelectrode Array
MEG	Magnetoencephalographic, Magnetoencephalography
ML	Maximum Likelihood
MLLS	Minimal Length Least Squares
MpWF	Markovian post-processed Wiener Filter
MRP	Movement-Related Potential
MSLM	Markov Switching Linear Model
MUA	Multi Unit Activity
NC	No-Control
NHP	Non-Human Primate
NMAE	Normalized Mean Absolute Error
NRMSE	Normalized Root Mean Squared Error
OLS	Ordinary Least Squares
PCA	Principal Component Analysis
PCC	Pearson Correlation Coefficient
PCR	Principal Component Regression
PLS	Partial Least Squares
PLSR	Partial Least Squares Regression
PLV	Phase-Locking-Value
PM	Premotor Cortex
PPC	Posterior Parietal Cortex
PRESS	Predicted Error Sum of Squares
pWF	Post-processed Wiener Filter
QDA	Quadratic Discriminant Analysis
RBF	Radial Basis Function
SCP	Slow Cortical Potential
SCP	Sensorimotor Rhythm
SL	Surface Laplacian
SLM	Switching Linear Model
SKF	Switching Kalman Filter
SMA	Supplementary Motor Area
SNR	Signal-to-Noise Ratio
SSVEP	Steady-State Visual Evoked Potential
STFT	Short Time Fourier Transform
SUA	Single Unit Activity
SVM	Support-Vector-Machine
SVR	Support-Vector-Regression
TN	True Negative
TNR	True Negative Rate
TP	True Positive
TPR	True Positive Rate

UKF	Unscented Kalman Filter
WF	Wiener Filter

Mathematical Notations

Matrices are denoted by bold upper-case romans (e.g., $\mathbf{X}$), vectors by bold lower-case romans (e.g., $\mathbf{x}$) and scalar by lower-cases italics (e.g., x). x_{ij} and x_i refer to the $(i,j)^{th}$ entry of matrix $\mathbf{X}$ and to the i^{th} element of vector $\mathbf{x}$, respectively.

Contents

Chapter 1

Introduction

Contents

1.1 Brain-Computer Interfaces

1.1.1 BCI overview

Brain-Computer Interfaces (BCI), also referred to as Brain-Machine Interfaces or Direct Brain Interfaces [Graimann et al., 2009], are systems which permit users to utilize their brain activity to control external devices without using their natural neuromuscular pathways [Leuthardt et al., 2006b] [Mak and Wolpaw, 2009]. BCIs are particularly being investigated for the sake of severely motor-impaired patients. Common causes for serious motor dysfuntions are neuromuscular disorders

like amyotrophic lateral sclerosis or cerebral palsy, spinal cord injury (paraplegia, tetraplegia) and stroke [Lebedev and Nicolelis, 2006]. The independence and well-being of severely motor-impaired patients are threatened by their inability to control muscles required for the execution of critical motor tasks, for example bowel control, respiration, limb movements or verbal communication. BCIs aim at overcoming some of these disabilities by establishing a new communication pathway between the patient's brain and an effector (e.g., a robotic arm, a speller or a wheelchair). Brain effector control [Mak and Wolpaw, 2009] is expected to help patients recovering the ability to interact with their environment. It is also anticipated that neurorehabilitation programs will benefit from the introduction of BCI-based approaches [Lew, 2012].

1.1.2 Brain pattern elicitation

The input of a BCI system is a measure of the user's brain activity. BCIs are based on the interpretation of brain activity patterns. Specific and measurable brain activity patterns must thus be generated to trigger the execution of a particular action by the effector. Different strategies are used to elicit this recognizable brain activity.

A distinction is drawn between externally- and internally-paced brain patterns [Waldert et al., 2009]. Externally-paced brain patterns are responses evoked by a visual, auditory or somatosensory stimulus (Evoked Potential) or by an event (Event-Related Potential, ERP). Steady-State Visual Evoked Potentials (SSVEP) and the P300 wave are two examples of externally-paced brain patterns regularly exploited for BCI control [Waldert et al., 2009]. The use of SSVEPs has for example been reported for spelling devices [Nakanishi et al., 2014a] [Yin et al., 2015], hand orthosis [Ortner et al., 2011] and wheelchair control [Müller et al., 2011] [Nguyen et al., 2013]. It is elicited when the user looks at a visual stimulus which flickers at a specific frequency. The stimulation frequency and its harmonics are observed in an oscillation mainly measured over the user's occipital brain area [Amiri et al., 2013]. A common BCI protocol thus consists of presenting the user with several targets flashing at a different frequency [Amiri et al., 2013]. Each target is associated with one action over the BCI effector, e.g. a specific wheelchair movement, or a letter of interest when the BCI is used for communication purposes. A target-specific SSVEP is elicited when the user directs his gaze towards the target of interest. A different protocol is utilized to elicit the P300 wave, which is typically exploited for communication BCIs (e.g., [Sellers et al., 2006] [Krusienski et al., 2006] [Kindermans et al., 2012] but has also permitted users to control wheelchairs [Rebsamen et al., 2006] and prostheses [Palankar et al., 2008]. This ERP systematically occurs 300ms after presentation of a rare but expected visual stimulus [Waldert et al., 2009]. To exploit the P300 wave, a grid filled in with letters is displayed on a screen. Users are asked to focus on the letter they wish to spell, and the grid columns and rows are successively highlighted [Amiri et al., 2013]. The P300 wave is elicited when a

column or row containing the letter of interest is highlighted [Amiri et al., 2013].

By contrast, internally-paced BCIs rely on brain patterns voluntarily elicited by users, e.g. Slow-Cortical Potentials (SCP) and Sensorimotor Rhythms (SMR) [Waldert et al., 2009]. SCPs are slow potentials centred on the upper surface of the brain, while SMRs are rhythms generated in the motor cortex, i.e. in brain areas particularly devoted to motor control. Voluntarily modulation of SCPs and SMRs is possible after training [Waldert et al., 2009], thus enabling users to control, for example, prosthetic devices [Pfurtscheller et al., 2000] [Wang et al., 2013c], cursors [Wolpaw and McFarland, 2004] or robots [Chae et al., 2012].

Brain pattern elicitation sometimes requires the execution of movements of unimpaired body parts [Amiri et al., 2013], e.g. head or eye movements [Mak and Wolpaw, 2009]. The corresponding BCIs are referred to as "dependent", in contrast with independent BCIs which don't require the ability to contract particular muscles [Wolpaw et al., 2002] [Mak and Wolpaw, 2009]. For example, the elicitation of visual evoked potential may require (possibly imperceptible) eye movements of the user towards the visual stimuli [Mak and Wolpaw, 2009]. The applicability of such dependent BCIs to severely motor-impaired patients with poor control over muscles is limited [Wolpaw et al., 2002].

1.1.3 BCI components

Several components are necessary to translate the user's cerebral activity into effector actions. The cerebral signal acquisition system, the transducer, the controlled effector and the feedback provided to the user are the main components of a BCI system [Schwartz et al., 2006] (**Figure 1.1**).

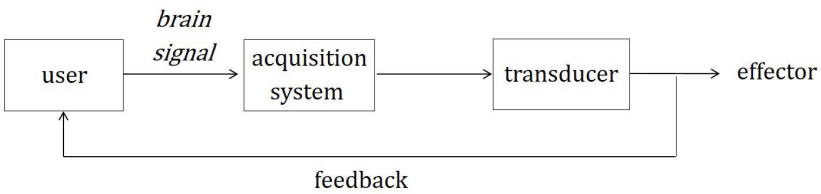

Figure 1.1: BCI main components: acquisition system, transducer, effector and feedback.

Signal acquisition The acquisition system is used to sample, amplify and digitize a measure of the user's cerebral activity [Homer and Nurmikko, 2013] as reflected by electrophysiological, magnetic or metabolic signals [Mak and Wolpaw, 2009].

Electrophysiological signals originate from the electrical currents generated by neurons. They are acquired by means of microelectrodes arrays (MEA), electrocor-

ticographic (ECoG) or electroencephalographic (EEG) arrays [Lebedev and Nicolelis, 2006]. These acquisition systems measure electrical fields at different distances from their sources, and therefore exhibit different degrees of invasiveness and spatial resolutions. MEAs are invasive arrays which directly sample neurons' electrical activity from within the brain (intracortical recordings) [Homer and Nurmikko, 2013]. Electrocorticographic (ECoG) arrays acquire the cerebral activity at the surface of the brain [Mak and Wolpaw, 2009]. In contrast with MEAs, ECoG arrays are said to be semi-invasive [Rak et al., 2012]. Finally, electroencephalographic (EEG) arrays measure the neural signals from the surface of the skull [Teplan, 2002]. EEG-based acquisition systems are therefore non-invasive [Rak et al., 2012].

Magnetoencephalographic (MEG) sensors sample the magnetic field generated by the brain activity at a few centimeters above the skull [Buzsáki et al., 2012].

The cerebral activity is also reflected by the blood oxygenation level in the brain, which is typically measured via fMRI (functional Magnetic Resonance Imaging) or fNIRS (functional Near-Infrared Spectroscopy) [Mak and Wolpaw, 2009].

Current MEG and fMRI acquisition systems suffer from high bulkiness, high cost and/or poor temporal resolution [Mak and Wolpaw, 2009] [Nicolas-Alonso and Gomez-Gil, 2012]. fNIRS acquisition systems, on the other hand, hold promise for BCI applications because of their non-invasiveness, portability and reasonable price [Mak and Wolpaw, 2009]. The feasibility of fNIRS decoding has been demonstrated in recent studies [Coyle et al., 2007] [Naseer et al., 2014] [Hong et al., 2015]. Offline distinction between 3 intentions was achieved in [Hong et al., 2015]; binary cursor control was reported in [Coyle et al., 2007] and [Naseer et al., 2014]. As 3D EEG-based neural control was provided to users in several studies [LaFleur et al., 2013], [McFarland et al., 2010], the performance of fNIRS-driven BCIs is not yet comparable with the performance of EEG-based BCIs. Consequently, BCIs mostly rely on electrophysiological sources of control [Mak and Wolpaw, 2009].

Transducer The BCI transducer translates brain activity measurements into estimates of the user's intention. It is generally composed of several signal processing blocks. A first, optional step consists of enhancing the raw cerebral signals. Algorithmical preprocessing methods permit to improve the Signal-to-Noise Ratio, and/or to remove (at least partially) artifacts [Bashashati et al., 2007a]. Features specifically related to the user's intentions are then extracted from the cerebral signals [Mak and Wolpaw, 2009]. A decoder, also referred to as "translation algorithm" [Yuan and He, 2014] or "feature translator" [Bashashati et al., 2007a], interprets the brain features and issues an estimate of the user's intention. Construction of the decoder is required before utilization of the BCI system. In other words, a model of dependence between neuronal patterns and associated intended actions has to be established [Homer and Nurmikko, 2013]. The hypotheses used to build a decoder depend on the problem at hand (signal quality and resolution, effector etc.). Most BCI systems rely on user-specific decoders, i.e. on decoders whose parameters are adapted for each user [Wolpaw et al., 2002]. This is typically done

by applying model identification algorithms on a data set that contains samples of cerebral signals along with the corresponding intentions of the subject. After being optionally enhanced by post-processing methods [Bashashati et al., 2007a], intention estimates are conveyed to the effector's controller.

Effector control The effector's controller converts intention estimates into commands which are sent to the effector. The type of effector integrated into a BCI system depends on the objective of the BCI. BCIs mostly aim at providing patients with the ability to communicate, exert control over their environment, displace themselves, or recover some motor control over their limbs [Mak and Wolpaw, 2009]. Beside the traditional spellers, BCIs for communication often give users control over a cursor displayed on a computer screen, so that they can write texts, select icons, surf the Internet etc. BCIs for environmental control also rely on cursor control for the adjustment of environmental variables [Mak and Wolpaw, 2009], e.g. light or position of a motorized bed [Cincotti et al., 2008]. Navigation (locomotion) BCIs provide users with control over wheelchairs [Mak and Wolpaw, 2009]. Finally, motor BCIs aim at restoring limb mobility. They generally embed upper- or lower-limb orthoses [Yuan and He, 2014] or prostheses. Considerable efforts are put into the restoration of upper-limb mobility, in particular for the execution of reaching movement with a brain-controlled robotic arm (e.g., [Wodlinger et al., 2015] [Aflalo et al., 2015]).

Feedback Natural volitional motor control is permitted by the perception and exploitation of different types of feedback [Suminski et al., 2010], e.g. proprioceptive, visual, auditory or tactile feedback. BCI systems consequently supply users with so-called closed-loop control over the effector, i.e. feedback is regularly delivered to users. Most BCI systems exclusively provide users with visual feeback on the effector state [Hochberg et al., 2012] [Collinger et al., 2013] [Wodlinger et al., 2015]. The potential benefits of alternative feedback types have nevertheless drawn the attention of several teams [Cincotti et al., 2007] [Wilson et al., 2012] [McCreadie et al., 2014] [Perruchoud et al., 2016]. Auditory feedback was provided to users in [McCreadie et al., 2014]; sensory feedback was delivered through vibrotactile or electrotactile stimulation on users' skin or tongue in [Cincotti et al., 2007] and [Wilson et al., 2012]. The utilization of more complex schemes like direct intracortical brain stimulation [O'Doherty et al., 2011] has additionally been reported [Lebedev and Nicolelis, 2006] [Flesher et al., 2016].

1.2 Motor BCIs

The present dissertation reflects work conducted within the framework of motor BCIs, i.e. BCIs providing users with control over an orthosis or prosthesis [Mak and Wolpaw, 2009]. Such BCIs raise hopes that limb mobility may be restored in severely impaired patients, for example patients with tetraplegia. Specific technical challenges

are associated with the control of physical orthoses or prostheses. The present section gathers a presentation of these challenges and of the technical solutions which have been proposed to address them.

1.2.1 Towards clinical motor BCIs

The following set of technical features is specially considered when designing a motor BCI.

Safety The safety of the signal acquisition system is critical for BCI applications [Leuthardt et al., 2006b]. Moreover, in contrast with communication BCIs and their virtual effectors, motor BCIs generally involve interactions between the user and a physical effector. Specific risks result from this physical interaction, for examples falls or the execution of anatomically impossible movements. As they jeopardize the safety of the BCI users, it is crucial to avert them.

Chronicity Long-term validity is commonly aimed at during the development of a BCI system [Lebedev and Nicolelis, 2006] [Carmena, 2013] [Leuthardt et al., 2006b]. It implies the chronic acquisition of brain signals (i.e., over several decades). Unavoidable signal instabilities, for example due to the user's fatigue, are one of the challenges associated with chronic BCIs [Wolpaw et al., 2002].

Decoding accuracy BCI systems benefit patients if the decoded actions reflect the user's intentions with an accuracy enabling him to interact with his environment [Lebedev and Nicolelis, 2006] [Leuthardt et al., 2006b] [Marathe and Taylor, 2011] [Marathe and Taylor, 2015] (see examples in **Figure 1.2**).

Multi-limb control Full independence of severely motor-impaired patients, e.g. patients with tetraplegia, is possible if they are provided with control over both lower- and upper-limb protheses or orthoses.

Degrees of Freedom Upper- and lower-limb prostheses have multiple controllable Degrees of Freedom (DoF). 3D control of the endpoint of an upper-limb prothesis or orthosis is a frequent objective of motor BCIs [Lebedev and Nicolelis, 2006].

Robustness Most BCI studies are currently completed in controlled laboratory environments. Robustness of the transducer, i.e. stability of decoding accuracy in noisy environments, is necessary for the practical BCIs [Leuthardt et al., 2006b].

Feedback It is crucial for BCI users to be regularly provided with feedback on the effector state. Specific feedback characteristics have been shown to have a dramatic impact on neural control performances. It has for example been demonstrated that using several feedback modalities, e.g. visual and proprioceptive feedback, can improve the quality of neural control [Suminski et al., 2010]. The ability to

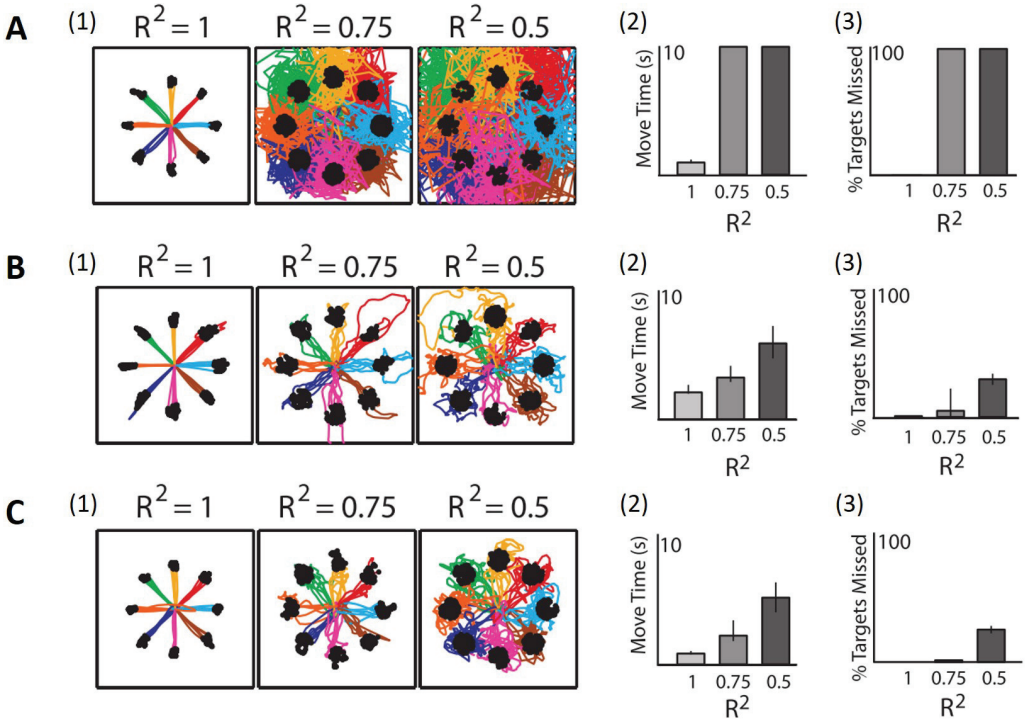

Figure 1.2: Ability to execute reaching movements depending on the correlation R between intended and decoded trajectory parameters (reproduced from [Marathe and Taylor, 2011]). Decoded position or velocity parameters were used to control either the position or the velocity of an effector. It was necessary for subjects to learn a position-to-velocity map when decoded position parameters were used to control the velocity of the effector. **(1)**: Examples of reaching trajectories. **(2)**: Average duration of a reaching movement. **(3)**: Percentage of missed targets. **A.** Position decoding, position control. **B.** Position decoding, velocity control. **C.** Velocity decoding, velocity control.

exert control over a motor BCI has also been found to be affected by unnatural, important latencies between the user's intention and the corresponding effector reaction [Willett et al., 2013][Marathe and Taylor, 2015]. The duration required for the execution of reaching tasks has been found to be increased by around 180ms for every 100ms-long pure delay introduced into the transducer [Willett et al., 2013]. Similar results were reported in [Marathe and Taylor, 2015], where the introduction of a 300ms-long delay into the transducer increased the duration required to execute reaching movements by more than 300ms. An important latency also limits brain plasticity, i.e. the ability of the user to adapt to the BCI transducer [Leuthardt et al., 2006b]. BCI developers therefore endeavour to limit the latency between movement intention and execution of this movement by the effector.

Real-time decoding update The effector controller is regularly fed with updated estimates of the user's intention. The suitability of update frequencies depends on the effector [Leuthardt et al., 2006b]. In contrast with communication BCIs which can be highly beneficial to patients even when they exhibit comparatively low update rates, motor BCIs require a high update rate to profit users. Most BCI systems rely on a frequency rate of 10 to 20Hz [Kim et al., 2006a] for cursor or prosthetic control in interaction-free environments. Lower frequency updates, e.g. 5Hz for arm prosthesis control [Yanagisawa et al., 2012a], were nevertheless reported. It has been suggested that an update rate of 20Hz is necessary for object manipulation [Kim et al., 2006a]. The update frequency defines the maximal duration the transducer can use to translate brain signals into user intention's estimates. The computational complexity of the transducer's processing blocks is consequently constrained by the update frequency and the hardware processing power.

Asynchronicity Most BCI demonstrations are conducted using synchronous, cue-paced control paradigms. User intentions are not processed outside predefined, cued windows [Graimann et al., 2009]. By contrast, an asynchronous or self-paced BCI decoder is continuously available to the user [Graimann et al., 2009]. Potential BCI users express their desire for stand-alone BCI systems, i.e. systems which utilization does not require the presence of a technician [Blabe et al., 2015]. Asynchronicity is thus a technical feature essential for practical motor BCIs [Wolpaw et al., 2002].

Mental load It is desirable for neural control to be associated with a light mental load [Yuan and He, 2014] [Mak and Wolpaw, 2009].

Embeddability Bulky hardware components for signal acquisition, processing or effector control are not practical for motor BCIs. Embeddable components are therefore developed for motor BCI applications. Because of this constraint, cumbersome acquisition systems like MEG are currently deemed unfit for motor BCIs [McFarland and Wolpaw, 2011].

These challenges drive the technical choices made when designing the components of motor BCIs.

1.2.2 Signal acquisition

The signal acquisition system is a crucial component of a BCI system. Its invasiveness determines the chronicity and safety of the system. Additionally, the control of several limbs with multiple degrees of freedom requires a highly informative content to be extracted from the acquired signals. The spatial and spectral characteristics of the acquired signals depend on the acquisition system, and directly impact the feasibility of accurate multi-limb, multi-DoF (Degrees of Freedom) effector control. To date, motor BCIs mostly rely on electrophysiological sources of control [Mak and Wolpaw, 2009], as opposed to magnetic and metabolic sources.

1.2.2.1 Cortical patterns for prosthetic control

Electrophysiologial signal acquisition systems mostly give access to the cerebral activity generated in the cortex [Buzsáki et al., 2012], i.e. in the $2-4$mm thick surface of the brain (**Figure 1.3.A**) [Buzsáki et al., 2012].

Cortical electro-physiological activity Cortical electrophysiological activities are generated by nerve cells called "neurons". Typical neurons are composed of a body and of two types of extensions, namely one axon and a variable number of dendrites [Squire et al., 2013] (**Figure 1.3.C**). Neurons are interconnected via their axon and dendrites, i.e. they are organized in networks. They are specialized in the reception, processing, and transmission of information encoded in electrical signals [Kandel et al., 2000]. Neurons have the capacity to generate a transient impulse called Action Potential (AP) in response to a simulus [Kandel et al., 2000]. The AP propagates inside the neuron and along the axon [Kandel et al., 2000]. At the axon's extremity, it is converted into chemical transmitters released in the extracellular environment, where they are captured by a cell, e.g. by the dendrites of a neighbouring neuron. The generation of APs is an all-or-nothing phenomenon: APs' amplitude and shape don't depend on the stimulus, and APs only occur if the stimulus is superior to a specific threshold [Kandel et al., 2000]. In other words, neurons units of communication are Boolean variables [Kandel et al., 2000]. Neurons are said to "spike" [Homer and Nurmikko, 2013] or to "fire". Information is thus exclusively encoded in the circulation of Boolean variables in neural networks [Kandel et al., 2000].

The cortical neural network is composed of 6 layers of aligned and interconnected neurons, referred to as "pyramidal neurons" (**Figure 1.3.B**). The pyramidal neurons also communicate with subcortical areas via their axon [Kandel et al., 2000]. Due to the alignment of the pyramidal neurons and to spiking synchronization phenomena, the joint activity of cortical neurons results in modulations (rhythms) observed in specific frequency bands. The following division of brain signals into bandwidths is

often used to describe neural population activity [Kandel et al., 2000] [Morshed and Khan, 2014]:

- Delta: $0.5\text{Hz} < f_\delta < 4\text{Hz}$
- Theta: $4\text{Hz} < f_\theta < 7\text{Hz}$
- Mu: $8\text{Hz} < f_\mu < 13\text{Hz}$ (over the motor cortex)
- Beta: $13\text{Hz} < f_\beta < 30\text{Hz}$
- Gamma: $30\text{Hz} < f_\gamma$

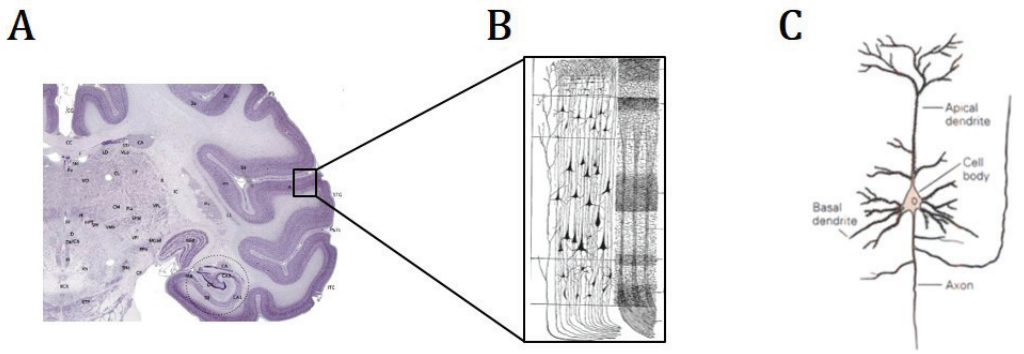

Figure 1.3: Cerebral cortex. **A.** Cerebral cortex (represented in dark purple) [BrainMaps, 2017]. **B.** Pyramidal neurons [Gray, 1918]. **C.** Pyramidal neuron [Kandel et al., 2000].

Functional organization Most motor BCIs exploit neural activities generated in localized areas of the cortex, from a few mm^2 [Collinger et al., 2013] [Wodlinger et al., 2015] [Kellis et al., 2012] to few cm^2 [Wolpaw and McFarland, 2004] [Baxter et al., 2013] [Wang et al., 2013c]. The brain presents a functional organization, i.e. spatial areas are specialized in specific functions. This functional organization permits to select areas relevant for neural control in motor BCIs.

The anatomical division of the brain into lobes (**Figure 1.4.A**) is often used to facilitate the description of its functional organization [Kandel et al., 2000]. Auditory stimuli are mainly managed in the temporal lobe [Kandel et al., 2000]. The occipital lobe receives and processes visual information sent by the eyes [Kandel et al., 2000]. The parietal lobe contains areas involved in space perception and processing of sensori information [Kandel et al., 2000]. The frontal lobe is in particular in charge of cognitive functions (thought, learning, memory, etc.) and motor skills (including speech) [Kandel et al., 2000].

The reported motor BCIs have exploited sources of control related to different cortex areas. Neural activity in the temporal lobe enabled users to control a cursor

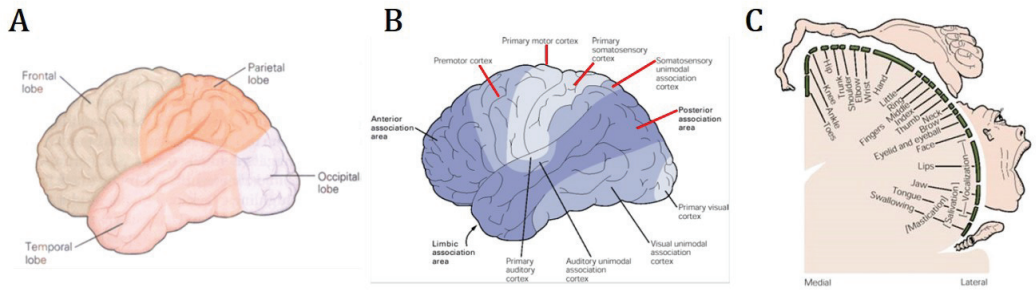

Figure 1.4: **A.** Brain's lobes [Kandel et al., 2000]). **B.** Cortical areas involved in motor control (red lines) [Kandel et al., 2000]. **C.** Somatotopic organization of the primary motor cortex (reproduced from [Kandel et al., 2000]).

using their speech network in [Leuthardt et al., 2011]. Users were provided with neural control over a robotic arm by means of visual evoked potentials measured over the occipital lobe in [Ortner et al., 2011]. A robotic device and a cursor were controlled by subjects' occipital activity in [Ferreira et al., 2008] and [Trejo et al., 2006], respectively. Finally, the majority of reported motor BCIs made use of activity elicited in the frontal and parietal lobes, and more specifically in the motor cortex [Collinger et al., 2013] [Wodlinger et al., 2015] [Kellis et al., 2012] [Wolpaw and McFarland, 2004] [Baxter et al., 2013] [Wang et al., 2013c].

The motor cortex is devoted to the organization of movements, and in particular of voluntary movements [Kandel et al., 2000]. It is composed of the primary motor area (M1) and of premotor areas, in particular the premotor cortex (PM), the supplementary motor area (SMA) and the Posterior Parietal Cortex (PPC) [Kandel et al., 2000] (**Figure 1.4.B**). Present knowledge about neuronal management of motor control remains limited [Lebedev and Nicolelis, 2006]. Some features of neural control have nevertheless been characterized. M1 neurons encode both low-level (e.g., muscle force) and high-level information about desired movements (e.g., direction of arm movements [Kandel et al., 2000]). The amount of neurons devoted to one limb or organ is not proportional to the organ size, but to the complexity of its control (somatotopic organization, illustrated in **Figure 1.4.C**). This spatial organization is exploited by numerous BCIs. Additionally, motor control is mainly contralateral, i.e. cortical neurons from left hemisphere are connected to motor neurons managing the right size of the body [Kandel et al., 2000]. Premotor areas are involved, e.g., in motor planning and limb coordination [Kandel et al., 2000]. Although the functions managed by the SMA have not yet been clearly identified, it has been suggested that it is involved in the control of voluntary movements and of sequences of movements [Kandel et al., 2000]. The Posterior Parietal Cortex manages in particular sensory information for motor planning [Kandel et al., 2000]. Whereas most BCIs indistinguishably exploit activity generated in various areas of the motor cortex, the specific use of M1 activity has been reported in motor BCIs

[Velliste et al., 2008] [Bouton et al., 2016] [Hochberg et al., 2012]. The feasibility of PPC-based neural control was additionally investigated in a recent study [Aflalo et al., 2015], and bimanual movements were controlled using M1, SMA, PPC and primary somatosensory activity in [Ifft et al., 2013].

Electro-physiological sources of control for motor BCIs Most reported motor BCIs were based on the elicitation of internally-paced patterns [Collinger et al., 2013] [Wodlinger et al., 2015] [Kellis et al., 2012] [Pfurtscheller et al., 2000] [Wolpaw and McFarland, 2004] [Baxter et al., 2013] [Wang et al., 2013c]. The use of externally-paced patterns (Evoked and Event-Related Potentials) has nevertheless been reported in several motor studies [Palankar et al., 2008] [Bell et al., 2008] [Ortner et al., 2011].

Cortical neurons The voluntary modulation of the firing rate of neurons in the motor cortex has been exploited in most recent motor intracortical BCI systems [Velliste et al., 2008] [Hochberg et al., 2012] [Collinger et al., 2013] [Wodlinger et al., 2015] [Bouton et al., 2016]. Neural control was generally based on the spiking activity measured in the primary motor cortex [Velliste et al., 2008] [Bouton et al., 2016] associated to hand [Hochberg et al., 2012] or arm movements. The bimanual movements reported in [Ifft et al., 2013] relied on a larger motor cortex area, namely on M1, SMA, PPC and primary somatosensory activity. A preliminary study has also demonstrated the possibility to decode users' intentions from PPC activity [Aflalo et al., 2015].

Sensorimotor rhythms Most semi- and non-invasive motor BCIs are based on the exploitation of sensorimotor rhythms [Pfurtscheller et al., 2000] [Wolpaw and McFarland, 2004] [McFarland et al., 2010] [LaFleur et al., 2013] [Baxter et al., 2013] [Milekovic et al., 2012] [Kapeller et al., 2015] [Hotson et al., 2016]. Sensorimotor rhythms (SMR) are patterns generated by neural populations of the sensorimotor cortex [Leuthardt et al., 2006b]. Although SMRs are not well understood [Yuan and He, 2014], some properties of these rhythms have been characterized.

Modulations synchronized with motor tasks have been observed in μ, β and γ frequency bands (**Figure 1.5**). Two seconds before the onset of a movement, an amplitude decrease is observed in μ and β bands (Event-Related Desynchronization). This phenomenon is mostly contralateral to the movement, and is spatially consistent with the somatotopic organization of the motor cortex. Simultaneously, an amplitude increase is observed in the γ band. This rhythm is more spatially focused than μ and β modulation [Schalk and Leuthardt, 2011]. After the movement, an amplitude increase occurs in μ and β frequency bands (Event-Related Synchronization). These rhythms can be observed in the contralateral sensorimotor area, with a peak 600ms after the movement offset. Although they have been initially observed during planning and execution of real movements, it has been shown that they also occur during motor imagery but with a smaller amplitude. Volitional control of SMRs

is thus possible after training, and the simplest way to elicit it is to use motor imagination [Leuthardt et al., 2006b] [Waldert et al., 2009].

Similarly, Movement-Related Potentials (MRPs) are synchronized with movements. These low-frequency potentials appear 1 to 1.5s before a movement [Waldert et al., 2009] [Bashashati et al., 2007a]. Their distribution is bilateral at the beginning of the movement, and becomes contralateral when close to the movement [Bashashati et al., 2007a]. Readiness Potential and Lateralized Readiness Potential are examples of MRPs. Volitional control of MRPs can be achieved, e.g. via motor imagination [Nazarpour et al., 2009].

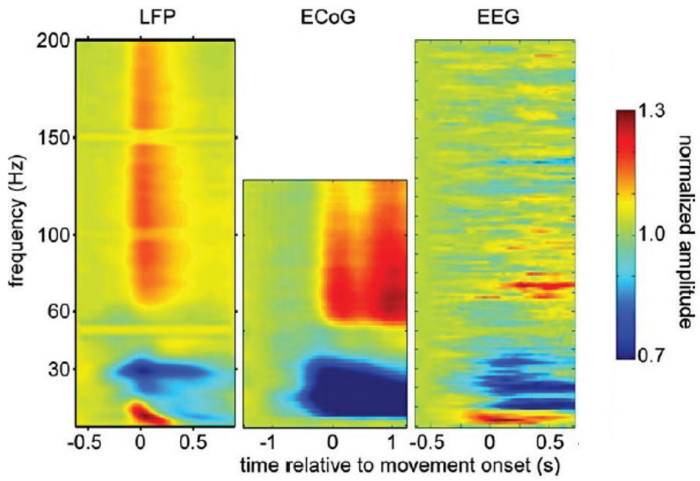

Figure 1.5: SMRs (reproduced and modified from [Waldert et al., 2009])

SMRs are widely exploited in motor BCIs, e.g. for prosthesis [Pfurtscheller et al., 2000] [Yanagisawa et al., 2012a] [Baxter et al., 2013] [Wang et al., 2013c] or virtual effector control [Wolpaw and McFarland, 2004] [McFarland et al., 2010] [Doud et al., 2011] [Kellis et al., 2012].

SCP Slow Cortical Potentials (SCPs) are slow potentials centered on the upper surface of the brain [Leuthardt et al., 2006b] [Bashashati et al., 2007a]. SCP are not necessarily related to motor control or to a mental specific task, but volitional modulation of SCPs is possible after training [Leuthardt et al., 2006b] [Waldert et al., 2009]. SCP-based neural control has been used for spelling systems [Birbaumer et al., 1999] and can be used for motor BCIs.

Response to mental tasks Brain pattern elicitation via mental tasks has been performed in a few cursor-control studies [Penny et al., 2000] [Vansteensel et al., 2010]. Discrimination between responses to mental tasks has also been investigated in feasibility studies, e.g. in [Anderson and Sijercic, 1996] [Ramsey et al., 2006], or in [Chiappa and Bengio, 2003] where both mental and motor tasks were classified

during offline analyses. Reported mental tasks consisted in solving equations [Ramsey et al., 2006], multiplying [Anderson and Sijercic, 1996] or substracting [Penny et al., 2000] [Vansteensel et al., 2010] numbers, writing a letter [Anderson and Sijercic, 1996], finding words beginning with a specific letter [Chiappa and Bengio, 2003], etc. The corresponding brain patterns were for example observed in the prefrontal cortex [Ramsey et al., 2006] [Vansteensel et al., 2010], or in the parietal, frontal and occipital cortex [Anderson and Sijercic, 1996].

Event-related potentials A few teams have reported P300-based neural control over robotic effectors, namely a robotic arm [Palankar et al., 2008] and a humanoid robot [Bell et al., 2008].

Evoked Potentials While evoked potentials have rarely been exploited for motor BCIs, SSVEP-based hand orthosis control has for example been achieved in [Ortner et al., 2011].

Other strategies Several motor BCIs have made use of several sources of control [Horki et al., 2011] [Pfurtscheller et al., 2010], e.g. hybrid SSVEP- and SMR-based hand prosthesis control in [Pfurtscheller et al., 2010] and [Horki et al., 2011].

Measuring cortical patterns The signals measured by electrophysiological acquisition systems, namely Microelectrode Arrays (MEA), Electrocorticographic (ECoG) and Electroencephalographic (EEG) acquisition systems, mainly correspond to the extracellular currents generated by the cortical neurons [Buzsáki et al., 2012]. Depending on the invasiveness of the acquisition system, sensors are located at a distance which ranges from a few μm (MEA) to several cm (EEG) from the cortical neurons generating the extracellular currents of interest [Waldert et al., 2009] (**Figure 1.6**). This distance impacts the size of the neuronal population observed by sensors. When signals are measured in the vicinity of neurons, they mostly reflect individual neural activity. By contrast, remote electrodes acquire the electric field generated by a large neuronal population, i.e. they measure the superposition of extracellular currents. The individual (spatial, spectral) characteristics of neuronal patterns are lost because of this spatial averaging. As the current amplitude depends on the inverse of the sensor-neuron distance [Buzsáki et al., 2012], this distance additionally impacts the Signal-to-Noise Ratio. High frequencies are particularly hard to observe since current amplitudes are also approximately inversely proportional to their frequency [Buzsáki et al., 2012]. They are, however, generally more spatially focused than lower frequencies because they are mostly generated by small neuron populations [Leuthardt et al., 2006a].

Acquisition systems thus correspond to different combinations of spectral and spatial resolution and invasiveness. While resolution impacts the feasibility of accurate, multiple DoF multi-limb control, the invasiveness of the acquisition system

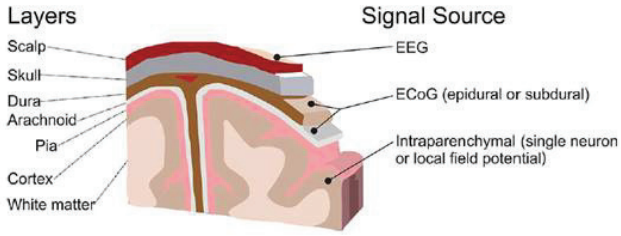

Figure 1.6: Location of MEA, ECoG and EEG sensors (reproduced from [Schalk and Leuthardt, 2011])

determines the chronicity and safety of the system. The resolution-invasiveness trade-offs associated with MEA-, ECoG and EEG- acquisition systems are detailed in the next sections.

1.2.2.2 Microelectrode arrays (MEA)

MEA-based motor BCIs have been deployed both in humans [Hochberg et al., 2006] [Kim et al., 2008] [Simeral et al., 2011] [Kim et al., 2011] [Hochberg et al., 2012] [Jarosiewicz et al., 2013] [Collinger et al., 2013] [Wodlinger et al., 2015] [Bouton et al., 2016] and primate subjects [Taylor et al., 2002] [Carmena et al., 2003] [Velliste et al., 2008] [Orsborn et al., 2011] [Orsborn et al., 2012] [Ifft et al., 2013] [Sadtler et al., 2014]. MEA, also referred to as "Intracortical Electrode Array", e.g. [Maynard et al., 1997] [Wodlinger et al., 2015], are grids of needle microelectrodes which are implanted into the cortex [Homer and Nurmikko, 2013].

Most reported MEA-based motor BCIs relied on the Utah array [Blackrock, 2016] (**Figure 1.7**) to acquire neural signals in human subjects [Simeral et al., 2011] [Kim et al., 2011] [Hochberg et al., 2012] [Jarosiewicz et al., 2013] [Collinger et al., 2013] [Wodlinger et al., 2015]. This commercialized array gathers 10×10 1.5mm-long electrodes on a $4 \times 4\text{mm}^2$ surface. The inter-electrode distance is 0.4mm. The size of the electrode tips is approximately 4 μm [Simeral et al., 2011], while the diameter of a neuron body is 50μm or more [Kandel et al., 2000]. Signal acquisition has nevertheless been performed with other MEAs in a few motor BCIs, e.g. [Orsborn et al., 2012] [Ifft et al., 2013]. The MEAs used in [Orsborn et al., 2012] and [Gowda et al., 2014] for example gathered 128 electrodes, with a 35μm diameter and a 0.5mm inter-electrode distance. Another team has reported bimodal control over virtual upper-limb prostheses by monkeys [Ifft et al., 2013] using volumetric arrays [Schwarz et al., 2014], i.e. arrays embedding electrodes of different lengths.

The spectral content of MEA signals stretches as far as 40kHz [Buzsáki et al., 2012]. Preprocessing permits to extract three signals from the raw signal acquired by an intracortical electrode, namely Single-Unit Activity (SUA), Multi-Unit Activity (MUA) and Local Field Potentials (LFP) [Waldert et al., 2009]. MUA and SUA signals reflect the spiking activity of the few neurons located in the immediate

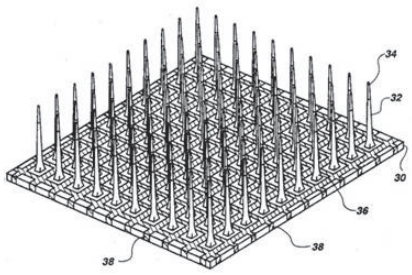

Figure 1.7: Utah microelectrode array (source: US Patent # 5,215,088)

vicinity of the electrode's tip (1 or 2 neurons typically) [Homer and Nurmikko, 2013] [Leuthardt et al., 2006a]. LFPs measure the superposed activity of a small population of neurons located in the neighborhood of the electrode's tip (spatial resolution around 1mm) [Homer and Nurmikko, 2013] [Leuthardt et al., 2006a]. Signals acquired by MEAs thus exhibit a high spatial resolution along with a large spectral content, but correspond to the neural activity generated within a restricted cortex area [Mak and Wolpaw, 2009]. Even in the case of multiple arrays implantation, the ability to exploit patterns associated with several brain areas remains limited [Mak and Wolpaw, 2009].

Because of the highly informative content of the signals measured by MEAs, MEA-based BCIs hold promise to support complex control strategies (in particular, accurate multiple DoF and multi-limb control). Tetraplegic patients controlled a robotic arm in [Hochberg et al., 2012]. 7D and 10D fine control over an upper-limb prosthesis was demonstrated for MEA-based motor BCIs [Collinger et al., 2013] [Wodlinger et al., 2015]. Several teams have additionally reported 2D cursor control by tetraplegic patients [Hochberg et al., 2006] [Kim et al., 2008] [Kim et al., 2011] [Simeral et al., 2011] [Jarosiewicz et al., 2013]. Finally, cursor and upper-limb prosthesis control has been achieved by Non-Human Primates during several BCI studies [Taylor et al., 2002] [Carmena et al., 2003] [Velliste et al., 2008] [Orsborn et al., 2011] [Orsborn et al., 2012] [Ifft et al., 2013] [Sadtler et al., 2014].

In spite of these promising studies, the invasiveness of intracortical arrays is for the moment a significant hindrance to their utilization for motor clinical BCI applications. The MEA implantation surgery causes a mechanical trauma in the brain [Polikov et al., 2005]. Despite optimizations of the implantation procedure (for example, of the implantation speed) and of the electrode shape, cell deaths or vessel ruptures are observed in the cortex [Polikov et al., 2005]. The MEA invasiveness also induces immune reactions which can be problematic for MEA-based chronic BCIs. Electrode encapsulation, i.e. the formation of a glial scar tissue around electrodes (glial cells being supportive cells surrounding the neurons [Squire et al., 2013]), occurs a few weeks after implantation [Polikov et al., 2005]. It causes an increase of impedance around electrodes along with an augmentation of the electrode tip-neuron distance [Polikov et al., 2005]. The subsequent decay of signal amplitude jeopardizes

signal acquisition after several months [Yuan and He, 2014], and is thus thought to limit the life expectancy of the MEA-driven BCIs [Leuthardt et al., 2006a]. In [Simeral et al., 2011] for example, neurons with a firing rate superior to 1Hz were associated with only 41 out of the array's 96 electrodes 1000 days after implantation. These observations suggest that further improvements may be necessary before MEAs fully support safe and chronic BCIs [Murphy et al., 2015].

1.2.2.3 Electrocorticography (ECoG)

ECoG-based motor BCIs have been investigated for both human [Leuthardt et al., 2004] [Leuthardt et al., 2006a] [Wilson et al., 2006] [Schalk et al., 2008] [Blakely et al., 2009] [Vansteensel et al., 2010] [Milekovic et al., 2012] [Kellis et al., 2012] [Yanagisawa et al., 2012a] [Wang et al., 2013c] [Fifer et al., 2014] [Kapeller et al., 2015] and primate subjects [Ashmore et al., 2012] [Sanchez et al., 2008] [Rouse and Moran, 2009] [Rouse et al., 2013] [Williams et al., 2013] [Marathe and Taylor, 2013] [Rouse et al., 2016]. Also referred to as "intracranial EEG" [Morshed and Khan, 2014], ECoG grids gather flat electrodes and are implanted onto the cortical surface, more precisely under or over the dura mater (respectively, subdural and epidural implantation) [Schalk and Leuthardt, 2011].

To date, most ECoG-based motor BCI clinical studies have been completed with subjects who had not been not fitted with an ECoG array because they participated in a BCI study [Vansteensel et al., 2010] [Milekovic et al., 2012] [Kellis et al., 2012] [Yanagisawa et al., 2012a] [Kapeller et al., 2015]. They were generally enrolled because they were undergoing a short-term ECoG implantation for the localization of an epileptic focus before a resection surgery. They were therefore implanted with clinical ECoG arrays which materials, configuration and size were optimized for the requirements of this monitoring objective [Schalk and Leuthardt, 2011]. The utilized clinical ECoG arrays generally consist in titanium, ball-shaped macro-electrodes placed on a grid (8 electrodes) or a strip (4 or 6 electrodes) [Schalk and Leuthardt, 2011] (**Figure 1.8**). They are mostly implanted under the dura mater (subdural implantation). Electrode diameters are typically of 4mm, for an inter-electrode of 1cm [Leuthardt et al., 2004] [Leuthardt et al., 2006a] [Wilson et al., 2006] [Schalk et al., 2008] [Blakely et al., 2009] [Vansteensel et al., 2010] [Milekovic et al., 2012] [Fifer et al., 2014]. The number of channels ranged from 15 [Yanagisawa et al., 2012a] to 128 electrodes [Hotson et al., 2016], possibly via the joint implantation of several arrays [Fifer et al., 2014]. Similar electrodes have been used for preliminary studies in human subjects [Schalk et al., 2007] [Chin et al., 2007]. A few teams have used high-density arrays, e.g. with a 0.5cm [Wilson et al., 2006] [Kapeller et al., 2015] or 3mm inter-electrode distance in [Hotson et al., 2016] for neural control over a cursor [Wilson et al., 2006], an humanoid robot [Kapeller et al., 2015] or a prosthetic hand [Hotson et al., 2016]. The ECoG arrays used to develop and test motor BCIs on primate models are generally less standardized. Custom arrays with an inter-electrode distance of 3.5mm have been used to measure signals analyzed in preclinical feasibility studies [Chao et al., 2010] [Shimoda et al., 2012].

Finally, the use of micro-ECoG (μECoG) arrays, i.e. arrays with an inter-electrode distance inferior to 4mm and an electrode diameter generally inferior to 100μm [Kellis et al., 2015], has recently been investigated with both human [Kellis et al., 2012] and primate subjects [Williams et al., 2013] [Rouse et al., 2013] [Rouse et al., 2016]. The inter-electrode distance of the tested μECoG arrays ranged from 1 to 3mm, with electrode diameters of 30-40 μm [Kellis et al., 2012] or between 300 μm [Williams et al., 2013] [Rouse et al., 2013] [Rouse et al., 2016] and 600μm [Williams et al., 2013]. The optimal inter-electrode distance is still an active field of research [Wang et al., 2009b] [Slutzky et al., 2010] [Bundy et al., 2014] [Kellis et al., 2015] [Wang et al., 2016]. Several studies have suggested that decoding performance is higher when dense arrays rather than generic arrays with 1cm-large inter-electrode distance are used [Bundy et al., 2014] [Wang et al., 2016].

Figure 1.8: Commercialized ECoG array [Blackrock, 2017].

ECoG spatial resolution is 1.25mm for subdural ECoG and 1.4mm for epidural recordings [Schalk and Leuthardt, 2011]. The synaptic currents generated by pyramidal neurons aligned in the superficial layers of cortex are thought to be the main sources of ECoG signals [Waldert et al., 2009]. ECoG bandwidth ranges from 0 to 500Hz [Schalk and Leuthardt, 2011]. The exploitation of high frequency patterns, which have been found useful in several motor studies [Anderson et al., 2012] [Nurse et al., 2015], is thus possible. Additionally, ECoG signals are significantly less sensitive to artifacts, in particular ocular ones [Ball et al., 2009a], than non-invasive acquisition methods like EEG.

Clinical and preclinical attempts at ECoG-based neural control have yielded encouraging results. Able-bodied epileptic patients have been reported to be able to control 1D or 2D cursors in several studies [Leuthardt et al., 2004] [Leuthardt et al., 2006a] [Wilson et al., 2006] [Schalk et al., 2008] [Blakely et al., 2009] [Vansteensel et al., 2010] [Kellis et al., 2012] [Milekovic et al., 2012]. Epileptic patients controlled a prosthetic arm, a humanoid robot and a prosthetic hand in [Fifer et al., 2014], [Kapeller et al., 2015] and [Hotson et al., 2016] respectively. Motor-impaired and able-bodied patients were able to manipulate a prosthetic hand in [Yanagisawa et al., 2012a], and 3D arm prosthesis control was achieved by a patient with tetraplegia in [Wang et al., 2013c]. 1D [Rouse et al., 2013] [Rouse et al., 2016] and 2D cursor control [Rouse and Moran, 2009] [Ashmore et al., 2012] [Williams et al., 2013] [Marathe and Taylor, 2013] [Rouse et al., 2016] has additionally been achieved by NHPs.

The implantation of an ECoG array requires a craniotomy. As the array is not introduced into the cortex but only positioned on it, the ECoG invasiveness is nevertheless limited when compared to the MEA's one. ECoG monitoring before a resective surgery only requires a short-term implantation, namely a few days or weeks. Most ECoG clinical arrays are therefore FDA-approved for a maximal duration of 28 days. For this reason, ECoG long-term stability has mainly been investigated in preclinical studies, e.g. [Chao et al., 2010] [Degenhart et al., 2016]. Chronic preclinical studies have shown that ECoG signals remain stable over months [Chao et al., 2010] [Shimoda et al., 2012] [Degenhart et al., 2016]. Chronic ECoG acquisition (over 7 years) has additionally been reported in a study completed with epileptic patients [King-Stephens et al., 2015]. These findings suggest that ECoG arrays may be profitably used to develop safe and chronic motor BCI systems.

1.2.2.4 Electroencephalography (EEG)

EEG-based control of prostheses [Pfurtscheller et al., 2000] [Onose et al., 2012] [Baxter et al., 2013], real limbs [King et al., 2015], humanoid robots [Chae et al., 2012] and virtual effectors [Wolpaw and McFarland, 2004] [Trejo et al., 2006] [Yuan et al., 2007] [McFarland et al., 2010] [Royer et al., 2010] [Doud et al., 2011] [LaFleur et al., 2013] has been achieved in the last twenty years. First designed in the 20s, EEG acquisition systems noninvasively measure electrical brain activities from electrodes placed on the scalp [Buzsáki et al., 2012].

EEG acquisition systems are widely used in clinical contexts, e.g. for the localization of epileptogenic zones or for coma diagnostics [Teplan, 2002]. For this reason, EEG systems approved for clinical applications are available on the market. Several EEG-based motor BCIs have made use of commercialized arrays, e.g. Electrocap International's [McFarland et al., 2010], Compumedics Neuroscan's [Chae et al., 2012] [Baxter et al., 2013] or Brain Products' systems [Onose et al., 2012]. EEG sensors are usually embedded in a tissue cap that maintains them on the scalp (**Figure 1.9**). Although motor BCI studies generally exploit montages with 2 [Pfurtscheller et al., 2000] to 64 [McFarland et al., 2010] [LaFleur et al., 2013] recording electrodes, high density systems with 264 embedded electrodes have been utilized in several preliminary BCI studies [Fruitet et al., 2010] [Gwin and Ferris, 2011]. A few teams have additionally reported the use of commercialized entertainment EEG systems (the Emotive array [Emotiv, 2016]), for example providing users with neural control over a robotic arm [Bhattacharyya et al., 2015] or over an upper-limb exoskeleton [Webb et al., 2012].

EEG acquires the superposed activities of a large neuronal population (> 100000 neurons, approximatively located in a 10 cm^2-region) [Waldert et al., 2009] [Buzsáki et al., 2012]. Similarly to ECoG, the main contributors to EEG signals are thought to be the neurons located in the superficial layers of the cortex [Buzsáki et al., 2012]. The activity of deep sources cannot be easily observed in EEG signals because of the sensor-source distance. The EEG spatial resolution, which amounts to a few cm, is limited when compared to the MEA's or ECoG's ones. EEGs' amplitude is

Figure 1.9: Example of a clinical EEG cap [Unicare, 2017]

additionally five times lower than ECoGs' amplitude [Leuthardt et al., 2006a]. EEG spectral content approximately ranges from 0 to 40Hz [Schalk and Leuthardt, 2011]. The absence of high frequencies is due to signal low-pass filtering when crossing the skull and skin [Lebedev and Nicolelis, 2006], and to the reduced power of high frequency components observed from a large distance to their source [Waldert et al., 2009] (see **section 1.2.2.1**). EEGs are very sensitive to artefacts, in particular ocular [Fatourechi et al., 2007a], muscular [Muthukumaraswamy, 2013] [Fatourechi et al., 2007a], cardiac [Lebedev and Nicolelis, 2006] and mechanical ones. Artefacts are liable to perturb neural control [Fatourechi et al., 2007a]. Strategies have been proposed to discard them, but at the risk of losing informative content if the spectral content of the artefacts is included in the frequency band of interest [Fatourechi et al., 2007a].

The feasibility of EEG-driven motor BCIs has been explored by several teams. Binary EEG-based control over a prosthetic and orthesis arm was achieved in [Onose et al., 2012] and [Webb et al., 2012], respectively. The aperture of a robotic arm was controlled by a tetraplegic patient in [Pfurtscheller et al., 2000]; 2D control of a prosthetic and robotic arm was reported in [Baxter et al., 2013] and [Hortal et al., 2015], respectively. Finally, 1D, 2D and 3D control over virtual effectors has been accomplished by several able-bodied subjects [Yuan et al., 2007] [Trejo et al., 2006] [McFarland et al., 2010] [Royer et al., 2010] [Doud et al., 2011] [LaFleur et al., 2013].

While these results are promising, the limited spatial resolution [Leuthardt et al., 2006a] and high-frequency content of EEG signals [Waldert et al., 2009] may complicate multi-DoF and multi-limb effector control. In particular, the adaptation of users to EEG-based decoders often requires extensive training [Leuthardt et al., 2006a], and, in the case of "BCI illiterates", may not permit users to satisfyingly control the effector [Milan and Carmena, 2010]. EEG-based BCI studies nevertheless constitute a significant part of motor BCI studies [Leuthardt et al., 2006a] because of their non-invasiveness [Teplan, 2002], of their attractive cost and of the relatively simple procedure which is required to put EEG caps into place [Milan and Carmena, 2010].

1.2.3 Transducers for motor BCIs

After acquisition, neural signals are processed by the transducer, which translates them into intention estimates in real time. The user is regularly provided with feedback on the transducer output ("closed-loop" experiments) [Héliot et al., 2010]. BCI transducers impact the capability of motor BCI systems to meet several of the technical challenges presented in **section 1.2.1**. Transducers must be designed so as to support real-time updates and to ensure a low feedback latency. These two requirements are generally achieved by using efficient algorithmic procedures and exploiting causal neural patterns. Importantly, the transducer also impacts the system decoding accuracy, which is determined by the transducer's ability to extract reliable information from the neural signals and to correctly interpret it. Transducer design permits to reach a high consistency, or accuracy, between true and estimated intentions. The mental load associated with multi-limb, multi-DoF control additionally depends on the decoding strategy the decoder relies on.

1.2.3.1 Transducer design

Before using a BCI transducer during closed-loop experiments, it is necessary to design it, i.e. to choose its signal components (namely, the pre-processing, feature extraction, decoder and post-processing blocks shown in **Figure 1.10**). Transducer design generally relies on two processes to reach a high consistency (accuracy) between user intention and transducer's output: transducer identification and/or user training [McFarland and Wolpaw, 2011].

The pre-processing, feature extraction and post-processing algorithms utilized in BCI systems are either user-specific [Marathe and Taylor, 2013] [Kapeller et al., 2015] or non user-specific [Yanagisawa et al., 2012b] [Wang et al., 2013c], i.e. they don't require specific adjustments for each BCI user. In the latter case, they are generally chosen in preliminary studies, and subsequently utilized regardless of the user. Their choice is mainly driven by the signal acquisition system embedded in the BCI system. By contrast, most decoders are user-specific because of the high variability of users' model of dependence between brain features and intended BCI outputs. Their structure is often chosen during a feasibility study and fixed for all possible users, but their parameters are, with a few exceptions (e.g., [Fazli et al., 2009]), optimized for each user.

Decoder identification is performed by analysing a data set of simultaneously acquired neuronal signals and intended effector movements. It is carried out after a decoder structure has been selected on the basis of preliminary studies, and consists of tuning this decoder so that it satisfyingly reflects the dependence between neural features and user's intentions within this training data set. This tuning phase is also referred to as decoder "adaptation" [McFarland and Wolpaw, 2011], "learning" (e.g., [Hudson and Burdick, 2007]), "training" (e.g. [Ifft et al., 2013]) or "calibration" (e.g., [Jarosiewicz et al., 2013]). The training data set is often collected during open-loop acquisition sessions, i.e. sessions during which the user is not given feedback on the

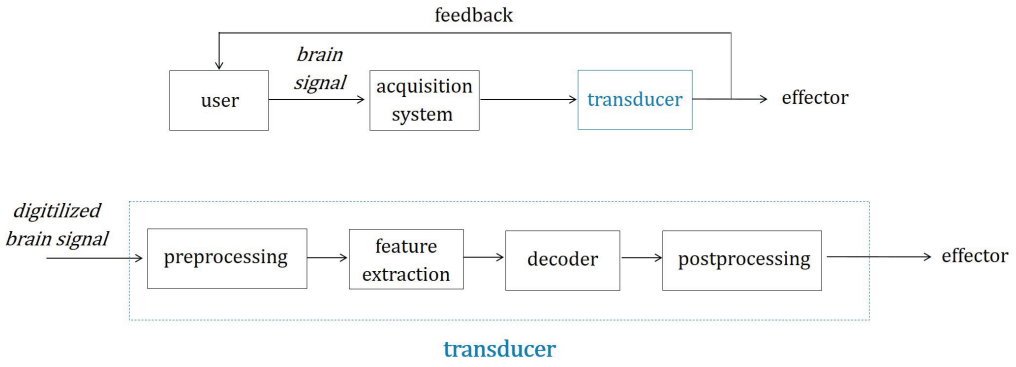

Figure 1.10: Transducer components.

output of the BCI transducer. The future BCI user is generally cued to repeatedly generate action-specific patterns, and his brain activity is recorded throughout this process [Homer and Nurmikko, 2013]. In the case of externally-paced motor BCIs, the user is simply exposed to the stimuli of interest. In both cases, decoders are optimized according to their performance for open-loop data decoding.

Decoder identification on open-loop data generally suffers from two limitations. As the acquired neural signals are noisy and only partially reflect the user brain activity, open-loop decoders rarely reach perfect decoding performance on the open-loop data set. Moreover, high open-loop decoding accuracy does not systematically result in optimal intention estimation in closed-loop settings [Jackson and Fetz, 2011] [Jarosiewicz et al., 2013]. Because of a change of context between open-loop and closed-loop neural signals, open-loop neural patterns differ from closed-loop patterns [Leuthardt et al., 2006a] [Jackson and Fetz, 2011] [Jarosiewicz et al., 2013]. Performance drops are regularly observed when an open-loop decoder is applied during closed-loop experiments [Tillery et al., 2003]. While it has been found that these shortcomings of decoder identification may not impact externally-paced systems [Duprès et al., 2014], they make user training indispensable for internally-paced motor BCIs.

User training (or adaptation) consists of leading the user to adapt to an imperfect decoder, i.e. to modify his brain patterns so as to compensate for the transducer's erroneous outputs. Thanks to the feedback, the BCI user can assess the difference between his true intention and the transducer's estimation of this intention, and progressively learn how to reduce it by modifying his neural activity. User adaptation exploits brain plasticity [McFarland and Wolpaw, 2011], i.e. the brain ability to reorganize to learn new tasks when it is provided with neuro-feedback. While user training is not required for externally-paced motor BCIs, it is a crucial phase of the deployment of internally-paced motor BCIs. It is most of the time completed after the identification of an open-loop decoder [Simeral et al., 2011].

Iterative decoder and user adaptation has been reported in preclinical and clinical motor BCIs [Shenoy and Carmena, 2014] (e.g., [Wang et al., 2013c] [Wodlinger et al., 2015]). After identification of an initial decoder on open-loop data, the user attempts to perform tasks during closed-loop experiments. The decoder is then re-identified so as to take into account modifications in the user's strategy. In most studies, several iterations of user-decoder adaptation are performed ("turn-taking" strategy) [Hochberg et al., 2012] [Wang et al., 2013c], for example 4 to 8 recalibrations for MEA-based prosthesis control [Hochberg et al., 2012].

1.2.3.2 Decoding approaches for motor BCIs

Two main approaches are used to provide users with control over orthoses or prostheses. They are thought to impact the user's ability to control multiple DoF and multi-limb effectors as well as the mental load associated with neural control.

Kinematic decoding A first decoding strategy consists of estimating the commands for effector control from the activity of neurons naturally devoted to the control of the corresponding limb. Continuously-valued kinematic parameters of the effector are generally directly extracted from the corresponding neural signals, for example the position or velocity of an orthosis endpoint. Several terms have been used in the literature to refer to such decoders, in particular "kinematic decoders" [Yuan and He, 2014] and "direct motor Brain Machine Interfaces" [Waldert et al., 2009]. The term "kinematic decoder" has here been arbitrarily chosen, and will be used in the remaining of this doctoral dissertation.

Kinematic transducers exploit neural features correlated with the kinematic parameters of the intended effector movement. Such features were first discovered in the spiking activity of monkeys performing reaching movements [Georgopoulos et al., 1982]. Some neurons of the monkeys' motor cortex were found to preferentially fire when the monkeys' arm wass in one specific direction, i.e. they were found to be directionnally tuned [Waldert et al., 2009] (**Figure 1.11**). Since then, firing rate tuning has been extended to other trajectory characteristics. Correlations have been observed between motor neurons' spiking activity and hand speed, position, velocity and acceleration, target localization, joint motion or muscle activation [Scott, 2008]. Neuron tuning has been found to persist when humans with tetraplegia where attempting to execute arm movements [Hochberg et al., 2006]. This finding suggested that the utilization of kinematic decoders is achievable by motor-impaired patients. While tuned features were first thought to be exclusively present in neurons' spiking activity, an increasing number of studies have investigated the existence of similar features in the activity of neuronal populations [Waldert et al., 2009], e.g., LFP [Mehring et al., 2003] [Mehring et al., 2004], EEG [Waldert et al., 2008] [Bradberry et al., 2010], ECoG [Gunduz, 2008] [Ball et al., 2009b] [Anderson et al., 2012] or MEG [Waldert et al., 2008] [Bradberry et al., 2009] signals. Correlations between neural features in the motor cortex and trajectory kinematics have been disclosed

in ECoG signals [Gunduz, 2008] [Ball et al., 2009b] [Anderson et al., 2012] [Nurse et al., 2015] [Bundy et al., 2016] and in low-pass filtered EEG [Waldert et al., 2008] [Bradberry et al., 2010] [Jerbi et al., 2011] signals in humans.

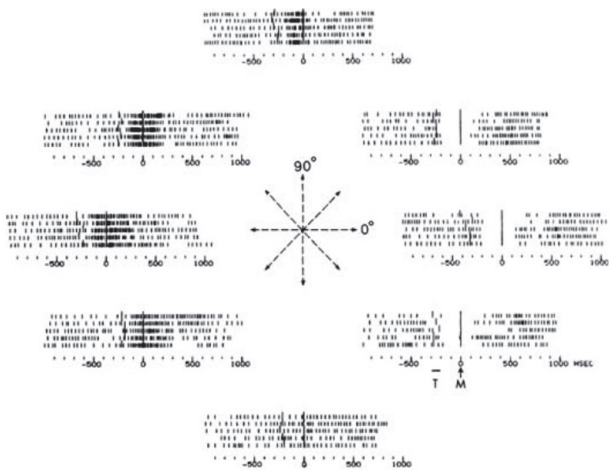

Figure 1.11: Directional tuning of 5 motor cortex neurons (reproduced from [Georgopoulos et al., 1982]).

To date, kinematic transducers have mainly been embedded in invasive BCI systems (e.g. [Ifft et al., 2013] [Wodlinger et al., 2015]) because tuned features have initially be found in neurons' spiking activity. While the feasibility or use of kinematic transducers has mostly been investigated for upper-limb effectors, results suggest that they may also be considered for MEA-driven lower-limb effector control [Fitzsimmons, 2009] [Ma et al., 2017]. When tuned features can be extracted from neural signals, they can be combined by continuous models to estimate the kinematic parameters of the intended trajectory. Typical continuous models include regression models [Collinger et al., 2013] [Wodlinger et al., 2015] [Bundy et al., 2016] and Kalman Filters [Pistohl et al., 2008] [Hochberg et al., 2012]. Tuned features have permitted accurate neural control over multiple degrees of freedom [Collinger et al., 2013] [Wodlinger et al., 2015], or fine movement reconstruction in offline analyses [Chao et al., 2010] [Ofner and Müller-Putz, 2012] [Shimoda et al., 2012] [Bundy et al., 2016].

Kinematic decoders are historically biomimetic, i.e. they aim at exploiting the mapping which related neuronal activity to limb movement before the patient began to suffer from motor disabilities. Their goal is to provide patients with natural control over the effector so as to minimize training. Another type of MEA-based kinematic decoders has nevertheless been explored by a few teams, namely biofeedback decoders [Ganguly and Carmena, 2009] [Sadtler et al., 2014]. Biofeedback decoders also focus on the activity of motor neurons, but they rely on user training rather than on the exploitation of the user's natural map between neuronal activity and limb kinematic parameters. The corresponding assumption is that BCI control can only partially

mimic natural motor control, in particular because MEAs only permit to acquire the activity of a very reduced subset of the neurons involved in motor control. The limits of brain plasticity, i.e. the extent to which a new, unnatural, initially non-intuitive map can be learned by users, have been investigated in a few studies [Ganguly and Carmena, 2009] [Sadtler et al., 2014]. While a monkey was able to learn how to proficiently control a 2D cursor using a decoder tuned with random parameters in [Ganguly and Carmena, 2009], a second closed-loop study suggests that learning cannot be achieved if the initial decoder is too far from the user's natural map [Sadtler et al., 2014]. The respective relevance of biomimetic and biofeeback decoders, in particular in terms of training duration, is still unclear, and is sometimes referred to as the "biomimetic vs. biofeedback" or "decoding vs. learning" dilemma [Carmena, 2013] [Jackson and Fetz, 2011]. To date, most MEA-based motor BCIs nevertheless relied on biomimetic kinematic decoders [Hochberg et al., 2012] [Wodlinger et al., 2015].

Mental-task decoding A second approach consists of using the activity elicited in brain areas which originally were not exclusively devoted to the control of the limb of interest. The brain patterns used to control the prosthesis or orthosis movements are elicited by mental tasks such as motor imageries and cognitive tasks. The term "mental-task decoding" has here been chosen to denote BCI control based on such mental tasks or strategies [Waldert et al., 2009]. Although it can be argued that this term may also apply to kinematic decoders, it has nevertheless been selected because no term has clearly arisen to refer to BCIs based on unnatural motor imageries and cognitive tasks. Incidentally, mental-task based decoders are the basis of the majority of internally-paced communication and navigation BCIs. Various mental tasks have been used to elicit intention-specific and distinguishable brain patterns for neural control in BCI systems [Waldert et al., 2009]. Motor imagery, i.e. the fact to imagine moving a limb, is routinely used for neural control in motor BCIs [Waldert et al., 2009] (e.g., [McFarland et al., 2010] [LaFleur et al., 2013]). Because of the somatotopic organization of the cortex, motor imageries associated with different limbs (e.g., tongue, foot, right arm, left arm etc.) generate patterns which are spatially distinguishable at a macroscopic scale (see **Figure 1.12**) [Waldert et al., 2009]. Mental tasks are not exclusively associated with patterns generated in the motor cortex [Jackson and Fetz, 2011]. Several studies have focused on the discrimination between cognitive tasks [Waldert et al., 2009], e.g. in [Curran et al., 2004] [Penny and Roberts, 1999].

Two types of decoders are used to infer the effector kinematics from the user's neural signals when the transducer relies on the distinction between mental tasks. A first strategy consists in considering discrete user's intentions, i.e. the mental task i is associated with a movement in the direction i. Continuously-valued commands are subsequently inferred from the decoder discrete or probabilistic output [Pistohl et al., 2008]. For example, movement direction can be extracted from the neural signals by the transducer while movement velocity is fixed at a value chosen by the

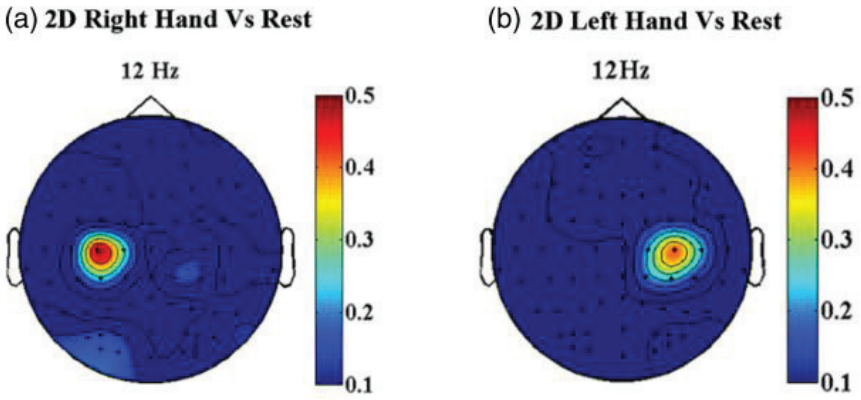

Figure 1.12: Right and left hand motor imagery (EEG) (reproduced from [LaFleur et al., 2013])

experimenter (e.g., [Yanagisawa et al., 2012a] for an ECoG-driven hand prosthesis, or [Leeb et al., 2007] for EEG-based 1D navigation). Alternatively, the effector velocity or position can be proportional to the class probability difference [Obermaier et al., 2001]. Distinction between 5 classes can thus permit navigation in a 2D space (up, down, right, left, rest). A second type of decoder directly relates the modulation of one feature or of a set of features to the effector movement along one axis (e.g., [Wolpaw and McFarland, 2004] [Schalk et al., 2008] for ECoG signals). Control is said to be proportional [Wang et al., 2013c]. The set of features is generally chosen during a screening procedure, which aims at finding features optimally modulated by mental tasks.

Non-invasive acquisition systems are generally associated with mental-task transducers [Waldert et al., 2009] [Milan and Carmena, 2010], which remain efficient when the acquired signals exhibit a limited spatial resolution. 3D EEG neural control over a quadcopter has for example been achieved in [LaFleur et al., 2013], using volitional modulation of SMRs elicited via motor imagery. Similarly, EEG-based neural control permitted users to perform 3D reaching movements in a virtual space in [McFarland et al., 2010]. Mental-task decoders have also been used for cursor or prosthesis control from SUA/MUA [Hochberg et al., 2006] or ECoG signals [Wang et al., 2013c].

Applicability and relevance of kinematic and mental-task decoders The relative performances of kinematic and mental-task decoders remain a matter of discussion. The neural control provided by kinematic transducers is expected to be more precise [Chin et al., 2007] [Nurse et al., 2015] than mental-task-based neural control. Whereas the users of mental-task decoders are likely to encounter difficulties to associate mental tasks with the effector movements, kinematic control is meant to be intuitive [Pistohl et al., 2008] [Schalk et al., 2007] [Ashmore et al., 2012] [Nurse

et al., 2015], to involve a lesser mental load [Yuan and He, 2014] and to require less user training [Waldert et al., 2009]. Although further studies are required to assess the limits of kinematic approaches, kinematic control is often presented as a feature generally desirable for BCI systems, including EEG- [Yuan and He, 2014] or ECoG-based [Chin et al., 2007] BCI systems. The feasibility of kinematic ECoG- and EEG-driven control is still unclear. In particular, it has been shown that the ability of users to execute reaching movements is degraded when estimated positions or velocities are not sufficiently correlated with the user's intentions, e.g. correlation is equal or inferior to 0.75 or 0.5 in the case of position and velocity decoding, respectively [Marathe and Taylor, 2011]. This finding suggests that an accurate extraction of kinematic parameters from the neural signals is required for kinematic decoding to benefit patients. Feasability studies for EEG- [Waldert et al., 2008] [Bradberry et al., 2010] [Jerbi et al., 2011] and ECoG-based kinematic decoders [Gunduz, 2008] [Ball et al., 2009b] [Anderson et al., 2012] [Nurse et al., 2015] [Bundy et al., 2016] have recently been completed. The interest of ECoG-driven kinematic decoders is of particular interest because of the ECoG potential for chronic and complex neural control.

Kinematic decoding of ECoG signals Research on ECoG-based BCI systems historically focused on mental-task decoding strategies [Anderson et al., 2012] (e.g., [Lal et al., 2005] [Hill et al., 2006] [Chin et al., 2007]), and most ECoG-based closed-loop studies have been completed with mental-task decoders [Leuthardt et al., 2004] [Leuthardt et al., 2006a] [Schalk et al., 2008] [Rouse and Moran, 2009] [Williams et al., 2013] [Wang et al., 2013c]. Recent studies have nevertheless suggested that the amount of directional tuning in ECoG signals is sufficient for kinematic decoding. The presence of upper-limb kinematic encoding, in particular position and velocity encoding, has been disclosed in various neural features.

Position encoding has been reported in several studies. In the 2D tracking task completed by human subjects in [Schalk et al., 2007], it permitted to reconstruct trajectories with a correlation of 0.51 and 0.47 for the x- and y-axis, respectively (average over 5 subjects). Position encoding was mainly observed in the Local Motor Potential (LMP, i.e. low-pass filtered ECoG signals) (axes x and y), and in the 140-190Hz and 8-12Hz frequency bands for the x- and y-axis, respectively. In [Bundy et al., 2016], 3D position encoding was mainly found in the LMP and high-gamma frequency bands, and yielded trajectory reconstructions with Pearson Correlation Coefficients (PCC) of 0.49, 0.28 and 0.38 for the x-, y- and z-axis, respectively (average over 5 subjects). In [Hammer et al., 2013], 1D position encoding was disclosed in the LMP features and in the phase and magnitude of frequency features extracted between 30 and 200Hz (for a maximum Pearson Correlation Coefficient of 0.53 at the peak of decoding accuracy, average over 3 subjects). Similarly, 1D position encoding was mainly found in LMP features in [Hammer et al., 2016]. In [Pistohl et al., 2008], a Kalman filter which state variable included position and velocity exploited low-frequency and gamma neural features, and yielded position

accuracy between 0.3 and 0.4 for x-axis, and 0.45 and 0.55 for the y-axis (3 subjects with non-compromised motor recordings). In [Hotson et al., 2014], the first Principal Component of subjects' wrist trajectory was decoded with a correlation coefficient around 0.7 for each one out of 3 subjects. Depending on the subjects, various frequency bands between 0 and 200Hz were found informative to predict the wrist position.

Velocity encoding has been reported in several studies, often in similar features as the position. Velocity encoding found in the LMP and in the 5-42Hz and 140-190Hz frequency bands permitted to reconstruct 2D trajectories with a correlation of 0.43 and 0.55 for the x- and y-axis, respectively (average over 5 subjects)[Schalk et al., 2007]. In [Anderson et al., 2012], neural features extracted from several channels were found to be significantly tuned to velocity during the execution of 2D movements. Average PCC of 0.45, 0.35 and 0.42 for the x-, y- and z-axis respectively were reported between 3D velocity profiles and their ECoG-based reconstructions. In [Hammer et al., 2013], 1D velocity encoding of LMP and various frequency features permitted to reconstruct subjects' wrist velocity with an accuracy comparable to the one obtained for position decoding. It was nevertheless reported in a later study that speed encoding was significantly more encoded than velocity [Hammer et al., 2016], again in LMP-related features. A similar phenomenon was observed in [Bundy et al., 2016] (average Pearson Correlation Coefficient between true and decoded speed of 0.82). Finally, a Kalman filter yielded velocity accuracies between 0.4 and 0.5 for x-axis, and 0.48 and 0.6 for the y-axis (3 subjects with non-compromised motor recordings) [Pistohl et al., 2008].

In [Bundy et al., 2016], speed encoding, where speed refers to the norm of the velocity, yielded reconstructions with PCC of 0.82, thus surpassing both position and velocity encoding (average over 5 subjects). While similar results were obtained in [Hammer et al., 2016], where speed was found to be more encoded than velocity and position, it was suggested in the tuning study completed by Anderson and colleagues [Anderson et al., 2012] that speed tuning is less prominent than 2D position and velocity.

Poor acceleration encoding has been reported in [Hammer et al., 2013] (PCC of 0.3 at the peak of decoding accuracy, average over 3 subjects) and [Hammer et al., 2016].

Finally, 2D directional encoding has been disclosed in high-gamma signals (85-250Hz) in 10% of the channels of arrays partly located over the motor cortex [Nurse et al., 2015] (distinction between 8 targets, i.e. angles of 45°). ECoG 2D directional tuning has also been reported in [Schalk et al., 2007] and [Anderson et al., 2012].

These findings suggest that kinematic decoders may be utilized to interpret ECoG signals during arm movements.

1.2.3.3 Decoders for asynchronous control

The ability of generic kinematic transducers to handle asynchronous decoding, i.e. to limit erroneous activations of the BCI system when the user doesn't desire to

control the effector, is generally poor (see, e.g. [Chao et al., 2010][Williams et al., 2013] [Velliste et al., 2014]). Similarly, mental-task decoders are regularly designed for synchronous decoding only [Doud et al., 2011] [Onose et al., 2012] [LaFleur et al., 2013]. Several studies have therefore been devoted to the development of efficient transducers for asynchronous control. The majority of these studies were completed with simplified frameworks, namely the discrimination between intentional control and idle states [Mason and Birch, 2000] [King et al., 2015] or the distinction of a limited number of mental tasks from idle states [Millán and Mouriño, 2003] [Bashashati et al., 2007b] [Bhattacharyya et al., 2015]. While a few asynchronous kinematic decoders have been proposed by the BCI community [Srinivasan et al., 2007] [Williams et al., 2013], their respective relevance and limits are still unclear. A detailed presentation of the technical solutions considered for asynchronous control is given in **Chapter 2**.

1.2.4 Effectors

The effectors embedded in motor BCI systems exhibit multiple degrees of freedom and/or multiple effectors.

Several prosthetic or orthotic devices have been considered for the restoration of upper-limb mobility. The use of Robotnik's 6-Degrees of Freedom (DoF) JACO robotic arm [Robotnik, 2016] has been reported in several motor BCI studies [Bougrain et al., 2012] [Baxter et al., 2013] [Bhattacharyya et al., 2015]. The Modular Prosthetic Limb, which has been developped at Johns Hopkins University and includes a hand and up to 16 controllable DoF, has been controlled by tetraplegic patients in several clinical trials [Collinger et al., 2013] [Wang et al., 2013c] [Wodlinger et al., 2015]. The DLR (respectively, the DEKA) robotic device utilized in [Hochberg et al., 2012] combines a 7-DoF (6-DoF, respectively) arm with a 15-DoF (4-DoF, respectively) hand. 6-DoF and 5-DoF robotic arms were controlled by a monkey in [Carmena et al., 2003] and [Velliste et al., 2008], respectively. Finally, a custom upper-limb orthosis has been developed within the framework of an EEG-based BCI project [Webb et al., 2012].

Both custom and commercialized hand orthoses and prostheses have been neurally manipulated by BCI users [Pfurtscheller et al., 2000] [Murguialday et al., 2007] [Chen et al., 2009] [Ortner et al., 2011].

Several teams also endeavor to restore lower-limb mobility via neural control over a lower-limb orthesis, e.g. the MINDWALKER, the RoGO, the Rex, the H2 and the Walk Again Project's exoskeleton in [Gancet et al., 2012], [Do et al., 2013], [Kwak et al., 2015], [López-Larraz et al., 2016] and [Nicolelis, 2014], respectively.

Finally, early results on the utilization of Functional Electrical Stimulation (FES) have been presented in a few studies [King et al., 2015] [Bouton et al., 2016] [Vidaurre et al., 2016]. FES-based BCIs aim at enabling users to regain control over their own limbs rather than over an external device. FES consists of stimulating the user's muscles so that they contract according to the user's movement intentions.

The control of physical effectors (e.g., orthoses or robotic arms) is necessary

to restore limb mobility. The utilization of virtual effectors has nevertheless been proposed in several studies, in particular because it permits to facilitate early training phases in clinical trials [Wodlinger et al., 2015]. Users have been provided with control over virtual effectors in many preclinical or clinical motor studies, e.g. cursors [Taylor et al., 2002] [Leuthardt et al., 2006a] [Kim et al., 2008] [Simeral et al., 2011] or simulated robotic arms in virtual reality environments [Wang et al., 2013c] [Ifft et al., 2013] [Wodlinger et al., 2015].

1.2.5 Feedback for prosthesis or orthosis control

The ability of the users to interpret and use the feedback conveyed to them is crucial for motor BCI systems. In the vast majority of motor BCIs, users are exclusively given visual feedback about the transducer output (e.g., MEA- [Kim et al., 2011] [Hochberg et al., 2012] [Collinger et al., 2013] [Wodlinger et al., 2015], ECoG- [Vansteensel et al., 2010] [Milekovic et al., 2012] [Kellis et al., 2012] [Yanagisawa et al., 2012a] [Wang et al., 2013c], and EEG-based clinical trials [Wolpaw and McFarland, 2004] [Yuan et al., 2007] [McFarland et al., 2010] [Doud et al., 2011] [LaFleur et al., 2013]). The addition of other types of feedback, e.g. kinesthetic feedback [Suminski et al., 2010], has nevertheless been shown to facilitate upper-limb prosthesis control. Haptic feedback was used to improve neural control over a hand prosthesis in [Murguialday et al., 2007]; a few teams have completed feasibility [Cincotti et al., 2007] or cursor control [Chatterjee et al., 2007] studies with vibrotactile feedback. Finally, combinations of feeback modalities [Suminski et al., 2010] and intracortical stimulation [O'Doherty et al., 2011] have been investigated.

1.2.6 Summary on current progress

To date, BCI clinical applications mostly consist of non-invasive systems for communication or environmental control [Mak and Wolpaw, 2009], e.g. EEG-driven spellers [G.tex, 2016]. Because of the difficulties encountered to meet the requirements presented in **section 1.2.1**, motor BCIs have not yet been deployed for everyday use [Mak and Wolpaw, 2009].

While EEG-based motor BCIs have the significant advantage of being safe, a long training process is generally necessary before the user is able to adapt to the mental-task decoder they embed. To date, the corresponding complexity of control additionally remains inferior to the one associated with MEA-based BCIs, for example 2D-control over a robotic arm with 4 possible directions [Hortal et al., 2015], 3D synchronous control [LaFleur et al., 2013], or 2D control decomposed into sequences of 1D movements [Bhattacharyya et al., 2015].

Invasive or semi-invasive acquisition methods (MEA, ECoG) may be a promising alternative to EEG for highly accurate, multiple DoF and multi-limb control [Lebedev and Nicolelis, 2006]. Although invasive and semi-invasive motor BCI clinical trials are still scarce [Mak and Wolpaw, 2009] because of the potential safety issues they are associated with, clinical proofs of concept studies have neverthess been completed

in laboratory environments [Hochberg et al., 2012] [Collinger et al., 2013] [Wang et al., 2013c] [Wodlinger et al., 2015].

The feasibility of 3D [Hochberg et al., 2012], 7D [Collinger et al., 2013] and 10D [Wodlinger et al., 2015] neural control over a robotic arm has been demonstrated in recent MEA-based studies. The integration of multi-limb effector control into motor BCIs has particularly been considered in a few studies [Hochberg et al., 2012] [Ifft et al., 2013] [Wodlinger et al., 2015] [Bouton et al., 2016]. In [Hochberg et al., 2012], both sequential or parallel MEA-based control over an upper-limb prosthesis endpoint and a prosthetic hand were achieved by users with tetraplegia. Parallel control over the wrist and hand of a robotic arm was additionally reported in [Wodlinger et al., 2015]. While numerous daily life tasks require bimanual movements [Swinnen and Wenderoth, 2004], bimanual control has only been reported over virtual effectors [Ifft et al., 2013]. While these studies suggest the relevance of MEA acquisition systems and of the reported decoder structure and training strategy, the issues pertaining to the MEA invasiveness, namely safety and chronicity [Vouga et al., 2017], are to date only partially addressed.

While ECoG arrays hold promise of chronic and stable signal acquisition, the reported ECoG-driven motor BCIs generally relied on mental-task decoders [Schalk et al., 2008] [Fifer et al., 2014] [Wang et al., 2013c] [Kapeller et al., 2015] and did not permit users to achieve complex effector control. Control over a set discrete commands was achieved using ECoG signals in [Fifer et al., 2014] and [Hotson et al., 2016], and 3D [Wang et al., 2013c], 2D [Schalk et al., 2008] and 1D control [Vansteensel et al., 2010] [Leuthardt et al., 2011] has been reported in a few studies. If kinematic control has not yet been completed in human subjects, 2D kinematic control was accomplished by monkeys in [Marathe and Taylor, 2013]. ECoG-based multi-limb control has mainly been considered in the case of multi-finger offline trajectory reconstruction [Wang et al., 2009b] [Wissel et al., 2013] [Saa et al., 2016] [Liang and Bougrain, 2012] [Kubánek et al., 2009] [Acharya et al., 2010] [Flamary and Rakotomamonjy, 2012] [Nakanishi et al., 2014b]. Individual finger ECoG-based control was nevertheless restored in [Hotson et al., 2016]. To date, the degree of complexity achieved with ECoG-driven control is consequently surpassed by the ones reported for MEA-based BCIs. Synchronous protocols have mainly been considered, and studies on ECoG-based effector chronic control are still lacking. The proof that chronic asynchronous ECoG control over multi-limb multi-DoF effectors is feasible thus remains to be established.

1.3 CLINATEC's motor BCI project

CLINATEC's BCI project is meant to contribute to the recent efforts towards motor BCI clinical applications [Eliseyev et al., 2014]. It aims at allowing patients with tetraplegia to chronically recover both mobility and arm control in the framework of a 5-year clinical trial recently approved by the French competent authorities [ClinicalTrials.gov, 2016] (see **Figure 1.13**).

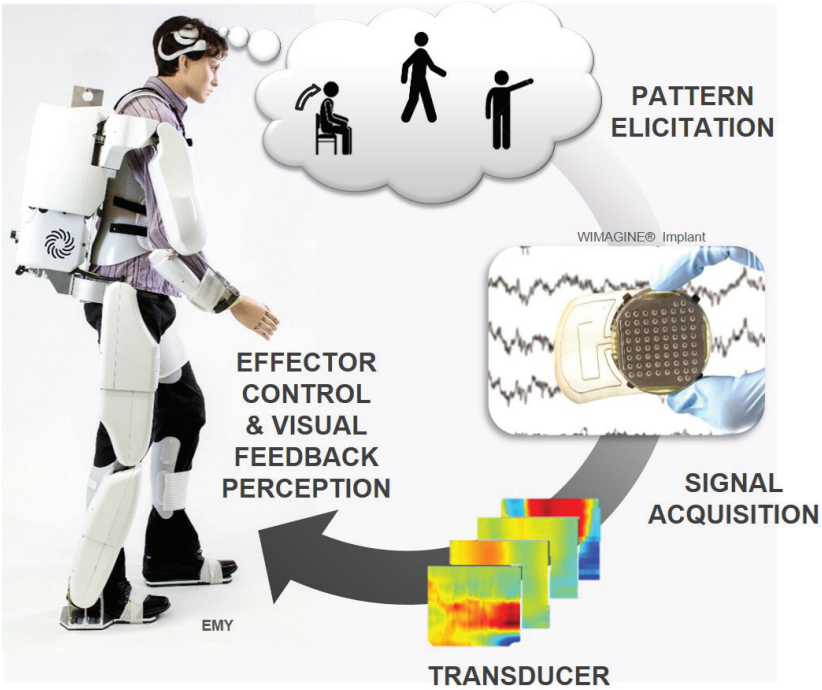

Figure 1.13: CLINATEC's BCI system (in particular, WIMAGINE implant®, transducer, EMY exoskeleton).

1.3.1 Signal acquisition

To date, obstacles to MEA-based chronic signal acquisition remain to be fully removed, and the performances of EEG-based motor BCIs may be limited by EEG spatial resolution and frequency content. An alternative approach has therefore been selected for CLINATEC's motor BCI system, namely an ECoG acquisition system. The ECoG capacity to address the challenges of motor BCI systems will be explored during CLINATEC's clinical trial. Because of the lack of chronic ECoG implants (see **section 1.2.2.3**), a wireless ECoG implant has been especially designed for long-term BCI applications [Mestais et al., 2015]. CLINATEC's ECoG implant WIMAGINE®, shown in **Figure 1.14**, gathers 64 electrodes on a 5cm-diameter surface. A distance of approximatively 4mm separates the electrodes' centres [Mestais et al., 2015]. CLINATEC's BCI system has been developed so as to rely on internally-paced neural control, and WIMAGINE® has been designed for an implantation on the user's motor cortex, i.e. for the acquisition of users' SMR. WIMAGINE®'s safety has been approved by the competent authorities [Mestais et al., 2015]. In particular, its implantation is performed by means of a standardized craniotomy procedure which limits post-implantation complications.

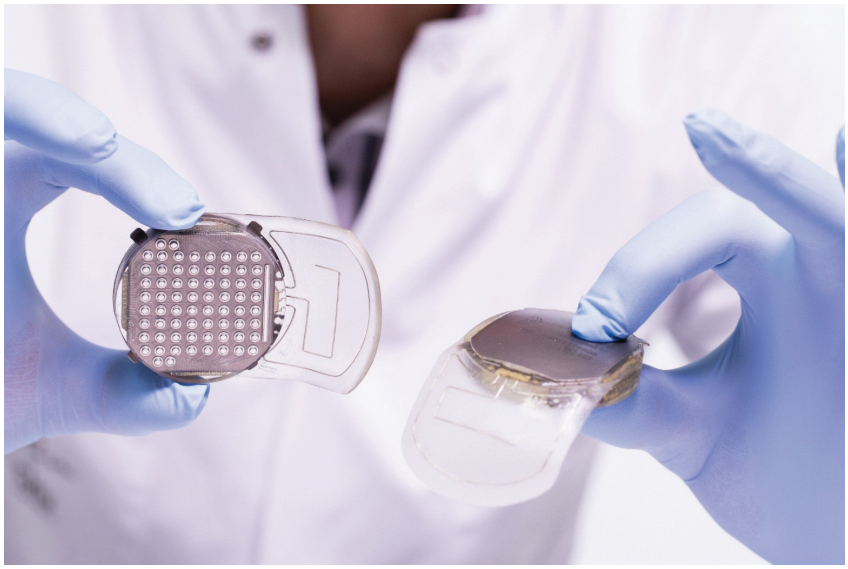

Figure 1.14: WIMAGINE® implants.

1.3.2 Transducer

The transducers developed in CLINATEC were mainly devoted to two signal processing challenges, namely accurate mono-limb and asynchronous control.

Kinematic decoding for accurate mono-limb control In view of the recent studies which have suggested that the amount of directional tuning in ECoG signals

may be sufficient for kinematic decoding (see **section 1.2.3.2**), CLINATEC's signal processing team has developed a kinematic decoding approach for the restoration of upper-limb mobility. The transducer first extracts high-dimensional time-frequency-space features from the neural signals. A preliminary pre-processing approach is used to discard artefacts where necessary [Eliseyev and Aksenova, 2014]. The kinematic parameters of interest are then regressed against the neural features using the efficient decoder training approaches designed in CLINATEC [Eliseyev and Aksenova, 2013] [Eliseyev and Aksenova, 2014]. Offline validation of the corresponding transducers has been completed by using ECoG signals to reconstruct the 3D kinematic parameters of monkeys' wrist and/or shoulder and elbow [Eliseyev and Aksenova, 2013] [Eliseyev and Aksenova, 2014].

State detection for asynchronous decoding First steps towards asynchronous decoding have additionally been completed by CLINATEC's team, namely the development of efficient binary switches for the discrimination between discrete mental states (e.g., idle states against desire to control the effector) [Eliseyev et al., 2011] [Eliseyev et al., 2012]. The considered binary transducers rely on the same feature extraction procedure than CLINATEC's kinematic transducers. While online preclinical studies have permitted to validate the relevance of these transducers for binary state prediction [Costecalde, 2012], their integration into kinematic decoders had not been explored yet.

1.3.3 Effector

Clinatec BCI system aims at providing users with multiple DoF control over the 4-limb exoskeleton EMY (for Enhancing MobilitY) [Morinière et al., 2015] [Eliseyev et al., 2014]. Developed at CEA-LIST, EMY (shown in **Figure 1.13**) presents up to 14 controllable degrees of freedom (4 for each arm [Morinière et al., 2015], 3 for each leg). The projected operating protocol consists in offering 7 degrees of freedom (DoF) asynchronous control to patients (3 DoF for each arm, a switch to activate or deactivate walking).

1.3.4 Feedback

Although the integration of alternative feedback types is envisaged for a future version of Clinatec's BCI system, neural control will initially rely on visual feedback only.

1.4 PhD objectives

The present doctoral work has been completed within the framework of CLINATEC's motor Brain-Computer-Interface (BCI) project. It tackles three challenges specific to the development of transducers for motor BCI clinical applications, namely

asynchronous decoding, multi-limb decoding and decoding accuracy during active states.

1.4.1 Asynchronous control

Although asynchronous control has been considered in a few motor clinical BCIs, most motor clinical trials have been completed using a synchronous protocol [Hochberg et al., 2006] [Wodlinger et al., 2015]. The deployment of synchronous BCI systems requires the presence of an operator to switch the system on and off. The impact of the BCI system on users' independence is therefore limited. By contrast, asynchronous BCIs are potential stand-alone systems. Generally, users alternate between periods of Intentional Control (IC) and of No-Control (NC), during which they don't intend to use the BCI system [Mason et al., 2006]. The limitation of erroneous activations of the BCI system during NC states is all the more important since users of motor BCIs physically interact with the effector, in contrast with BCIs based on the control of a virtual effector (e.g., a cursor on a computer screen). False activations are likely to be particularly disturbing and stressful to users. If BCI outputs are not accurate (e.g., neutral) during the NC epochs, users need to actively force them to neutral values. A high mental load is required by such constantly-engaged control paradigms. While NC support is highly desirable for motor BCIs [Leeb et al., 2007], its integration into kinematic decoders has only been partially addressed in the literature. A decoder, referred to as Switching Markov Linear Model, has been designed in the present doctoral work to perform asynchronous kinematic decoding with a limited number of false activations.

1.4.2 Multi-limb control

Multi-limb control is desirable for patients with tetraplegia, in particular right and left upper- and lower-limb control and hand and/or finger control. The integration of multi-limb effector control into motor BCIs has only been considered in a few studies [Hochberg et al., 2012] [Ifft et al., 2013] [Wodlinger et al., 2015] [Bouton et al., 2016]. In particular, the issue of asynchronous sequential multi-limb control has rarely been tackled. The proposed decoder, the Switching Markov Linear Model, has been developed so as to address the challenge of sequential multi-limb asynchronous decoding in addition to the one of mono-limb asynchronous decoding, i.e. to prevent parallel limb activations.

1.4.3 Accurate control

An accurate extraction of kinematic parameters from the neural signals is required for kinematic decoding to enable patients to interact with their environment (see **section 1.2.3.2**). Focus has thus been set on the improvement of decoding accuracy during intentional control states. A specific configuration of the proposed decoder has been utilized to improve decoding accuracy during IC periods.

Proper transducer design is necessary to achieve accurate, asynchronous multi-limb control. The next chapter (**Chapter 2**) focuses on the technical solutions which have been considered by the BCI community to address these issues. The transducer proposed in the present doctoral work, namely the Markov Switching Linear Model, is presented in **Chapter 3**. **Chapter 4** consists in a description of the data sets used to assess the performance of the Markov Switching Linear Model. The procedure completed to measure the decoder performance and the benchmark decoders are exposed in **Chapter 5**. Details on the transducer implementations are given in **Chapter 6**. Results are reported in **Chapter 7**. Finally, perspectives are presented in **Chapter 8**.

Chapter 2

Data-driven transducers for motor BCIs

Contents

Accurate asynchronous neural control over multi-limb prostheses or orthoses requires relevant online processing methods to be applied on the user's neural signal. Translation of the user's brain activity into control signals is performed by the BCI transducer [Bashashati et al., 2007a]. The present chapter presents the algorithmic tools which have been utilized in motor BCI transducers to address the challenge of accurate, asynchronous and/or multi-limb control. As CLINATEC's motor BCI relies on internally-paced neural patterns, focus has been set on data-driven transducers, i.e. on transducers which decoder is user-specific and trained on neural data before utilization.

2.1 Notations and processing blocks

Let $\mathbf{s}^t \in \mathbb{R}^{nc}$ be a multi-channel raw neural signal, where $nc \in \mathbb{N}^*$ is the number of channels. Let $\mathbf{y}^t$ denote the corresponding user movement intention, $\mathbf{y}^t \in \mathbb{N}$ if discrete-valued intentions are considered (e.g., movement towards the right or the left) and $y^t \in \mathbb{R}^n$ if continuously-valued intentions are to be decoded. When the BCI system is turned on ("executed"), its transducer regularly issues an estimate $\hat{\mathbf{y}}^t$ of the user's intention $\mathbf{y}^t$ from neural signals $\mathbf{s}^t$. The computation of $\hat{\mathbf{y}}^t$ is usually decomposed into several processing steps, namely signal pre-processing, feature extraction, decoder application and post-processing (see **Figure 2.1.B**).

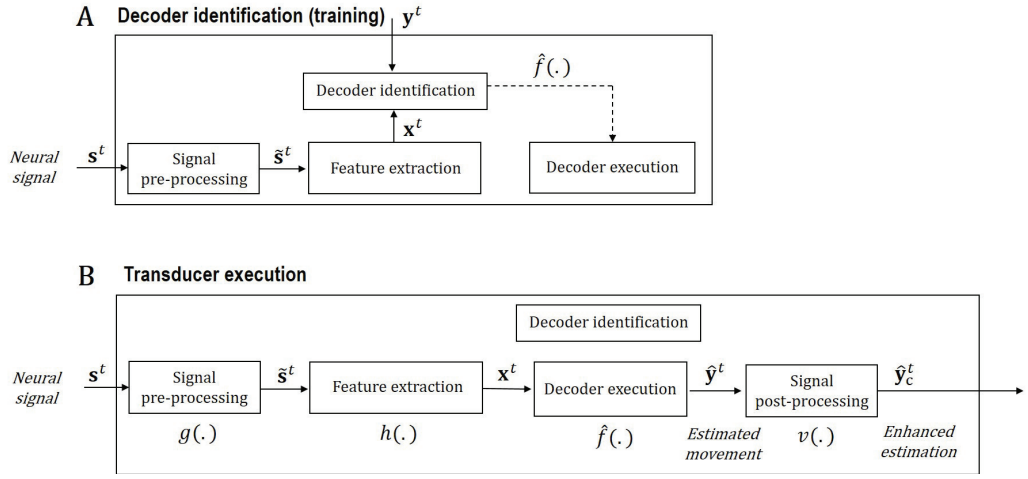

Figure 2.1: General processing block. **A**. Decoder identification stage. **B**. Transducer execution stage.

First, an optional pre-processing transform is applied on the instantaneous raw signal $\mathbf{s}^t$ or on an τ_1-long epoch of raw signal $\mathbf{s}^{(t+1-\tau_1):t}$. It outputs an enhanced signal $\tilde{\mathbf{s}}^t \in \mathbb{R}^{nc}$, for example a signal with an increased Signal-to-Noise Ratio. A transform $h : \mathbb{R}^{nc \times \tau_2} \to \mathbb{R}^m$ is then used to extract features $\mathbf{x}^t \in \mathbb{R}^m$ from an τ_2-long epoch of enhanced neural signal $\tilde{\mathbf{s}}^{(t+1-\tau_2):t}$. Feature extraction permits to build a new representation of neural signals, bringing out signals' informative attributes and discarding redundant or irrelevant characteristics. Next, a function $\hat{f} : \mathbb{R}^m \to \mathbb{R}^n$ or $\hat{f} : \mathbb{R}^m \to \mathbb{N}$ is applied on the neural features $\mathbf{x}^t$. This "decoder" permits to infer an estimate $\hat{\mathbf{y}}^t = \hat{f}(\mathbf{x}^t)$ of the user intention $\mathbf{y}^t$ from neural features. The decoder output $\hat{\mathbf{y}}^t$ is then optionally post-processed. The enhanced estimate $\hat{\mathbf{y}}^t_c$ is sent to the effector controller. The latter drives the effector according to the decoded user intention.

Before executing the BCI transducer, it is necessary to choose optimal transforms for each of its components. In particular, the decoder $\hat{f}$ is generally identified using a training data set $(\mathbf{X}, \mathbf{Y})$, where $\mathbf{X} = \{\mathbf{x}^t\}_{t=1}^T$ and $\mathbf{Y} = \{\mathbf{y}^t\}_{t=1}^T$. This

preliminary modelling phase is illustrated in **Figure 2.1.A**. It is referred to as "decoder identification", "training" or "calibration".

2.2 Pre-processing

Decoders are liable to issue erroneous intention estimates when they are trained and/or applied on neural signals which are noisy or corrupted by artefacts. Pre-processing filters aim at discarding artefacts and/or increasing the Signal-to-Noise-Ratio (SNR). They are profitably used for various decoding tasks, e.g. kinematic or asynchronous binary decoding.

Artefact management Artefacts are signals which are irrelevant for neural control, e.g. signals with non-cerebral sources like ocular movements or heartbeats [Fatourechi et al., 2007a]. The EEG sensitivity to artefacts, in particular artefacts with ocular, cardiac or muscular origins, has been demonstrated in several studies [Fatourechi et al., 2007a] [Whitham et al., 2007] [Muthukumaraswamy, 2013]. Albeit to a lesser extent, ECoG signals are also liable to be corrupted by artefacts, e.g. blink artefacts in subdural ECoG signals [Ball et al., 2009a] or chewing artefacts in epidural ECoG signals [Shimoda et al., 2012]. Finally, the presence of common-noise artefacts has been reported in intracortical recordings [Paralikar, 2010].

Artefact rejection and artefact removal methods have been developed to limit artefact-induced perturbations during decoder training and/or execution. Manual or automatic artefact rejection consists of discarding corrupted samples [Fatourechi et al., 2007a] and is mainly performed for offline cleaning of training data before model identification (e.g., [Kubánek et al., 2009] [López-Larraz et al., 2016]). Visual inspection [Kubánek et al., 2009] or outlier detection [López-Larraz et al., 2016] are example of strategies which were used to reject artefacts before decoder training. By contrast, artefact removal methods aim at correcting neural signals corrupted by artefacts and are suited for online application.

Main artefact removal methods are temporal filtering, linear regression and spatial filtering [Fatourechi et al., 2007a].

Temporal filtering is used to reject frequency bands containing artefacts [Fatourechi et al., 2007a], e.g. low-pass filtering to discard muscular artefacts which are mainly observed between 20 and 300Hz [Muthukumaraswamy, 2013]. Low-pass filtering below 30Hz has for example been reported in several EEG-based motor BCI systems [Sadeghian and Moradi, 2007] [Herman et al., 2008]. Band-stop Notch filters are regularly applied to remove power line artefacts at 50 or 60Hz, e.g. in ECoG signals [Anderson et al., 2012] [Nurse et al., 2015] [Spüler et al., 2016]. The potential rejection of frequency bands of interest is a shortcoming of temporal filtering [Fatourechi et al., 2007a]. In EEG-based BCIs, temporal filtering-based rejection of muscular activity results in the elimination of gamma activity, which has been found relevant for limb kinematic reconstruction from ECoG signals [Leuthardt et al., 2004] [Anderson et al., 2012] [Nurse et al., 2015] [Bundy et al., 2016].

Removal methods based on linear regression [Fatourechi et al., 2007a] don't suffer from this drawback (e.g., [Trejo et al., 2006]). They are applicable as long as a signal containing the artefact is acquired simultaneously with the neural signals [Urigüen and Garcia-Zapirain, 2015], e.g. the user's ElectroOculoGraph (EOG) and ElectroCardioGraph (ECG) in case of ocular and cardiac artefacts, respectively [Wallstrom et al., 2004] [Waser and Garn, 2013]. The enhanced signals are computed by subtracting the reference artefactual signal from the neural signals, i.e. they are a linear combination of the corrupted signals and of the reference artefactual signal. This approach is suboptimal if the reference artefactual signal contains some neural signals of interest [Fatourechi et al., 2007a].

Spatial filtering is another strategy which use has been reported for artefact removal [Fatourechi et al., 2007a], for example from ECoG signals in [Liu et al., 2004] or EEG signals in [Brunner et al., 2007]. Online acquisition of pure artefactual signals is not necessary when artefact removal is completed with spatial filters such as Principal Component Analysis (PCA) and Independent Component Analysis (ICA) [Urigüen and Garcia-Zapirain, 2015]. Artefact extraction relies on the projection of the neural signals onto a new basis ("virtual channels"). The subspace A spanned by the first k virtual channels is associated with the signals of interest (e.g., with neural sources), and the complementary subspace with signals with artefactual origin. Online artefact removal consists of projecting the neural signals onto the subspace A. The corrupted virtual channels are thus discarded [Urigüen and Garcia-Zapirain, 2015]. The subspaces associated with the signals of interest or the artefacts are identified during preliminary offline analyses. The use of PCA, which yields uncorrelated virtual channels of decreasing variance, has for example been considered for subspace identification within the framework of magnetoencephalographic data cleaning [Kelly et al., 2011]. A stronger assumption of statistical independence between the signals of interest and the artefacts is exploited by ICA-based algorithms. ICA-based virtual channels have for example been used in several EEG-based BCI studies [Vigário, 1997] [Iriarte et al., 2003] [Brunner et al., 2007].

The relevance of alternative artefact removal approaches has been investigated in several BCI studies [Eliseyev and Aksenova, 2014] [Foodeh et al., 2016]. Chewing artefacts in ECoG signals were for example detected as outliers in time-frequency features and replaced by neutral values in [Eliseyev and Aksenova, 2014].

Despite several demonstrations of the presence of strong artefacts in EEG signals [Fatourechi et al., 2007a] [Muthukumaraswamy, 2013], the use of artefact removal methods is rarely indicated in EEG-based motor BCI studies. For example, attempts at artefact monitoring and/or management was not mentionned in the EEG-driven motor BCI systems developed in [Chae et al., 2012] [Webb et al., 2012] [Baxter et al., 2013] [Karin and Andres, 2014] and [Vidaurre et al., 2016]. Similarly, artefact removal is infrequently mentioned in ECoG-based motor BCIs. In [Trejo et al., 2006], however, ocular artefacts were subtracted from EEG signals during a virtual effector (namely, a cursor) online control task. The use of simple approaches like visual signal monitoring [Doud et al., 2011] [Onose et al., 2012] [LaFleur et al., 2013] or the exclusion of channels liable to include artefactual patterns [Fifer et al., 2014]

[Milekovic et al., 2012] [Lew, 2012] have also been reported in motor EEG- and ECoG-based BCIs.

Signal enhancement Similarly, both spatial or temporal filters can be applied to increase the SNR.

Most reported spatial filters are the Common Average Reference (CAR), the Surface Laplacian (SL) and the Common-Spatial-Pattern (CSP) filters, along with the already mentioned PCA and ICA [Bashashati et al., 2007a]. The use of CAR filtering has been reported in numerous EEG- [Lew, 2012] [Galán et al., 2008] and ECoG-based [Chao et al., 2010] [Shin et al., 2012] [Hammer et al., 2013] [Schalk et al., 2007] [Hotson et al., 2014] [Fifer et al., 2014] motor BCI studies. It consists of computing the average value of the neural signals and in subtracting this value from all channels. SL-based pre-processing is also regularly performed in motor EEG-based BCI systems [McFarland et al., 2010] [Chae et al., 2012]. Signals' second spatial derivatives are extracted to discard patterns which are similar for all electrodes, and are therefore associated with deep sources [Carvalhaes and De Barros, 2015]. By contrast, local and superficial sources are preserved. Finally, CSP filtering is frequently applied for neural signal preprocessing before EEG classification [Townsend et al., 2004] [Sadeghian and Moradi, 2007] [Onose et al., 2012]. Its use has also been reported for offline [Wei and Tu, 2008] and online [Onaran et al., 2011] [Marathe and Taylor, 2013] [Kapeller et al., 2015] ECoG processing. The CSP constructs virtual channels under the criterion that their variance ratio is maximized [Blankertz et al., 2008].

In addition to their application for artefact filtering, temporal filters are used to avoid aliasing effects. Anti-aliasing filters are low-pass filters which permit to respect the Nyquist-Shannon sampling theorem (Nyquist frequency), which states that sampling frequency should be at least twice superior to the maximal frequency present in the signal. Anti-aliasing filters have been applied in numerous EEG-based motor BCIs, e.g. [Flotzinger et al., 1994] [Sadeghian and Moradi, 2007] [Herman et al., 2008]. They were as well utilized in motor ECoG-based studies, e.g. [Schalk et al., 2007] [Bundy et al., 2016].

Finally, low-pass (< 300Hz) and high pass (> 300Hz) filtering of MEA signals is performed to extract LFP and enhance MUA/SUA signals, respectively [Waldert et al., 2009].

2.3 Feature extraction

The application of efficient feature extraction algorithms is a prerequisite for all decoding tasks, e.g. synchronous or asynchronous kinematic reconstruction as well as mental task discrimination.

2.3.1 Neural feature extraction

The application of several classes of transforms $h : \mathbb{R}^{nc \times \tau_2} \rightarrow \mathbb{R}^m$ has been reported for the extraction of neural features.

2.3.1.1 Spike count

Spike counts in short time bins (i.e., neuron instantaneous firing rate) are often used to characterize SUA/MUA signals in intracortical motor BCIs (e.g., [Collinger et al., 2013] [Wodlinger et al., 2015]). Spike detection is generally performed by thresholding the neural signal. In the case of MUA signals, an additional spike sorting step is sometimes carried out to decouple the activity of each observed unit. Although spike-based features have been found to encode information relevant for trajectory decoding, their extraction cannot be performed from ECoG signals.

2.3.1.2 Temporal features

Temporal features, i.e. sequences of instantaneous neural signal characteristics, have been exploited in several BCI systems. Reported features include signals' amplitude at particular instants [Palankar et al., 2008] [Flamary and Rakotomamonjy, 2012] and signals' latency with respect to a stimulus [Mak and Wolpaw, 2009]. Raw signals are possibly low-pass filtered [Flamary and Rakotomamonjy, 2012] or averaged [Palankar et al., 2008] beforehand to increase their SNR.

The amplitude of neural signals after low-pass filtering has been found to encode useful information for the reconstruction of limb kinematic parameters from ECoG signals (see **section 1.2.3.2**). It is thus regularly exploited in ECoG-based motor BCIs [Hammer et al., 2016] [Pistohl et al., 2008] [Ball et al., 2009b] [Kellis et al., 2012] [Milekovic et al., 2012] [Wang et al., 2012] [Hotson et al., 2014] [Schalk et al., 2007]. Low-pass filtering is generally performed by means of a Butterworth [Hammer et al., 2016], Savitzky-Golay [Pistohl et al., 2008] [Ball et al., 2009b] [Kellis et al., 2012] [Milekovic et al., 2012] or Moving Average [Wang et al., 2012] [Hotson et al., 2014] filter.

Most reported ECoG-based BCIs more precisely relied on the combination of temporal, amplitude-based features with time-frequency features [Schalk et al., 2007] [Pistohl et al., 2008] [Wang et al., 2012] [Hotson et al., 2014].

2.3.1.3 Time-frequency and time-scale features

Neural signals are often characterized by their spectral content in short time windows, especially in the case of field potentials-based motor BCIs (i.e. LFP [Aggarwal et al., 2013] [Flint et al., 2013], ECoG [Wang et al., 2013a] [Chin et al., 2007] [Yanagisawa et al., 2012b] and EEG [Pfurtscheller et al., 2000] [Trejo et al., 2006] [LaFleur et al., 2013] [Hortal et al., 2015]). A sine wave is fully described by its frequency, its amplitude and its phase. Both amplitude- and phase-based features have been used in motor BCIs. Their temporal evolution within specific frequency bands is usually

considered. Most reported internally-paced ECoG- and EEG-based motor BCIs relied on such time-frequency features. Different methods have been utilized for the extraction of time-frequency features from neural signals.

Amplitude Both power- and magnitude-based features have been extracted to characterize neural signals in motor BCIs [Wang et al., 2013a] [Chin et al., 2007] [Yanagisawa et al., 2012b] [Eliseyev and Aksenova, 2014]. The instantaneous power of a signal x^t is defined as its squared amplitude $|x^t|^2$, and its magnitude by its absolute value $|x^t|$. In most BCI studies, the power or magnitude average value is then computed in temporal windows and used as a neural feature [Ball et al., 2009b]. As power features are extracted from some specific frequency bands, they are often referred to as "band-power" features. Different strategies have been used to extract amplitude information in motor BCIs.

Bank filter is one of the approaches which use has been reported for the extraction of amplitude-based time-frequency features from EEG [Bashashati et al., 2015] and ECoG [Wang et al., 2013b] [Wang et al., 2013a]. It consists of using a set of real-valued band-pass filters, e.g. Butterworth filters [Shin et al., 2012], to extract signal components included in specific frequency bands [Brodu et al., 2011]. The instantaneous or average power or magnitude within a particular frequency band is then estimated from the amplitude of the filtered signals. In some studies, the amplitude of the filtered signal was post-processed [Pistohl et al., 2008].

The application of the Short-Time-Fourier-Transform (STFT) has also been reported for feature extraction from ECoG [Chin et al., 2007] [Yanagisawa et al., 2012b] signals. The STFT is obtained by applying a temporal window w (e.g., a Hamming [Chin et al., 2007], Hann, Bartlett or Gaussian window) on neural signals before computing their Fourier transform. The squared amplitude of the Fourier descriptors $|STFT(t, w, f)|^2$, where f is the frequency of interest and t characterizes the considered instant, is used to compute the signal spectrogram.

Another approach reported in several EEG-, MEA- ECoG-based BCI studies consists of applying a wavelet transform to compute a time-frequency representation of neural signals [Chao et al., 2010] [Bhattacharyya et al., 2011] [Shimoda et al., 2012] [Eliseyev and Aksenova, 2014] [Bashashati et al., 2015] [Bouton et al., 2016]. The use of different wavelets has been investigated, e.g. Daubechies [Bhattacharyya et al., 2011] [Bouton et al., 2016], Meyer [Eliseyev et al., 2012], Haar [Kousarrizi et al., 2009] or Morlet wavelets [Lemm et al., 2004] [Chao et al., 2010] [Eliseyev and Aksenova, 2014] [Bashashati et al., 2015]. Wavelets are real- or complex-valued. Both real [Chao et al., 2010] and complex Morlet [Eliseyev and Aksenova, 2014] wavelets have for example been used for 3D wrist trajectory decoding from ECoG signals. Complex-valued wavelets can be used to extract phase information from the analysed signals [Fernandes et al., 2003] (see next paragraph). In contrast with the STFT, the temporal resolution of the wavelet transform depends on the considered frequency. Temporal resolution is lower for low-frequency bands, but is compensated

by a higher frequency resolution [Brodu et al., 2011]. Wavelet-based extraction of the instantaneous power [Chao et al., 2010] [Shimoda et al., 2012] or magnitude [Eliseyev et al., 2012] [Eliseyev and Aksenova, 2014] at specific instant has for example been reported for the extraction of kinematic parameters from ECoG signals. The squared amplitude of wavelet transform is referred to as "scalogram", and the corresponding features as "time-scale" features.

The Hilbert transform has also been used in BCI studies, for example to analyse the power spectral density in several frequency bands of ECoG signals [Yanagisawa et al., 2012a].

Finally, parametric spectrum estimation is a popular approach for the characterization of both ECoG [Lal et al., 2005] [Leuthardt et al., 2004] [Hill et al., 2006] [Felton et al., 2007] [Schalk et al., 2007] [Schalk et al., 2008] [Blakely et al., 2009] [Leuthardt et al., 2011] [Ashmore et al., 2012] [Wang et al., 2012] [Wang et al., 2013c] [Fifer et al., 2014] and EEG signals [Schlögl et al., 2005] [Argunşah and Çetin, 2010]. The Auto-Regressive (AR) coefficients of the neural signals are estimated, for example via the Yule-Walker [Herman et al., 2008] or Burg method [Ashmore et al., 2012] [Fifer et al., 2014]. Spectrum estimation is inferred from the AR parameters. When it is based on Burg AR paramers, it is referred to as Maximum-Entropy Spectral Estimation. Maximum-Entropy Spectral Estimation has been performed in several ECoG studies [Anderson et al., 2012] [Bundy et al., 2016] [Spüler et al., 2016].

The above-mentioned time-frequency and time-scale features have rarely been compared on a common data set. Two extensive studies have nevertheless been completed on EEG signals elicited by motor imageries [Herman et al., 2008] [Brodu et al., 2011]. The periodogram and parametric power estimation approaches permitted to extract features associated with the best classification accuracy in [Herman et al., 2008], whereas Morlet wavelet transforms surpassed alternative methods in [Brodu et al., 2011]. These inconsistent results seem to suggest that the relevance of different time-frequency and time-scale features partially depends on the data sets at hand.

Phase Most motor BCI systems don't exploit the phase information of the neural signals, which can for example be extracted using Fourier Analysis [Krusienski et al., 2012], Hilbert transforms [Hamner et al., 2011] and complex Wavelet transforms [Le Van Quyen et al., 2001] [Bruns, 2004]. Two classes of phase-based features have nevertheless been explored in offline and online studies.

The first class of features is formed by the phase information associated with each channel [Hammer et al., 2013], e.g. the instantaneous phase value of each channel. While the use of such features has seldom been reported, they have been shown to contain useful information for 1D kinematic offline recontruction from ECoG signals [Hammer et al., 2013].

A second approach consists in considering the phase difference between signals rather than the absolute phase of each signal. This phase difference is used to

characterize the coupling between two channels [Bruns, 2004]. The extraction of the Phase-Locking-Value (PLV), which measures the degreee of phase locking between channel k and channel l, has been reported for the classification of mental tasks in several BCI studies, e.g. [Wang et al., 2006] [Wei et al., 2007] [Pourbakhtiar et al., 2013] [Gysels and Celka, 2004] [Loboda et al., 2014]. The PLV of channels k and l is defined as the average of the instantaneous phase difference in a T-long temporal windows, i.e. $PLV(f) = \left|\frac{1}{T}\sum_{t=1}^{T} e^{j\Delta\phi_{kl}(t)}\right|$, where f is the frequency of interest and $\Delta\phi_{kl}(t)$ is the instantaneous phase difference between the two channels. The relevance of alternative phase-based features, e.g. the instantaneous or mean phase difference between two channels, has also been investigated [Hamner et al., 2011].

Although phase features permitted to control a 3-class virtual effector in [Brunner et al., 2006], they have mainly been utilized in offline studies. Coherence-, Fourier-based and phase features were compared in [Krusienski et al., 2012] for motor imagery offline classification from EEG signals (acquired during 1D cursor control). Phase and coherence features did not lead to an improved classification accuracy when compared to Fourier features. By contrast, phase features have been shown to outperform magnitude features for 1D kinematic offline recontruction from ECoG signals [Hammer et al., 2013]. The relevance of phase-related features thus remains unclear.

2.3.1.4 Time-frequency-space features

The integration of spatial information into time-frequency features has been reported for both EEG- [Onose et al., 2012] [Vidaurre et al., 2016] and ECoG-based [Marathe and Taylor, 2013] [Kapeller et al., 2015] motor BCI transducers, for example via the application of Common Spatial Pattern filters to neural signals filtered in specific frequency bands [Wu et al., 2008].

2.3.1.5 Other features

The use of alternative features like fractal dimension, Horjth parameters or temporal sequence modelling has been proposed in EEG-based studies [Boostani and Moradi, 2004] [Boostani et al., 2007] [Vidaurre et al., 2009] [Coyle et al., 2005].

Amplitude coupling between a pair of channels has also been used in BCI studies [Wei et al., 2006] [Wei et al., 2007] [Krusienski et al., 2012]. The Magnitude-Squared Coherence features between channels k and l used in [Krusienski et al., 2012] were defined as $\frac{|P_{kl}(f)|^2}{P_{kk}(f)P_{ll}(f)}$, where $P_{kk}(f)$ is the power spectral density and $P_{kl}(f)$ is the cross power spectral density between channels k and l. When compared to classic band-power features for the task of motor imagery classification, they did not permit to improve classification accuracy.

2.3.2 Features for effector control

Both discrete and continuous dependent variables can be extracted from neural signals to control prosthesis and orthosis movements.

Continuous dependent variables Continuous variables traditionally consist of position and/or velocity of the effector's endpoint, e.g. the wrist kinematic parameters in the case of an upper-limb orthosis [Li, 2014], or of the angular characteristics of effector joints [Ajiboye et al., 2012]. Wrist speed and acceleration [Hammer et al., 2013] [Hammer et al., 2016], force profil [Carmena et al., 2003] [Chen et al., 2014a] and muscular activity [Carmena et al., 2003] [Koike et al., 2006] [Choi et al., 2009] [Shin et al., 2012] have nevertheless been reconstructed in offline preliminary studies. The principal components of the effector's position or velocity have also been estimated from neural signals in offline studies [Acharya et al., 2010] [Wong et al., 2013] [Hotson et al., 2014].

Discrete dependent variables Discrete variables in particular include the direction of the effector's movement [Hortal et al., 2015] [Bhattacharyya et al., 2015], the finger of interest [Hotson et al., 2016] or the open/closed state in the case of hand prostheses or orthoses [Pfurtscheller et al., 2000]. Binary dependent variables are also regularly used to characterize the state of the user during asynchronous decoding, i.e. an Intentional Control (IC) or Non-Control (NC) state [Mason and Birch, 2000] [Müller-Putz et al., 2010]. Binary decoding has for example been considered for EEG-based 1D virtual navigation [Leeb et al., 2007] and for the restoration of walking patients with tetraplegia by means of Functional Electrical Stimulation [King et al., 2015].

It should finally be noticed that the extraction of such discrete or continuous variables from neural signals can be complemented by a re-mapping step performed by the user [Marathe and Taylor, 2011]. It has been shown that users are able to learn how to control the velocity of an effector even when position parameters are extracted from their neural signals, i.e. that position decoding is compatible with velocity control [Marathe and Taylor, 2011]. While velocity control is generally easier than position control with an equivalent degree of accuracy [Marathe and Taylor, 2011], satisfying extraction of velocity parameters from neural feature is not systematically possible (see, e.g., [Marathe and Taylor, 2013]). A combination of position decoding, velocity control and user's position-to-velocity remapping is liable to result in an improved control performance. In [Marathe and Taylor, 2013] for example, position parameters were extracted from monkeys' neural signals, and they were able to learn a position-to-velocity map to control the velocity of 2D cursor [Marathe and Taylor, 2013].

2.3.3 Dimensionality reduction

High dimensional and/or correlated features are liable to disrupt decoders training. Reduction of the independent variable dimension is mainly performed via projection or feature selection methods.

2.3.3.1 Projection methods

Projection methods are utilized to reduce the dimension of the feature space by projecting the features $\mathbf{x}^t \in \mathbb{R}^m$ onto a subspace of lower dimension F, $F < m$. Principal Component Analysis and Partial-Least-Squares are examples of projection methods which are popular within the BCI community.

PCA [Bishop, 2006] is widely used for unsupervised dimension reduction of neural signals, for example MUA/SUA [Wu et al., 2003b] [Aggarwal et al., 2008] [Kao et al., 2013] [Kao et al., 2017], ECoG [Wang et al., 2009b] and EEG signals [Ke and Li, 2009] [Suk and Lee, 2010] [Argunşah and Çetin, 2010] [Bhattacharyya et al., 2011]. PCA is an orthogonal projection of the features onto a subspace which maximizes the variance of the projected features [Bishop, 2006]. This low-dimensional subspace is spanned by the first F eigenvectors of the feature covariance matrix [Bishop, 2006]. The use of PCA's variants has also been considered in BCI studies, e.g. nonlinear PCA for motor imagery classification from EEG signals [Devulapalli, 1996].

In contrast with PCA, Partial-Least-Squares (PLS) [Höskuldsson, 1988] permit to perform supervised dimensionality reduction. Data are projected onto a low dimensional subspace which maximizes the covariance between the respective projections of the independent and dependent variables. PLS-based regression has for example been performed for efficient trajectory decoding from ECoG signals [Shimoda et al., 2012] [Eliseyev and Aksenova, 2014] [Eliseyev and Aksenova, 2016] [Bundy et al., 2016] [van Gerven et al., 2012]. The use of a combined PLS-PCA projection strategy has also been reported for SUA/MUA decoding [Kim et al., 2006c].

Both PCA and PLS are parametrized by the dimension F of the reduced feature subspace. F was chosen so as to maximize decoding accuracy on a validation data set in [Argunşah and Çetin, 2010]. Similarly, a cross-validation approach was utilized in [Kim et al., 2006c] [Shimoda et al., 2012] [Eliseyev and Aksenova, 2016] [Bundy et al., 2016]. Another approach consists in choosing F so as to retain a specific percentage of variance, e.g. 95% [Aggarwal et al., 2008] or 85% [Suk and Lee, 2010].

2.3.3.2 Feature selection methods

Whereas projection methods rely on the creation a set of F new informative features, feature selection methods permit to extract a subset of F existing features from the full feature set (m features). Different feature selection methods have been utilized in BCI studies.

A naive approach consists in computing a model for each possible feature subset, and in choosing the best subset with respect to the corresponding decoding

performance on a validation data set. Probably because exhaustive search is liable to be time-consuming, its use has not been reported for feature selection in BCI studies. Several heuristic methods were, however, exploited for dimensionality reduction in BCI studies [McFarland et al., 2010] [Wang et al., 2012] [Liang and Bougrain, 2012] [Hotson et al., 2014] [Wang et al., 2015] [Kelly et al., 2012] [Fazli et al., 2011] [Fatourechi et al., 2007b] [Garrett et al., 2003] [Graimann et al., 2004].

Wrapper methods have been used to select informative features in both EEG- and ECoG-based motor BCIs [McFarland et al., 2010] [Wang et al., 2012] [Liang and Bougrain, 2012] [Hotson et al., 2014]. They consist in fitting a model on a particular feature subset, and in monitoring the corresponding decoding accuracy on an independent data set. In contrast with exhaustive search, all possible feature subsets are not considered. Heuristics approaches, e.g. stepwise regression, are utilized to iteratively approach the optimal subset. In the case of the stepwise forward selection procedure used in [Wang et al., 2012] [Liang and Bougrain, 2012] [Hotson et al., 2014], the feature subset is initially empty. The feature associated with the highest decoding accuracy on a validation data set is chosen. At each iteration, the feature which most increases the prediction accuracy is added to the growing feature subset. Another stepwise selection procedure has been utilized for EEG neural feature selection in [McFarland et al., 2010]. Forward selection was combined with backward selection, i.e. features which had become useless after the addition of a new feature were regularly removed from the feature set [McFarland et al., 2010].

Feature selection embedded in the model identification procedure has been utilized in several BCI studies. Sparse models were for example obtained by LASSO training (Least Absolute Shrinkage and Selection Operator) in [Wang et al., 2015], [Kelly et al., 2012] and [Fazli et al., 2011].

The use of so-called filter methods has also been reported in several closed-loop [Schalk et al., 2007] [LaFleur et al., 2013] and offline [Spüler et al., 2016] motor BCI studies. In contrast with wrappers methods which are computationally expensive, filter methods do not require to train more than one model. A criterion between feature and dependent variable, e.g. correlation, is considered. Features associated with the highest criterion values are selected *a priori*.

The use of Genetic Algorithms has been reported in several EEG- [Flotzinger et al., 1994] [Fatourechi et al., 2007b] [Garrett et al., 2003] [Graimann et al., 2004] [Boostani et al., 2007] and ECoG-based BCIs [Wei et al., 2006]. Genetic Algorithms are inspired from Darwin's natural selection. Different initial features subsets are considered. The most promising subsets (with respect to some criterion which assesses decoding accuracy on the validation data set) are combined using random processes so as to create a new subset generation.

Alternative feature selection approaches, e.g. Distinctive Sensitive Learning Vector Quantization [Flotzinger et al., 1994] [Scherer et al., 2008] or selection based on the Davies–Bouldin index [Wissel et al., 2013], have additionally been used in BCI studies.

2.4 Decoders for motor BCIs

Different technical solutions have been considered to design and train BCI decoders, for example for accurate kinematic decoding or asynchronous mental-task discrimination. Let $\mathbf{x}^t \in \mathbb{R}^m$ be an independent, input variable and $\mathbf{y}^t \in \mathbb{R}^n$ or $y^t \in \mathbb{Z}$ denote a dependent, output variable. Let us assume the existence of a model f, $f : \mathbb{R}^m \rightarrow \mathbb{R}^n$ or $f : \mathbb{R}^m \rightarrow \mathbb{Z}$, such that $\mathbf{y}^t \approx f(\mathbf{x}^t)$. Let $\hat{f}$ be an estimate of the unknown model f. When motor BCIs rely on the decoding of continuous variables $y^t \in \mathbb{R}^n$ [Hochberg et al., 2012] [Collinger et al., 2013] [Wodlinger et al., 2015], the corresponding $\hat{f} : \mathbb{R}^m \rightarrow \mathbb{R}^n$ is referred to as continuous decoder (a regression model for example). Such continuous decoders are typically used to build kinematic decoders, mainly in synchronous frameworks. In the case of discrete dependent variables, a discrete decoder $\hat{f} : \mathbb{R}^m \rightarrow \mathbb{Z}$ (classifier) is applied on neural features [Sing et al., 2007] [Yanagisawa et al., 2012b] [Hotson et al., 2016]. Discrete decoders are generally used for the task of mental-task-based accurate mono-limb or multi-limb control. Their utilization is also regularly considered to embed NC support into both mental-task and kinematic decoders.

2.4.1 Generalities on decoder training

Most motor BCIs rely on user-specific decoders with data-driven training. Machine learning methods are used to build a relevant decoder $\hat{f}$ to model the dependence between neural features $\mathbf{x}^t$ and user intentions $\mathbf{y}^t$.

Decoder training consists in building an estimate $\hat{f}$ of the unknown function f. First, the set of admissible models $\mathbf{G}$ is restricted by making hypotheses about the function f (e.g., linearity) and/or about the noise distribution. The estimation of the optimal value of a set Θ of parameters is generally necessary to fully characterize each admissible model $f_G \in \mathbf{G}$. The parameters of admissible models $f_G \in \mathbf{G}$ are therefore identified on "training" samples $(\mathbf{X}, \mathbf{Y})$, where $\mathbf{X} = \{\mathbf{x}^t\}_{t=1}^T$ and $\mathbf{Y} = \{\mathbf{y}^t\}_{t=1}^T$. Each sought-after model is trained so that the estimates $\hat{\mathbf{Y}} = \{\hat{\mathbf{y}}^t\}_{t=1}^T = \left\{\hat{f}_G(\mathbf{x}^t)\right\}_{t=1}^T$ are optimal with respect to a specific criterion, for example the Mean Squared Error $\left\|\hat{\mathbf{Y}} - \mathbf{Y}\right\|_2$, where $\|.\|_2$ is the ℓ^2-norm. Model selection is required when more than one admissible model is considered. Model selection procedures usually aim at choosing the decoder associated with the best generalization performance, i.e. with the most accurate estimates $\hat{\mathbf{y}}^{t_{new}} = \hat{f}_G(\mathbf{x}^{t_{new}})$ for samples $\{\mathbf{x}^{t_{new}}, \mathbf{y}^{t_{new}}\}$ not used for training. The performance of a decoder on training samples is known to be an optimistic estimate of its generalization performance (for example evaluated by means of the test error) [Friedman et al., 2001]. For this reason, admissible models are generally not directly compared on the basis of their performance on training samples. Two main approaches are used to estimate the test error associated with a model [Friedman et al., 2001]. Model selection criteria, for example Mallow's C_p or information criteria like the Bayesian Information Criterion and the Akaike Information Criterion, estimate the test error from the training error, i.e. they

attempt to correct the optimism of the training error [Friedman et al., 2001]. The model which minimizes the considered information criterion is selected. By contrast, Cross-Validation (CV) is based on a direct estimation of the test error. The training data set is typically divided into k folds. Models are trained on (k-1) folds, and tested on the remaining fold [Friedman et al., 2001]. The procedure is repeated with all possible combinations of (k-1) folds. The model with the best average test performance is selected, and re-trained on the full training data set. Whereas model selection criteria are applicable to models which are linear in their parameters [Friedman et al., 2001], CV approaches can be used for all classes of models. Their main shortcoming is that they involve repeated model identification procedures, and are therefore computationally expensive.

2.4.2 Discrete decoding - classifiers

Let us consider the case of a discrete-valued dependent variable $y^t \in \mathbb{N}$. y^t is usually referred to as "class label". The associated observation $\mathbf{x}^t$ is said to belong to the "class" identified by its label. The classification of neural patterns, e.g. the discrimination between different mental tasks such as motor imageries or NC (idle) states [LaFleur et al., 2013] [Fifer et al., 2014], has been the basis of several EEG- and ECoG-driven motor BCIs [Sing et al., 2007] [Yanagisawa et al., 2012b] [Hortal et al., 2015] [Hotson et al., 2016]. Various classifiers have been embedded into motor BCI transducers.

2.4.2.1 Generative and discriminative classifiers

Both generative and discriminative classifiers have been used in EEG- [Chae et al., 2012] [Hortal et al., 2015] and ECoG-based [Yanagisawa et al., 2012b] [Fifer et al., 2014] [Kapeller et al., 2015] motor BCIs.

Generative classifiers The use of generative classifiers has been reported in both EEG- [Sing et al., 2007] [Pfurtscheller et al., 2000] [Chae et al., 2012] [Vidaurre et al., 2016] and ECoG-based motor BCIs [Fifer et al., 2014] [Kapeller et al., 2015] [Hotson et al., 2016]. The relevance of generative classifiers has additionally been investigated in offline EEG- [Chiappa and Bengio, 2003] [Hasan and Gan, 2009] [Bhattacharyya et al., 2011], ECoG- [Wang et al., 2016] and MUA/SUA-based [Hatsopoulos et al., 2004] preliminary studies, e.g. for the classification of real movements [Hatsopoulos et al., 2004] [Wang et al., 2016] or of mentals tasks [Chiappa and Bengio, 2003] [Hasan and Gan, 2009] [Bhattacharyya et al., 2011].

Generative classifiers model the way independent variables are generated within a class i, i.e. they model the joint probability $P(\mathbf{x}^t, y^t = i)$ [Ng and Jordan, 2002]. Once the joint probability has been fitted for each class, the classification of a new observation sample $\mathbf{x}^t$ is performed by computing the posterior probability $P(y^t = i|\mathbf{x}^t)$ with respect to each class [Ng and Jordan, 2002]. Using Bayes rule, this posterior probability is proportional to $P(\mathbf{x}^t, y^t = i)$ for class i [Ng and Jordan, 2002].

The most likely class label is assigned to the considered observation sample [Ng and Jordan, 2002]. The characteristics of the decision boundary (i.e., the equiprobable hypersurface defined by $P(y^t = i|\mathbf{x}^t) = P(y^t = j|\mathbf{x}^t)$) are not explicitly chosen, but result from the distribution used to model data generation within each class. The majority of generative classifiers reported in motor BCIs or preliminary studies relied on multivariate Gaussian distributions (e.g., [Bhattacharyya et al., 2011] [Lemm et al., 2004] [Wang et al., 2016]) or Gaussian Mixtures Models [Chiappa and Bengio, 2003] [Hasan and Gan, 2009], i.e. $P(\mathbf{x}^t|y^t = i) = \mathcal{N}(\mu_i, \mathbf{\Sigma}_i)$ or $P(\mathbf{x}^t|y^t = i) = \sum_k \mathcal{N}(\mu_{k,i}, \mathbf{\Sigma}_{k,i})$. By contrast, the MUA/SUA firing rates of two NHPs were modelled by Poisson distributions in [Hatsopoulos et al., 2004].

High-dimensional neural features are frequently considered in motor BCIs [Kim et al., 2011] [Bhattacharyya et al., 2015] (up to 285 and 630 features, respectively). In high dimension, fitting a multivariate distribution is impractical [Fan et al., 2011]. Several teams have therefore investigated the application of naive Bayes classifiers, for left and right hand movement classification from more than 800 EEG features in [Bhattacharyya et al., 2011], and to classify 8 wrist directions on the basis of the firing rate of 32 to 143 neurons in [Hatsopoulos et al., 2004]. Naive Bayes classifiers assume that features are independent conditionally to the class. The modelling of the corresponding conditional probability is simplified, and classifier training is consequently facilitated. Its potential advantages over traditional Bayes classifiers have particularly been illustrated in [Bhattacharyya et al., 2011], where a naive Bayes classifier surpassed a Gaussian-based generative classifier for the decoding of EEG signals, both when a high-dimensional EEG feature set (871 features) was used and when its dimensionality had been preliminary reduced (91 features).

Finally, the training of generative classifiers can be performed on partially labelled training data sets [Sutton and Mccallum, 2012], e.g. for EEG features unsupervised classification [Hasan and Gan, 2011].

Discriminative classifiers Discriminative classifiers have been utilized in both EEG- and ECoG-driven motor BCIs, e.g. for EEG-based displacement of a robotic arm's endpoint [Hortal et al., 2015] or for ECoG-based control of a prosthetic arm [Yanagisawa et al., 2012b]. Their performance has also been assessed in offline preliminary studies, i.e. for the classification of motor imageries [Schlögl et al., 2005] [Hill et al., 2006] [Chin et al., 2007].

Discriminative classifiers directly model the posterior class probability $P(y^t = i|\mathbf{x}^t)$ [Bishop, 2006]. In contrast with generative classifiers, the intermediary probability $P(\mathbf{x}^t|y^t = i)$ is not modelled [Ng and Jordan, 2002]. This makes the use of discriminative classifiers advantageous when this distribution cannot be approximated with classical distributions. Discriminative classifiers are particularly relevant when $\mathbf{x}^t$ is high-dimensional or includes redundant (correlated) neural features [Sutton and Mccallum, 2012], because non-discriminant features are not considered during model training. Support-Vector-Machine (SVM) classifiers [Lal et al., 2005] [Hill et al., 2006] [Yanagisawa et al., 2012a][Schlögl et al., 2005] [Sadeghian and

Moradi, 2007] [Bhattacharyya et al., 2011] [Hortal et al., 2015], Logistic Regression (LR) [Tomioka et al., 2007] [Bashashati et al., 2015] [Chen et al., 2014b] [Bundy et al., 2016], k-Nearest Neighbors (kNN) [Chin et al., 2007] [Kayikcioglu and Aydemir, 2010] and Artificial Neural Networks (ANN) [Haselsteiner and Pfurtscheller, 2000] [Navarro et al., 2005] [Nakayama and Inagaki, 2006] [Hatsopoulos et al., 2004] are some of the discriminative classifiers which use has been reported in motor BCIs and/or considered in preliminary studies.

The superiority of generative or discriminative classifiers for mental states classification cannot be established *a priori*. Their relevance and performance depend on the problem at hand, in particular on the validity of the assumed distributions in the case of generative classifiers, on the dimension of the neural feature representation, on the availability of labelled and/or unlabelled training samples etc. The reported popularity of decoders within the BCI community is consistent with this observation, as Linear Discriminant Analysis (LDA) and SVM, which are both widely used in BCI studies [Bashashati et al., 2015] [Nicolas-Alonso and Gomez-Gil, 2012], are a generative and discriminative classifier, respectively.

Generative and discriminative classifiers can be associated with linear or non-linear decision boundaries.

2.4.2.2 Linear and non-linear classifiers

The comparative relevance of linear and non-linear classifiers has been frequently investigated in preliminary studies focusing on the recognition of mental tasks from neural signals (e.g., [Müller et al., 2003] [Garrett et al., 2003]).

Linear classifiers Different linear classifiers have been applied and tested for online and offline mental states recognition, for example multiple Intentional Control and No-Control states.

Linear Discriminant Analysis (LDA) classifiers are generative classifiers based on multivariate Gaussian distributions. Their linear decision boundary [Bishop, 2006] is obtained by using the same covariance matrix for all classes. LDA classifiers have been embedded in several motor BCIs, e.g. in BCIs providing users with control over hand prostheses or orthoses [Pfurtscheller et al., 2000] [Fifer et al., 2014] [Hotson et al., 2016], lower-limb orthoses [Vidaurre et al., 2016] or humanoid robots [Kapeller et al., 2015]. LDA has also been used for offline motor imagery classification in EEG [Bhattacharyya et al., 2011], and for cognitive state estimation in LFP [Aggarwal et al., 2013] and MUA/SUA [Velliste et al., 2014] signals.

Another linear classifier, namely the Support-Vector-Machine (SVM) classifier, has frequently been applied in motor BCI studies [Schlögl et al., 2005] [Yanagisawa et al., 2012b] [Hortal et al., 2015] The SVM's linear decision boundary is chosen so as to maximize its margin with the nearest training samples ("support-vectors") [Bishop, 2006]. SVM-based classification has for example enabled users to control a prosthetic hand [Yanagisawa et al., 2012b] and a robotic arm [Hortal et al., 2015].

SVMs are also regularly used for offline motor imagery classification in ECoG [Lal et al., 2005] [Hill et al., 2006] [Demirer et al., 2009] [Yanagisawa et al., 2012a] and EEG [Schlögl et al., 2005] [Sadeghian and Moradi, 2007] [Bhattacharyya et al., 2011] signals. SVMs are attractive for neural signal decoding [Lotte et al., 2007] because of their good generalization abilities [Schlögl et al., 2005] and of their robustness in high-dimensional settings [Friedman et al., 2001].

The application of a threshold on the output of a linear regression model has been reported in motor BCI studies, e.g. for ECoG-driven asynchronous 2D cursor control [Williams et al., 2013]. This classification approach has also been considered in offline asynchronous studies [Eliseyev et al., 2011] [Eliseyev et al., 2012].

Logistic Regression (LR) is a discriminant classifier based on Generalized Linear Models, which extend linear models in that a non-linear link function g is applied on a linear combination of features [Bishop, 2006]. In contrast with linear regression-based classifiers, LR considers a discrete dependent variable and assumes that $P(y^t|\mathbf{x}^t)$ follows a Bernoulli distribution. Although LR has rarely been used in closed-loop studies [Penny et al., 2000], several teams have investigated its relevance for the distinction betwen mental tasks from EEG [Tomioka et al., 2007] [Gouy-Pailler et al., 2009] [Bashashati et al., 2015] and ECoG signals [Chen et al., 2014b] [Bundy et al., 2016].

The previously mentioned classifiers are based on Euclidian distances, i.e. on the ℓ^2-norm. Classifiers based on the Mahalanobis distance have been applied in several offline BCI studies, e.g. in [Cincotti et al., 2003] [Sadeghian and Moradi, 2007] [Bai et al., 2011] [Schlögl et al., 2005]. Efficient classification of EEG patterns has additionally been achieved by exploiting Riemannian distance [Barachant et al., 2010] [Barachant et al., 2012].

LDA has been regularly used to provide users with neural control over prostheses, orthoses and robotic devices, and it is particularly popular for EEG offline linear classification [Bashashati et al., 2015]. No clear superiority of LDA decoding performance has, however, been reported in offline comparative studies [Schlögl et al., 2005] [Wang et al., 2009b] [Bashashati et al., 2015]. In [Bashashati et al., 2015], a LDA classifier was slightly but not significantly surpassed by a LR-based classifier for asynchronous and synchronous EEG decoding . In [Wang et al., 2009a], LDA and SVM performed similarly for both motor imagery and finger movement classification from EEG signals. In another comparative study [Schlögl et al., 2005], LDA was significantly outperformed by a SVM for 4-class motor imagery classification in EEG signals. By contrast, it performed better than a SVM when applied on low-dimensional EEG features in [Bhattacharyya et al., 2011]. Its comparatively low robustness in high dimensions was also illustrated in the same study [Bhattacharyya et al., 2011], as LDA performance diminished when the dimension of the EEG features had not been reduced beforehand [Bhattacharyya et al., 2011].

Non-linear classifiers Non-linear classifiers have mainly be applied in offline preliminary studies [Schlögl et al., 2005] [Navarro et al., 2005] [Nakayama and

Inagaki, 2006] [Bhattacharyya et al., 2011], for example to discriminate between mental tasks associated with different Intentional Control states [Schlögl et al., 2005] [Bhattacharyya et al., 2011] or between No-Control and Intentional Control states [Mason and Birch, 2000].

Several teams have reported EEG mental task classification by means of Quadratic Discriminant Analysis (QDA) classifiers, i.e. Gaussian-based generative classifiers with class-specific covariance matrices [Schlögl et al., 2005] [Bhattacharyya et al., 2011].

The use of non-linear SVM has been investigated for EEG feature classification in [Bhattacharyya et al., 2011] and [Bashashati et al., 2015]. Non-linear SVMs were designed by means of non-linear kernels, typically Radial Basis Functions (RBF), in [Bhattacharyya et al., 2011] [Bashashati et al., 2015].

Artificial Neural Networks (ANNs) have been used for the offline, non-linear classification of mentals states [Haselsteiner and Pfurtscheller, 2000] [Mahmoudi and Erfanian, 2002] [Navarro et al., 2005] [Nakayama and Inagaki, 2006] or real movements [Hatsopoulos et al., 2004]. ANNs attempt to mimic information encoding in biological neuron networks [Bishop, 2006] by applying cascaded non-linear functions on weighted combinations of features x_i^t, resulting in a highly non-linear model [Bishop, 2006]. The flexibility of ANNs makes them attractive for the complex problem of neural signal modelling. It has nevertheless been reported that they can suffer from a few shortcomings, namely difficulties to select the optimal network architecture, to avoid overfitting [Kayikcioglu and Aydemir, 2010], and to interpret results. As a result, it has been observed in comparative studies that the accuracy of ANN-based mental task classification is not systematically better than the one obtained with simple non-linear models [Wang et al., 2009a] [Garrett et al., 2003]. In [Kayikcioglu and Aydemir, 2010], an ANN was outperformed by a non-linear SVM for different training data set sizes. In [Garrett et al., 2003], where non-linear SVMs and ANNs were compared for a 5-class discrimination task with EEG signals, the ANN was bettered by the SVM. Similar results were obtained on two EEG data sets in [Wang et al., 2009a].

Finally, the use of the k-Nearest Neighbors (kNN) classifier has been investigated for offline detection of mental tasks from EEG features [Mason and Birch, 2000] [Schlögl et al., 2005] [Wang et al., 2009b] [Kayikcioglu and Aydemir, 2010] [Bhattacharyya et al., 2011], and has been applied for real movement classification from ECoG data in [Chin et al., 2007]. In contrast with previously reported classifiers, the kNN classifier is not parametric. A new sample is assigned with the label which is the most represented among its k nearest training samples [Bishop, 2006]. Thus, kNNs don't require a time-consuming training procedure to be completed before their application. A high computational load can, however, be associated with kNN application, which is based on the computation of the distance between a new sample and the training samples. The computation of these distances is computationally expensive when large training data sets are necessary to properly model the data structure. This shortcoming may limit its applicability for motor BCIs relying on high-dimensional features, as online kNN-based classification may introduce a large

delay in the system. Mixed results have been reported in the comparative studies completed on non-linear classifiers [Wang et al., 2009a] [Kayikcioglu and Aydemir, 2010]. In [Wang et al., 2009a], a kNN performed similarly to a SVM with a RBF kernel for the discrimination between EEG motor imageries, and was only slightly surpassed by the same SVM-based classifier for finger movement decoding from EEG signals. This satisfying performance was obtained with low-dimensional input features (respectively of 2 and 14). In [Kayikcioglu and Aydemir, 2010], a similar comparison was drawn between a kNN, a RBF-based SVM and an ANN for 2-class classification in the context of EEG-based up-down neural control of a cursor. The kNN outperformed both the MLP and SVM for this specific classification task, and its performance was best maintained when the researchers attempted to reduce the training data set size. The comparison was nevertheless performed in a setting particularly favourable to the kNN, as the input features were only of dimension two. By contrast, in [Bhattacharyya et al., 2011], the kNN was outperformed by a RBF-based SVM for two sizes of independent variable (namely, 871 and 91 features).

The respective advantages and limits of the above mentioned classifiers are unclear. First, most of them have not been used for online pattern classification. Additionally, offline comparisons have generally been completed for two or three classifiers only, and the statistical significance of the results has seldom been established. A few studies have nevertheless endeavoured to assess the relative interest of linear and non-linear classifiers for offline discrimination between neural patterns [Wang et al., 2009a] [Bhattacharyya et al., 2011] [Bashashati et al., 2015].

Linear models exhibit a lesser modelling ability. In [Bhattacharyya et al., 2011], a RBF-based SVM was found to outperform a linear SVM as well as the other linear classifiers implemented in this study. The same result were obtained in [Wang et al., 2009a], where a linear SVM was bettered by a RBF-based SVM for both motor imageries and finger movement classification.

The superiority of non-linear classifiers has not systematically been reported in BCI preliminary studies. For example, QDA did not outperform LDA in two comparative studies [Wang et al., 2009a] [Bhattacharyya et al., 2011]. In [Garrett et al., 2003], a LDA classifier was compared to non-linear SVMs and to an ANN for the classification of 5 mental tasks. The performance of the non-linear classifiers was found to be only slightly better than the LDA's one for this EEG classification task. Training pitfalls were illustrated in [Schlögl et al., 2005], where a kNN was significantly outperformed by a linear SVM and by LDA for 4-class motor imagery discrimination from EEG signals. For this reason, it has been advocated in [Müller et al., 2003] to use linear methods except for some specific cases with "complex, large" data sets. Correspondingly, linear classifiers like LDA [Bashashati et al., 2015] are regularly chosen over non-linear models despite their lesser modelling ability. In particular, most recent clinical motor BCIs have relied on linear classifiers, e.g. LDA [Sing et al., 2007] [Fifer et al., 2014] [Kapeller et al., 2015] [Hotson et al., 2016] [Vidaurre et al., 2016] or SVM [Hortal et al., 2015].

Despite some trends, the relevance of a linear or non-linear classifier ultimately

depends on the problem at hand. In [Bashashati et al., 2015] for example, the two top classifiers for self-paced data decoding were a linear and a non-linear classifier, namely a LR classifier and an ANN. Classifier performance is in particular related to the characteristics of the extracted neural features, for example to their type [Bashashati et al., 2015] or dimension [Bhattacharyya et al., 2011]. In [Bashashati et al., 2015], classifiers' performance for synchronous data decoding was not similar when classifiers were fed with band-pass- or with wavelet-based features. In [Bhattacharyya et al., 2011], differences in performance ranking were observed if classifiers were applied on a high-dimensional input variable or on the same variable after PCA-based dimensionality reduction. Similarly, a RBF SVM and a kNN were identified as the best classifiers for a task of motor imagery decoding in [Wang et al., 2009a], but LDA reportedly equalled a RBF-based SVM for finger decoding in the same paper.

Finally, to the best of our knowledge, only limited comparative studies have been completed on ECoG data [Shenoy et al., 2008]. The respective relevance of the above-mentioned classifiers thus remains to be ascertained for ECoG data.

2.4.2.3 Static and sequential classifiers

The previously mentioned classifiers are static, i.e. they don't take into account possible dependencies between successive independent or dependent variables. Formally, they assume that pairs $(\mathbf{x}^t, y^t)$ are temporally independent and identically drawn from the distribution $P(\mathbf{x}^t, y^t)$ [Dietterich, 2009].

This assumption is typically violated in motor BCI studies. A few teams have therefore investigated the interest of taking into account the sequential nature of the independent or dependent variable [Obermaier et al., 2001] [Chiappa and Bengio, 2003] [Argunşah and Çetin, 2010]. One strategy regularly utilized in BCI studies, e.g. [Kim et al., 2011] [Flamary and Rakotomamonjy, 2012] [Eliseyev et al., 2012], consists in extracting features from several time segments to build a temporal sequence of feature vectors. This sequence is then fed to a static classifier [Lotte et al., 2007] [Dietterich, 2009]. Another approach, namely the application of dynamic classifiers, has been reported for neural pattern classification in SUA/MUA [Darmanjian et al., 2003], EEG [Obermaier et al., 2001] [Argunşah and Çetin, 2010] and ECoG-based [Onaran et al., 2011] [Saa et al., 2016] BCI studies. Dynamic classifiers directly exploit time series temporal behaviour [Lotte et al., 2007].

Dynamical classification of neural signals has been performed by means of Hidden Markov Models (HMMs) in EEG [Obermaier et al., 2001] [Gouy-Pailler et al., 2009] [Argunşah and Çetin, 2010], ECoG [Onaran et al., 2011] and SUA/MUA preliminary studies [Darmanjian et al., 2003] [Wissel et al., 2013], and in a few motor BCIs [Fifer et al., 2014] [Hotson et al., 2016].

HMMs consider a hidden state $z^t \in \mathbb{Z}$ which is generated by a first order Markov process, i.e. such as $P(z^{t+1} = k|z^{1:t}) = P(z^{t+1} = k|z^t)$ [Rabiner, 1989]. The value of the observation $\mathbf{x}^t \in \mathbb{Z}^m$ or $\mathbf{x}^t \in \mathbb{R}^m$ depends on the corresponding hidden state value z^t via the conditional probability $P(\mathbf{x}^t|z^t)$ [Rabiner, 1989]. Efficient recursive

algorithms permit to infer the most likely state label $\hat{z}^t$ by combining prior knowledge about the previous hidden state z^{t-1} with the likelihood of the current observed features $\mathbf{x}^t$ [Rabiner, 1989].

One reported approach to HMM-based classification consists in associating one hidden state value $z^t = i$ to each class label $y^t = i$ [Kemere et al., 2008] [Fifer et al., 2014] [Hotson et al., 2016]. The Markovian hypothesis thus models the class label succession. This strategy has for example been used for offline target estimation from SUA/MUA signals in [Kemere et al., 2008]. HMM-based classifiers have also been applied for robust online state detection in several closed-loop motor BCIs [Fifer et al., 2014] [Hotson et al., 2016] [Kao et al., 2017]. States were for example associated with NC and IC classes [Fifer et al., 2014] [Hotson et al., 2016].

An alternative approach has been investigated in offline preliminary studies [Obermaier et al., 2001] [Darmanjian et al., 2003] [Argunşah and Çetin, 2010] [Onaran et al., 2011] [Wissel et al., 2013]. One HMM was associated with each considered class, and several states were thus used to model feature dynamic within each class. Classification was performed by feeding each HMM with a sequence of N consecutive observations, and by computing the associated probability $P(\mathbf{x}^{t-N+1:t}|y^t = i)$. The sequence was assigned the class i which maximized $P(\mathbf{x}^{t-N+1:t}|y^t = i)$. HMMs have been used for offline modelling of the variations of neural features within NC and IC states in SUA/MUA [Darmanjian et al., 2003] and ECoG [Onaran et al., 2011], within finger movements in ECoG signals [Wissel et al., 2013], or within motor imageries in EEG signals [Obermaier et al., 2001] [Argunşah and Çetin, 2010].

The use of several HMM's variants has been proposed for the classification of EEG and ECoG mental tasks [Chiappa and Bengio, 2003] [Hasan and Gan, 2010] [Hasan and Gan, 2011] [Delgado Saa and Cetin, 2012] [Saa and Çetin, 2013].

Input-Output Hidden Markov Models (IOHMM) were applied on EEG signals to discriminate between 3 mental tasks in [Chiappa and Bengio, 2003]. In contrast with HMMs, IOHMMs are trained to distinguish between classes composed of several hidden states, and directly map input features to the non-stationary classes [Bengio and Frasconi, 1996].

Conditional Random Fields (CRFs) are discriminative undirected graphical models [Sutton and Mccallum, 2012], and linear-chain CRFs are more particularly the discriminative counterpart of HMMs [Sutton and Mccallum, 2012]. CRFs have been used for EEG offline modelling and decoding [Hasan and Gan, 2010] [Hasan and Gan, 2011] [Delgado Saa and Cetin, 2012] [Saa and Çetin, 2013], and for finger movement detection in ECoG signals [Saa et al., 2016]. If they have a better ability to model long-term time dependencies [Lafferty et al., 2001], their training is computationally expensive [Dietterich, 2009].

Dynamic Bayesian Models (DBN) are probabilistic graphical models which permit to take into account the dependence between several random variables [Murphy, 2002]. HMMs are a specific case of DBNs, and are therefore less flexible than DBNs. The dynamic of EEG [Shenoy, 2005] and ECoG [Wang et al., 2012] signals has been exploited by means of DBN [Murphy, 2002] in offline studies.

Finally, the use of a time-dependent ANN was reported for EEG dynamical classification in [Haselsteiner and Pfurtscheller, 2000].

Most dynamic classifiers which were embedded in motor BCIs were generic HMMs with state-class correspondence [Fifer et al., 2014] [Hotson et al., 2016] [Kao et al., 2017]. While the respective performances of more complex dynamical classifiers has been investigated and compared in offline studies [Chiappa and Bengio, 2003] [Delgado Saa and Cetin, 2012] [Saa and Çetin, 2013], they have not been assessed in closed-loop settings. In [Delgado Saa and Cetin, 2012], HMM surpassed CRFs for the classification of EEG signals, but were outperformed by a CRF variant, namely a hierarchical CRF. On the contrary, in [Saa and Çetin, 2013], HMM-based EEG classification accuracy was inferior to the CRF-based one. In [Chiappa and Bengio, 2003], IOHMMs were found to outperform HMMs for EEG dynamic classification. As both dynamic classifiers performed similarly to their static counterparts (namely, a Gaussian Mixture Model-based Bayes classifier and an ANN), the authors concluded on the superiority of ANN over Gaussian Mixture Model-based generative classification for the considered EEG data set. It has correspondingly been suggested that multi-state dynamic classifiers could be suboptimal for asynchronous decoding [Lotte et al., 2007].

A few additional papers have investigated the respective relevance of static and dynamic classifiers when applied to offline data, e.g. [Cincotti et al., 2003] [Saa et al., 2016]. In [Cincotti et al., 2003], HMMs were significantly outperformed by ANNs (and Mahalanobis Distance) for the classification of right and left hand motor imageries in EEG signals. In [Saa et al., 2016], an extension of CRFs improved the discrimination between finger movements from ECoG signals when compared to LR and sparse linear regression. The interpretability of this result is nevertheless limited, because the considered static and dynamic classifiers did not belong to the same class of models, in contrast with the static-dynamic pairs compared in [Chiappa and Bengio, 2003].

2.4.3 Continuous decoding

Let the dependent variable $\mathbf{y}^t$ be continuously-valued, $\mathbf{y}^t \in \mathbb{R}^n$. Decoding of continuous dependent variables is mainly performed within the framework of MEA- and ECoG-based motor BCI systems. Continuous dependent variables typically characterize the position or velocity of the effector's endpoint, e.g. the wrist kinetics or kinematics in the case of an upper-limb orthosis [Li, 2014] (see **section 2.3.2**). The use of different classes of models $\hat{f} : \mathbb{R}^m \to \mathbb{R}^n$ has been explored in BCI studies.

2.4.3.1 Linear and non-linear regression models

Both linear and non-linear regression models have been applied for kinematic parameter reconstruction from neural signals.

Linear regression models Neural control over prostheses or cursors has been achieved by means of linear models in several motor BCIs, both with human [Hochberg et al., 2006] [Collinger et al., 2013] [Wodlinger et al., 2015] and primate subjects [Taylor et al., 2002] [Carmena et al., 2003] [Velliste et al., 2008] [Suminski et al., 2010] [Williams et al., 2013] [Willett et al., 2013]. Offline trajectory reconstruction has also been performed by means of linear models in several EEG-, ECoG- and MUA/SUA-based preliminary studies [Bradberry et al., 2010] [Liang and Bougrain, 2012] [Eliseyev and Aksenova, 2014] [Bundy et al., 2016] [Koyama et al., 2010b]. Linear models rely on the assumption that the dependent variable is a (noisy) linear combination of the independent variable components, i.e. of the neural features:

$$\mathbf{y}^t = \mathbf{B}\mathbf{x}^t + \epsilon^t$$

where $\mathbf{B} \in \mathbb{R}^{n \times m}$ and $\epsilon^t \in \mathbb{R}^n$ is the observation noise, and where the neural features $\mathbf{x}^t$ can embed a history of instantaneous neural features $\tilde{\mathbf{x}}^t$, i.e. $\mathbf{x}^t = \tilde{\mathbf{x}}^{(t+1-\tau_2):t}$. Neural feature temporal concatenation is one of the approaches which permit to exploit neural signal temporal characteristics [Dietterich, 2009] [Lotte et al., 2007].

A particular linear model, namely the Population Vector Algorithm (PVA), has more specifically been used for kinematic decoding in several MEA-driven motor BCI systems [Taylor et al., 2002] [Velliste et al., 2008] [Collinger et al., 2013] [Wodlinger et al., 2015]. The PVA is based on the cosine directional tuning model [Georgopoulos et al., 1986], which states that neurons of the motor cortex fire preferentially in one specific direction. The instantaneous firing rate of each neuron is used to weight the corresponding preferred direction.

The use of different identification algorithms has been reported in motor BCIs and offline reconstruction studies. Linear regression models have frequently been trained using Ordinary Least Squares (OLS) in BCI studies [Li et al., 2009]. While OLS corresponds to the Maximum Likelihood estimator when the measurement noise is Gaussian, the OLS estimator is unstable when the input variable $\mathbf{x}^t$ is high dimensional or composed of correlated explanatory features [Friedman et al., 2001]. The use of penalized approaches such as pace regression [Kubánek et al., 2009], ridge regression [Willett et al., 2013] [Li et al., 2009] [Suminski et al., 2010] [Shanechi et al., 2013] and sparse linear regression [Williams et al., 2013] has therefore been proposed for model identification. Performant decoding from ECoG high dimensional feature representations has additionally been reported using Partial Least Squares and variants [Shimoda et al., 2012] [Eliseyev and Aksenova, 2014] [Eliseyev and Aksenova, 2016] [Bundy et al., 2016] [van Gerven et al., 2012].

Non-linear regression models Linear regression models rely on simplistic assumptions about information encoding in motor neural signals, which complexity has been suggested in numerous studies [Scott, 2008]. Several teams have in particular investigated the use of non-linear models for neural signal decoding [Li, 2014], assuming that $\mathbf{y}^t = f(\mathbf{x}^t) + \varepsilon$ with f non-linear. These studies mainly consisted of

offline trajectory reconstructions [Kim et al., 2006b] [Eliseyev and Aksenova, 2014] [Spüler et al., 2016].

Generalized Linear Models (GLMs) constitute one class of non-linear models which interest for offline trajectory reconstruction has been explored in BCI studies, e.g. from ECoG signals in [Eliseyev and Aksenova, 2014]. GLMs extend linear models by applying a non-linear function g^{-1} on the output of a linear filter β, i.e. $\mathbf{y}^t = g^{-1}(\beta \mathbf{x}^t)$. The non-linear function g^{-1} , referred to as "link" function, is chosen among standard functions (e.g., the logarithm) or fitted on training data [Eliseyev and Aksenova, 2014]. A similar approach, namely a cascaded Wiener filter, has been applied on SUA/MUA offline data sets in [Flint et al., 2012] and [Scheid et al., 2013]. Generalized Additive Models (GAM) were alternatively utilized for trajectory reconstruction from primate ECoG signals in [Eliseyev and Aksenova, 2014]. Generalized Additive Modelling consists in applying a non-linear function g_i on each component $\mathbf{x}_i^t$ of the independent variable. A linear model β subsequently combines the outputs of the non-linear functions $g_i(\mathbf{x}_i^t)$. The application of non-linear regression models such as Support Vector Machine Regression (SVR) [Kim et al., 2006b] [Mehring et al., 2003] or ANN models [Sanchez et al., 2002] [Kim et al., 2006c] [Kim et al., 2006b] [Hatsopoulos et al., 2004] has been additionally proposed for SUA/MUA decoding, and tested in offline preliminary studies. SVR-based offline trajectory reconstruction has also been reported in an ECoG-driven preliminary BCI study [Spüler et al., 2016].

The findings of several offline preliminary studies are consistent with the idea that non-linear regression models are likely to be more realistic than linear ones for kinematic decoding: linear decoders were outperformed by both GLM and GAM approaches for ECoG signal decoding in monkeys [Eliseyev and Aksenova, 2014] signals, and by SVR in simulated primate SUA/MUA signals [Kim et al., 2006b]. Because non-linear models are more flexible than linear ones, and therefore more prone to overfit, fine identification procedures were often required for non-linear models proper training. The training of the GAM proposed in [Eliseyev and Aksenova, 2014] relied on algorithms from the PLS family. Difficulties were reported for the training of the ANN used in [Kim et al., 2006b] for trajectory decoding. They were presented as a possible cause for the superior decoding performance of the SVR, which is yet less flexible than ANNs [Kim et al., 2006b]. A specific early-stopping procedure was utilized to prevent overfit during ANN training in [Hatsopoulos et al., 2004]. The complexity of ANNs' possible structures (e.g., number of layers and number of neurons per layer) additionally makes their optimization time-consuming, which is the reason why proper optimization of the ANN structure was not performed in [Hatsopoulos et al., 2004]. Under these conditions, linear- and ANN-based trajectory reconstructions from MUA/SUA signals yielded similar results in this study [Hatsopoulos et al., 2004].

Thus, linear models are sometimes chosen over their non-linear counterparts in spite of their simplistic assumptions, in particular in the case of MUA/SUA- [Taylor et al., 2002] [Velliste et al., 2008] [Collinger et al., 2013] [Wodlinger et al., 2015] and ECoG-driven [Schalk et al., 2008] [Wang et al., 2013c] motor clinical

BCIs. Up to 10D- and 3D-control has been achieved by means of linear filtering of MUA/SUA and ECoG signals, respectively [Wodlinger et al., 2015] [Wang et al., 2013c]. Linear models have additionally been shown to be reasonably efficient for position, velocity, acceleration, speed etc. offline decoding [Wang et al., 2007] [Bundy et al., 2016] [Hammer et al., 2016], and generally involve simpler training procedures than non-linear models.

Since the last decade, another class of decoders, namely dynamic models, has gained popularity in motor BCIs.

2.4.3.2 Dynamic models

Stochastic state-space models Linear or GAM-, GLM-, SVR- and ANN-based decoders are static regression models, i.e. they assume the existence of a (parametric or non-parametric) linear or non-linear model f so that $\mathbf{y}^t \approx f(\mathbf{x}^t)$. By contrast, the dynamic models utilized for cursor or prosthesis control in several motor BCIs [Hochberg et al., 2012] [Ifft et al., 2013] consider stochastic state-space models, i.e.:

$$\mathbf{y}^{t+1} = g(\mathbf{y}^t) + \mathbf{w}^t \ , \tag{2.1}$$

$$\mathbf{x}^t = h(\mathbf{y}^t) + \mathbf{v}^t \ . \tag{2.2}$$

The noise processes $\mathbf{w}^t$ and $\mathbf{v}^t$ are generally independent and identically distributed sequences of random variables [Krishnamurthy, 2016]. The continuous response variable $\mathbf{y}^t \in \mathbb{R}^n$ is here composed by the trajectory coordinates and derivatives (velocity, acceleration etc.). The transition equation (2.1) explicitly describes the dynamic of the hidden sequence $\mathbf{y}^t \in \mathbb{R}^n$ ("movement model" [Li, 2014]). As expressed in (2.1), movement models traditionally rely on first-order Markovian temporal dependencies. Free or typical upper-limb movements like reaching movements cannot be precisely characterized by a first-order model [Kim et al., 2006c]. For this reason, movement models are generally limited to random-walk models which permit to constrain the trajectory smoothness [Koyama et al., 2010b] [Brockwell et al., 2004].

The dependence between measurements $\mathbf{x}^t \in \mathbb{R}^m$ and hidden state value $\mathbf{y}^t \in \mathbb{R}^n$ is described by the emission equation (2.2), where $\mathbf{v}^t$ is the observation noise. As the emission equation models how neural features are generated conditionally to a given trajectory point, state-space models are sometimes referred to as "generative model" [Wu et al., 2002] [Gao et al., 2003] [Kim et al., 2006c].

Recursive Bayesian estimation procedures are generally used to infer the hidden trajectory $\mathbf{y}^t \in \mathbb{R}^n$ from the sequence of noisy measurements $\mathbf{x}^t \in \mathbb{R}^m$ [Bishop, 2006].

Recursive Bayesian estimation: Kalman filter The Kalman Filter (KF) is a recursive estimation procedure which has been frequently utilized for online and offline trajectory reconstruction. It was first applied for 2D offline hand trajectory decoding from SUA/MUA signals in monkeys [Wu et al., 2002] [Wu et al., 2003a],

where it was found to surpass linear filtering [Wu et al., 2003a]. It has since then permitted to provide users with MUA/SUA-based control over prostheses [Hochberg et al., 2012]. It has additionally been applied for trajectory decoding from ECoG signals in online and offline studies [Pistohl et al., 2008] [Kellis et al., 2012] [Marathe and Taylor, 2013] [Wang et al., 2013b]. KF applies to linear Gaussian state-space models [Bishop, 2006]:

$$\mathbf{y}^{t+1} = \mathbf{A}\mathbf{y}^t + \mathbf{w}^t \ , \tag{2.3}$$

$$\mathbf{x}^t = \mathbf{C}\mathbf{y}^t + \mathbf{v}^t \ . \tag{2.4}$$

Here, the emission and transition models are linear, and the corresponding noises are Gaussian: $P(\mathbf{w}^t) \sim \mathcal{N}(0, \mathbf{\Gamma}), \mathbf{\Gamma} \in \mathbb{R}^{n \times n}$, $P(\mathbf{v}^t) \sim \mathcal{N}(0, \mathbf{\Sigma}), \mathbf{\Sigma} \in \mathbb{R}^{m \times m}$.

After training, typically performed using Ordinary Least Squares [Wu et al., 2002], the KF issues the estimate $\hat{\mathbf{y}}^t = E(\mathbf{y}^t | \mathbf{x}^{1:t})$.

Alternative recursive estimation procedures To the best of our knowledge, dynamical modelling of ECoG data has been restricted to Gaussian state-space models, i.e. Kalman Filtering procedures. However, further investigations have been carried out to ascertain the interest of non-linear and/or non-Gaussian state-space modelling of MUA/SUA data. If non-linear and/or non-Gaussian state-space representations integrate more realistic emission and noise models (e.g., Poisson noise for spiking counts), the associated trajectory estimation procedures are often approximate and/or computationally expensive [Koyama et al., 2010a] (e.g., Unscented Kalman Filter (UKF) in [Li et al., 2009] [Ifft et al., 2013], particle [Brockwell et al., 2004], point-process or Laplace-Gaussian Filtering (LGF) [Velliste et al., 2014] in the case of Poisson noise).

The relevance of non-linear emission models has in particular been studied for MUA/SUA offline decoding in [Gao et al., 2003]. The emission model was modelled using linear models, GLM or GAM associated with Poisson noises [Gao et al., 2003]. Non-linear models, and particularly GAM-based emission models, were found to improve the quality of trajectory estimation. In [Koyama et al., 2010b], KF and the LGF (i.e., procedure for non-linear emission models and Poisson noise) performed similarly for offline trajectory reconstruction from primate SUA/MUA signals. An additional closed-loop study suggested a slight superiority of the LGF over the KF [Koyama et al., 2010b]. Analogously, the UKF proposed in [Li et al., 2009] surpassed traditional KF for a task of trajectory reconstruction from MUA/SUA signals.

These last few years, dynamical models have emerged as a promising and efficient alternative to static (typically linear) models [Li, 2014] [Srinivasan et al., 2007]. Since its first application in 2002, the Kalman filter and its variants have been increasingly applied for both online and offline SUA/MUA decoding [Wu et al., 2002] [Hochberg et al., 2012] [Aggarwal et al., 2013]. Because of the deterrent computational burden of its variants, recursive estimation during closed-loop experiments has mainly been

achieved via Kalman Filtering [Li et al., 2009]. While the respective performance of Wiener and Kalman filters appeared to depend on the decoding task at hand in [Kim et al., 2006c], a few systematic comparisons of static and dynamic models have given a steady ground to the popularity of the Kalman filter: static linear models were outperformed by dynamical ones for open-loop [Aggarwal et al., 2009] and/or closed-loop SUA/MUA decoding [Koyama et al., 2010b]. KF embedded approach for smoothing has been identified as particularly efficient in both open- and closed-loop studies [Koyama et al., 2010b].

By contrast, the respective suitability of static and dynamic continuous models for ECoG decoding is still unclear. ECoG-based neural control has been achieved by means of a linear decoder both in Non-Human Primates [Williams et al., 2013] and human subjects [Wang et al., 2013c] (2D and 3D effector control, respectively). The use of linear decoding has also been reported for ECoG offline trajectory reconstruction [Schalk et al., 2007] [Chao et al., 2010] [Shimoda et al., 2012] [Liang and Bougrain, 2012] [Nakanishi et al., 2013] [Williams et al., 2013] [Hammer et al., 2013] [Wang et al., 2013c] [Hotson et al., 2014] [Bundy et al., 2016]. Up to 7DoF have been reconstructed in offline feasibility studies led on Monkeys [Chao et al., 2010] [Shimoda et al., 2012] and human subjects [Nakanishi et al., 2013] [Schalk et al., 2007]. On the other hand, Kalman filtering has permitted to reconstruct 2D kinematic parameters from ECoG signals [Pistohl et al., 2008] [Kellis et al., 2012] [Wang et al., 2013b] [Marathe and Taylor, 2013]. In a comparative study performed on ECoG data [Eliseyev and Aksenova, 2016], static models outperformed Kalman Filtering for the reconstruction of kinematic parameters from high-dimensional time-space-frequency feature representations. One reason for these findings could be a lesser relevance of generative approaches for high-dimensional ECoG data. In another study led on ECoG data [Marathe and Taylor, 2013], however, Kalman-based cursor control was more precise than linear-decoder-based control.

2.5 Post-processing

Post-processing techniques rely on *a priori* knowledge about specific characteristics of $\mathbf{y}^t$ to improve the estimates $\hat{\mathbf{y}}^t$. They can be used to improve the quality of both discrete-valued and continuously-valued intention estimates, and their application has for example been reported for asynchronous binary [King et al., 2015] and/or kinematic decoding [Wang et al., 2011].

2.5.1 Discrete output

Post-processing is mainly used to take into account the *a priori* knowledge that fast switches between classes are unlikely. Typical post-processing methods are filtering of the classifier output [Mason and Birch, 2000][Millán and Mouriño, 2003] [Bashashati et al., 2007b] [King et al., 2015], triggering a state transition after successive identical state estimates only [Townsend et al., 2004] [Pfurtscheller et al.,

2010], or blocking state transitions for a predefined duration after a performed transition [Townsend et al., 2004] [Pfurtscheller et al., 2010].

2.5.2 Continuous output

Post-processing of a continuous estimate $\hat{\mathbf{y}}^t$ generally aims at improving the smoothness of the corresponding decoded trajectory. The use of temporal filtering has been investigated in [Marathe and Taylor, 2015]. The improvement of other trajectory characteristics requires the existence of an *a priori* trajectory model which is seldom available. A complex *a priori* finger trajectory model based on a switching non-linear dynamic model was built in [Wang et al., 2011]. In particular, this model integrated *a priori* knowledge about the succession of rest (NC), flexion and extension states and about the maximal amplitude of finger movements. The switching post-processing model was applied on the output of a linear decoder fitted on both NC and IC samples, and permitted both to support NC state and accurately decode multi-limb trajectories [Wang et al., 2011].

2.6 Transducers for accurate asynchronous kinematic decoding

Kinematic decoders are considered in the present doctoral work. Asynchronous kinematic decoders are suited for clinical motor BCIs if they associate null-velocity kinematic estimates with each limb during NC states. In the case of sequential multi-limb decoding, it is additionally desirable for the estimated velocity of limbs $j,\ i \neq j$ to be null when limb i is moving. Finally, decoding accuracy during IC states is crucial for kinematic decoders to be profitable to patients.

Several transducers have been proposed to handle accurate asynchronous mono- and/ multi-limb decoding. While they generally relied on generic pre-processing and feature extraction approaches, they embedded specific decoders and/or post-processing operators. These decoders and post-processing operators exploited and possibly combined the algorithmic tools presented in the previous sections, for example continuous decoders and/or discrete decoders.

2.6.1 Asynchronous mono-limb decoding

NC states were not supported in the majority of kinematic motor BCIs, which were deployed using synchronous paradigms [Hochberg et al., 2006] [Wodlinger et al., 2015] [Hochberg et al., 2012]. Although generic linear static and dynamic models are favored for motor BCI kinematic decoders (see **section 2.4.3**), they usually fail to output zero-velocity (neutral) estimates when they are used for asynchronous decoding and are applied to NC states [Chao et al., 2010] [Shimoda et al., 2012] [Velliste et al., 2014]. The development of specific algorithmic strategies is thus required to integrate NC support into kinematic decoders, and thereby limit or eliminate spurious movements during NC states. The use of different decoding

strategies has been reported. While most of them were validated with offline analyses, the integration of NC support into a kinematic decoder was nevertheless utilized for asynchronous control of a robotic arm in monkeys in [Suway et al., 2013].

Decoder-based strategies A first approach to asynchronous decoding consists in managing NC states at the level of the decoder. While the use of generic non-linear models has been considered in a few studies (e.g., GLM or GAM [Eliseyev and Aksenova, 2014]), the most popular decoding approach consists in switching between continuous models. Switching models rely on a latent discrete variable to introduce state-specific non-linearities into a generic continuous decoder [Srinivasan et al., 2007] [Wood et al., 2005] [Bundy et al., 2016]. More specifically, the dependence between $\mathbf{x}^t$ and $\mathbf{y}^t$ is conditioned on a discrete latent variable z^t, i.e. the decoder switches between models according to the value of z^t. When these non-linearities are associated with NC and IC states, i.e. $z^t = 0$ for NC states and $z^t = 1$ for IC states, the resulting decoder is expected to correspond to limited spurious activations during NC states. Both static [Williams et al., 2013] [Bundy et al., 2016] and dynamic [Wood et al., 2005] [Srinivasan et al., 2007] switching models have been considered for asynchronous mono-limb decoding.

Static switching models rely on both discrete and continuous algorithmic tools, as a discrete decoder is used to infer the value of the switching state variable. In [Bundy et al., 2016] for example, the output of a logistic regression was used to switch between 2 linear models. One of these models was dedicated to kinematic decoding during IC states, and the other one yielded neutral kinematic estimates during NC states. While they were not presented as switching models, other decoders can be formulated as switching models, for example the decoder used in [Williams et al., 2013]. The latter applied a linear kinematic model on neural features when IC states had been detected by a discrete decoder.

The use of dynamic switching models has also been proposed for the integration of NC support into generic dynamical models [Wood et al., 2005] [Srinivasan et al., 2007]. A Switching Kalman Filter and a Switching Particle Filter were implemented and tested in [Srinivasan et al., 2007] and [Wood et al., 2005], respectively. The value of the latent variable was used to switch between observation and/or transition models. The transition model associated with NC states explicitly modelled the fact that null-velocity estimates are expected during NC states.

Post-processing-based strategies A second approach to asynchronous mono-limb decoding consists in using a single model to describe the dependence between neural signals and kinematic parameters during both NC and IC states [Wang et al., 2013b] [Velliste et al., 2014]. A binary discrete variable is explicitly decoded using a discrete decoder. The output of the single kinematic model is overwritten with null-velocity (neutral) estimates when a NC state is detected by the discrete decoder. Post-processing-based integration of NC support into kinematic decoders has been explored for both SUA/MUA [Velliste et al., 2014] [Aggarwal et al., 2013] and ECoG

signal (e.g., [Wang et al., 2013b]) decoding. Different pairs of discrete-continuous decoders have been considered.

Kalman filters and (non-Gaussian) variants were gated by LDA in SUA signals [Velliste et al., 2014] [Aggarwal et al., 2013] or Bayes classifier in ECoG signals [Wang et al., 2013b]. A Kalman filter or one of its variants was continuously applied on neural signals, and its output was possibly overwritten with neutral values when an independent classifier detected a NC state [Aggarwal et al., 2013] [Wang et al., 2013b] [Velliste et al., 2014].

Finally, a dynamic switching post-processing model permitted to integrate NC support into a finger kinematic decoder in [Wang et al., 2011]. It was applied on the output of a linear decoder fitted on both IC states and NC samples located near movement onsets and offsets.

2.6.2 Asynchronous sequential multi-limb decoding

The design of continuous decoders for asynchronous sequential multi-limb neural control has rarely been considered, and is not straightforward. Although one limb only should be active at each time moment, the activation of one limb is liable to result in residual movements of the other limbs. Such noisy outputs were for example observed in [Nakanishi et al., 2014b], where the displacement of a finger resulted in small-amplitude movements in the estimations of the other fingers' position.

Decoder-based strategies A switching model was considered for sequential asynchronous multi-finger decoding in [Flamary and Rakotomamonjy, 2012]. One linear model was devoted to each finger, and applied when deemed appropriate by a multi-class discrete decoder. The linear model associated with a finger was exclusively trained on samples acquired during the movement of this specific finger. Such switching models intrinsically prevent parallel activations as only one active limb model is chosen at each instant.

Post-processing-based strategies While a post-processing approach has been applied to improve the reconstruction of sequential multi-finger movements in [Wang et al., 2011], the considered post-processing operator did not specifically addressed the issue of sequential multi-limb decoding, as each finger was independently considered. Thus, to the best of our knowledge, the use of post-processing approaches has not yet been considered for asynchronous sequential multi-limb decoding.

2.6.3 Accurate decoding during IC states

Decoding accuracy during IC states is a typical objective for kinematic decoders.

Decoder-based strategies The improvement of decoding accuracy is often attempted by using complex non-linear models, for example ANN [Kim et al., 2006c], Support-Vector-Regression [Spüler et al., 2016] or Unscented Kalman Filters [Ifft

et al., 2013]. The use of static switching models has in particular been explored for accurate trajectory decoding during active states [Kim et al., 2003]. In [Kim et al., 2003] for example, a discrete latent variable permitted to switch between 10 linear models dedicated to different phases of a reaching movement, and led to an increase of the decoding accuracy for a task of MEA-based trajectory reconstruction. Trajectory decoding from MEA signals was improved by combining reaching movement models in [Kemere et al., 2004b] [Kemere et al., 2004a] [Yu et al., 2007]. Finally, the use of a switching state-space filter has been reported in [Wu et al., 2004]. The objective lied in improving MEA decoding accuracy by using emission models conditioned on an unspecified latent cognitive state.

Post-processing-based strategies Post-processing-based approach to the improvement of decoding accuracy are generally limited to the improvement of the smoothness of the kinematic estimates. In [Wang et al., 2011] however, a complex post-processing Bayes filter was used to enhance the estimation of the amplitude of finger flexions and extensions.

2.7 Conclusion

In this chapter, algorithmic tools utilized by the BCI community to build transducers for motor BCIs have been presented. The decoders used to extract discrete or continuous variables from neural signals have in particular been introduced.

Asynchronous mono-limb kinematic decoding is one of the objectives of the present doctoral work. Kinematic transducers rely on continuous decoders, e.g. static linear models or dynamical models like the Kalman Filter. As generic continuous decoders generally fail to issue zero-velocity estimates during No-Control (NC) states, specific decoding and/or post-processing strategies have been proposed to integrate NC support into kinematic decoders, i.e. to limit spurious system activations during NC states. Two main approaches to mono-limb asynchronous decoding have emerged, namely decoder-based and post-processing-based strategies. The introduction of NC support within the decoder is generally performed by means of non-linear continuous decoders. The use of switching models has particularly been reported. Switching models rely on the estimation of a latent discrete variable to introduce state-specific non-linearities into a generic continuous decoder. Discrete decoders are possibly exploited to infer the value of this latent discrete variable. By contrast, a single continuous model is trained and/or applied on both NC/IC (Intentional Control) states when NC support is introduced at the post-processing stage. The estimation of the value of a latent discrete variable is again used to select the post-processing operator applied on NC and IC states.

Asynchronous sequential multi-limb decoding has rarely been considered in BCI studies. The use of switching models has nevertheless been reported for accurate multi-finger control. Parallel limb activations are avoided because one limb only is detected as active at each instant.

Finally, attempts to improve decoding accuracy during IC states often focus on the utilization of non-linear decoders, in particular generic non-linear models or piece-wise linear models. Post-processing has also been used to improve different characteristics of the kinematic estimates, e.g. their smoothness or spatial accuracy.

In the next chapter, the solution which has been chosen in the present doctoral work to address the issue of accurate asynchronous mono- and multi-limb control is presented. This solution relies on a decoder-based approach, and has been developed within the framework of switching models.

Chapter 3

Switching models for accurate asynchronous trajectory decoding

Contents

A switching decoder, namely a Markov Switching Linear Model (MSLM), has been proposed and designed for the task of accurate, asynchronous multi-limb ECoG decoding.

3.1 Switching models for BCI control

The proposed MSLM decoder has been developed to ensure efficient asynchronous multi-limb decoding in the presence of context-related modifications of a linear dependence between neural features and kinematic parameters.

3.1.1 Context-related modifications of kinematic encoding in neural features

The existence of several motor control internal models in humans has been disclosed in several studies [Imamizu et al., 2007] [Rouse and Schieber, 2015]. Consistently, context-related modifications of kinematic encoding in neural features have been

observed in a few BCI studies [Williams et al., 2013] [Velliste et al., 2014] [Ifft et al., 2013]. Ifft and colleagues reported in [Ifft et al., 2013] that modifications of the tuning properties of individual neurons were induced when NHPs performed unimanual or bimanual reaching movements. In [Williams et al., 2013] [Velliste et al., 2014], they were triggered when NHPs switched between NC and IC states. It was reported in [Velliste et al., 2014] that NC samples did not fit the emission model of a Kalman filter trained on IC samples. In [Williams et al., 2013], it was observed that the baseline around which NHPs BCI users modulated their neural features was different during NC and IC periods. These differences resulted in a poor state discrimination when NC states were inferred for the estimated velocity of the considered effector. Improvement of kinematic decoding accuracy during IC periods has been reported when mixtures of models were applied to MUA/SUA signals [Kim et al., 2003] [Wu et al., 2004].

These reports suggest that the framework of switching models may be relevant to perform accurate asynchronous, multi-limb ECoG decoding for motor BCIs.

3.1.2 Switching models

The use of switching models has been reported in numerous fields. Switching Auto-Regressive models have notably been applied for economic forecasting [Quandt, 1958] [Goldfeld and Quandt, 1973] [Quandt and Ramsey, 1978] [Kim, 1994], for cognitive state detection from neural signals [Liehr et al., 1999] and for efficient control of dynamical systems subject to regime changes [Alessandri et al., 2008]. In particular, switching models have been utilized in BCI studies for accurate [Kim et al., 2003] [Wu et al., 2003a] [Wu et al., 2004], asynchronous [Williams et al., 2013] [Bundy et al., 2016] [Srinivasan et al., 2007] and/or multi-finger kinematic decoding [Flamary and Rakotomamonjy, 2012]. In these studies, several continuous models were applied on the input variable. Their relevance at time t was dependent on the value of an unobserved (hidden, latent) discrete variable. The resulting kinematic estimates were combined by a gate which estimated the value of the latent variable. Two main frameworks have been used to develop these switching models, namely switching state-space filters and switching regression models.

Generative switching models: switching state-space filters The use of switching state-space (or Bayes) filters has been reported in several BCI studies [Wu et al., 2004] [Wood et al., 2005] [Yu et al., 2007] [Srinivasan et al., 2007]. Bayes filters include a model $P(\mathbf{y}^{t+1}|\mathbf{y}^t, \Theta)$ of the dynamic of the state variable $\mathbf{y}^t$, and are inasmuch dynamic filters. Moreover, the emission probability $P(\mathbf{x}^t|\mathbf{y}^t, \Theta)$ is utilized to characterize the dependence between observed features and latent trajectories. Such state-space models are thus generative [Wu et al., 2002] [Gao et al., 2003] [Kim et al., 2006c]. The switching state-space filters applied in BCI studies extended well known state-space (Bayes) filters, e.g. Kalman or point-process filters (see **section 2.4.3.2**), by conditioning their emission and/or transition models on a discrete latent switching variable. As state-space models rely on the hypothesis that the

generation of neural features is conditioned on the continuously-valued kinematic variable, the probability associated with each possible switching state depends on the extent to which the corresponding couple emission/transition models can explain neural features observed at time t (**Figure 3.1**), e.g. in [Wu et al., 2004] [Yu et al., 2007] [Srinivasan et al., 2007]).

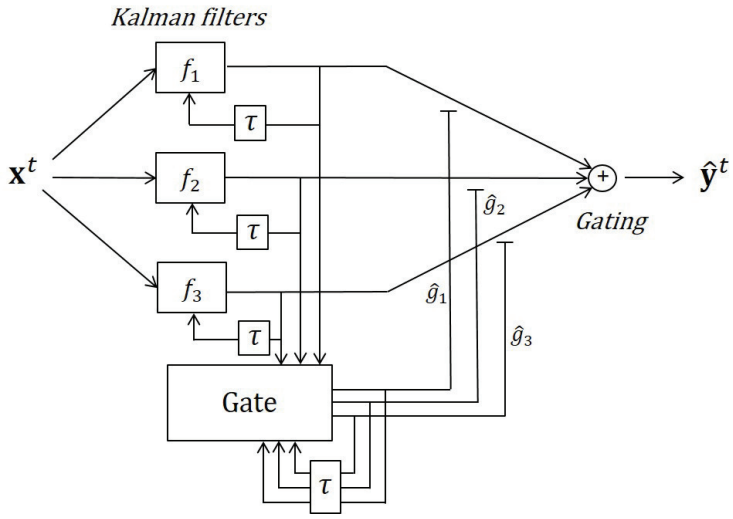

Figure 3.1: Data flow during the application of a switching state space filter, here a Switching Kalman Filter (e.g., [Wu et al., 2003b] [Wu et al., 2004] [Srinivasan et al., 2007]). The probability associated with each possible switching state depends on the extent to which the corresponding couple emission/transition models can explain neural features observed at time t.

Discriminative switching models: switching regression models Switching modelling for kinematic motor BCIs has also been investigated in the framework of regression models [Kim et al., 2003] [Kemere et al., 2004b] [Kemere et al., 2004a] [Flamary and Rakotomamonjy, 2012] [Williams et al., 2013] [Bundy et al., 2016]. Regression models consider the probability $P(\mathbf{y}^t|\mathbf{x}^t, \Theta)$, i.e. they are discriminative static models. In contrast with switching state-space (Bayes) filters, the distribution of the neural features is assumed to be conditioned on the latent switching state variable. The estimation of the value taken by the hidden state variable directly relies on the current neural features (see **Figure 3.2**, e.g. [Bundy et al., 2016] [Flamary and Rakotomamonjy, 2012]).

The MSLM has been developed in the framework of switching regression models. This framework permits to introduce the key hypothesis that switching (cognitive, internal) states are associated with different distributions of the neural features. Such modifications are regularly observed, and are exploited in the majority of BCI classification studies. Moreover, regression models are suitable for the decoding

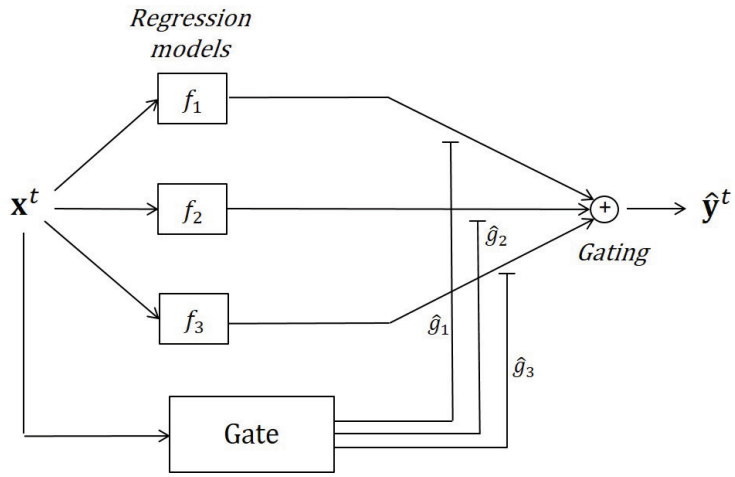

Figure 3.2: Data flow during the application of switching regression models (e.g., [Kim et al., 2003] [Bundy et al., 2016]). The gate directly exploits the distribution of the current neural features to estimate the value of the latent switching state.

of potentially high-dimensional neural features. By contrast, the applicability of generative models like the Kalman filter may be limited in high-dimensional settings because of deterrent computational costs (see **Appendix C**). A preliminary dimensionality reduction is regularly performed when state-space models are used to model the generation of high-dimensional data [Wu et al., 2003b] [Kao et al., 2013].

The Mixture of Experts (ME) framework [Jacobs et al., 1991] [Waterhouse, 1997] can be used to unify the switching regression models proposed in BCI studies.

3.1.3 Mixture of Experts structure for discriminative switching models

First introduced by [Jacobs et al., 1991], MEs combine several functions ("experts") to model the dependence between input and output variables space [Jacobs et al., 1991] [Waterhouse, 1997]. Experts are weighted according to the input variable [Bishop, 2006]. Optimal weights are computed for each sample according to a given criterion.

Let $\mathbf{x}^t \in X \subset \mathbb{R}^m$ be the explanatory (input, independent) variable, and $\mathbf{y}^t \in Y \subset \mathbb{R}^n$ be the continuous response (output, dependent) variable. Samples are indexed by $t \in \mathbb{N}$. MEs assume that the input space X is partitioned into K (possibly overlapped) regions: $X = \bigcup_{k=1}^{K} X_k$, and that a different sub-process generates the output vector from the input vector in each region [Waterhouse, 1997]. Thus, Mixtures of Experts model the dependence between $\mathbf{x}^t$ and $\mathbf{y}^t$ by means of a set of K local functions $\{f_k\}_1^K$, $f_k : X \to Y$, referred to as "experts". Expert k is

relevant for the region X_k:

$$\mathbf{y}^t = \sum_{k=1}^{K} \delta_{z^t,k}\ f_k(\mathbf{x}^t) + \boldsymbol{\epsilon}^t \tag{3.1}$$

where $\delta_{z^t,k} = 1$ if $z^t = k$ and $\delta_{z^t,k} = 0$ otherwise. Here, $z^t = k$ if $\mathbf{y}^t$ has been generated by expert k, and $\boldsymbol{\epsilon}^t$ is the observation noise.

The Bayes estimate $\hat{\mathbf{y}}^t = E(\mathbf{y}^t|\mathbf{x}^t)$ [Bishop, 2006] of target variable $\mathbf{y}^t$ is computed via the decomposition of conditional expectation [Waterhouse, 1997]:

$$E(\mathbf{y}^t|\mathbf{x}^t) = \sum_{k=1}^{K} E(\mathbf{y}^t, z^t = k|\mathbf{x}^t) \tag{3.2}$$

$$= \sum_{k=1}^{K} P(z^t = k|\mathbf{x}^t)\ E(\mathbf{y}^t|\mathbf{x}^t, z^t = k) \tag{3.3}$$

$$= \sum_{k=1}^{K} g_k^t \hat{\mathbf{y}}_k^t \tag{3.4}$$

Here, $\hat{\mathbf{y}}_k^t = E(\mathbf{y}^t|\mathbf{x}^t, z^t = k)$ is the estimation issued by expert k.

The mixing coefficients $g_k^t = P(z^t = k|\mathbf{x}^t)$ combine ("gate") the experts' estimates $\hat{\mathbf{y}}_k^t$. In contrast with generic mixtures of regression models which rely on fixed prior probabilities to combine the regression models' outputs [Bishop, 2006], mixing coefficients are here input-dependent. They can be interpreted as the conditional probability of expert f_k having generated the output value $\mathbf{y}^t$, given the value of $\mathbf{x}^t$ [Waterhouse, 1997]. Mixing coefficients satisfy the constraints $g_k^t \in [0, 1]$, $\sum_k g_k^t = 1$ [Bishop, 2006]. The structure which computes the mixing coefficients is often referred to as "gating network". Linear or nonlinear experts and gating networks can be integrated into the ME flexible structure [Yuksel et al., 2012].

ME-based approaches have been used in several BCI studies which didn't support NC periods or multilimb decoding, e.g. for multi-model wrist movement decoding from monkeys' SUA/MUA signals [Kim et al., 2003]. The switching regression models previously implemented in asynchronous and/or multi-finger BCI studies, e.g. [Williams et al., 2013] [Flamary and Rakotomamonjy, 2012] [Bundy et al., 2016], can also be formulated as static mixtures of regression models. In the majority of these ECoG-based BCI studies, the estimate yielded by a continuous decoder was overwritten by neutral values when a classifier dedicated to state detection indicated that the current state was most likely a NC state [Williams et al., 2013] [Flamary and Rakotomamonjy, 2012] [Bundy et al., 2016]. More precisely, one or several Wiener filters were weighted by the outputs of a logistic regression [Bundy et al., 2016] or of a linear regression of state labels on neural signals [Flamary and Rakotomamonjy, 2012] [Williams et al., 2013]. These Wiener filters can be interpreted as linear experts, and the classifiers used to combine them can be interpreted as

gates making use of soft [Chen et al., 2014b] or hard (Winner-takes-All) [Flamary and Rakotomamonjy, 2012] [Williams et al., 2013] [Bundy et al., 2016] probabilistic strategies to mix the experts' outputs, where the application of hard combination rules consists of associating a weight of 1 to the most likely current state, and a null weight to all other possible states.

Although recurrent gates were embedded into the previously mentioned switching state-space filters [Wu et al., 2004] [Srinivasan et al., 2007], the switching regression models which application has been reported for BCI kinematic decoding relied on static gates [Flamary and Rakotomamonjy, 2012] [Bundy et al., 2016]. Such static gates are based on the assumption that pairs $(\mathbf{x}^t, z^t)$ are temporally independent and identically drawn from the distribution $P(\mathbf{x}^t, z^t)$ [Dietterich, 2009]. As the classes associated with internal states are likely to be overlapped in BCI studies, this discrimination process is liable to issue erroneous estimates of the internal states. More realistic hypotheses about temporal dependencies in the state sequences have been exploited in BCI studies (see **section 2.4.2.3**), and in particular in a few classification-based motor BCIs [Fifer et al., 2014] [King et al., 2015] [Hotson et al., 2016]. When compared to static gates, dynamic (recurrent) gates are expected to be less liable to output spurius detections (e.g., [Saa et al., 2016]). The MSLM was thus designed as a switching regression model with recurrent gating for robust estimation of the user's latent switching state (**Figure 3.3**).

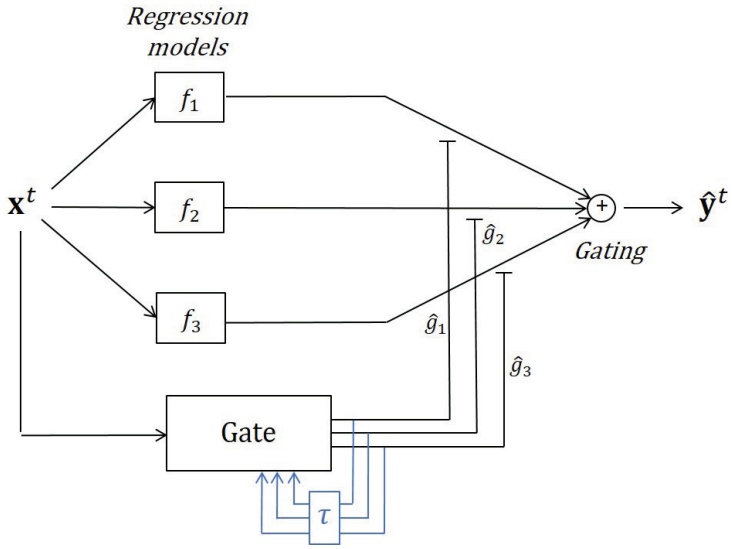

Figure 3.3: Data flow during the application of the MSLM.

3.2 Markov Switching Linear Models (MSLM)

The MLSM is built on the assumption that the continuous model which relates neural features $\mathbf{x}^t$ and kinematic parameters $\mathbf{y}^t$ is a linear model conditioned on a discrete latent variable $z^t \in \mathbb{N}$ (switching state). A probabilistic rule is used to combine the linear models respectively associated with each value z^t. The discrete latent variable z^t is assumed to be generated by a first-order Markov chain. The Mixture of Experts (ME) framework [Jacobs et al., 1991] [Waterhouse, 1997], which was used in **section 3.1.3** to formalize static mixtures of regression models, was here extended to support dynamic mixtures of regression models.

3.2.1 Dynamic gating

The MSLM embeds a dynamic gate, which is liable to reduce the number of erroneous state estimates.

A parallel can be drawn between efficient state classifiers and efficient ME gates if gate optimization is momentarily considered independently of expert application. The potential sub-optimality of static classifiers for cognitive state detection has been taken into account by the BCI community, inasmuch as the use of several strategies has been reported for dynamic state detection in several BCI studies. A first approach consists in integrating the *a priori* knowledge that fast switches between internal states are unlikely at the post-processing stage. Several strategies were considered and possibly combined, namely filtering of a NC/IC classifier output [Mason and Birch, 2000] [Millán and Mouriño, 2003] [Bashashati et al., 2007a] [King et al., 2015], triggering a state transition after successive identical state estimates only [Townsend et al., 2004] [Pfurtscheller et al., 2010], and/or by blocking state transitions for a predefined duration after a performed transition [Townsend et al., 2004] [Pfurtscheller et al., 2010]. A second approach consists in using dynamic classifiers (see **section 2.4.2.3**), which explicitly integrate a model of the state succession dynamic. In particular, dynamic classifiers such as Hidden Markov Models (HMM) [Kemere et al., 2008] [Hotson et al., 2016] and Conditional Random Fields (CRF) [Hasan and Gan, 2011] [Saa et al., 2016] were used for the estimation of NC/IC states, possibly with several NC- or IC-related sub-states, in SUA/MUA [Kemere et al., 2008], EEG [Hasan and Gan, 2011] and ECoG signals [Saa et al., 2016].

Dynamic gating has additionally been considered in the ME framework. Several extensions have been specifically proposed to adapt MEs for sequential data modelling [Yuksel et al., 2012]. Recurrent gating networks [Cacciatore and Nowlan, 1994] [Meila and Jordan, 1996] were integrated into ME and for example applied for movement modelling [Meila and Jordan, 1996] and efficient control of switching systems [Cacciatore and Nowlan, 1994]. Particularly, Markov Mixtures of Experts (MME) [Meila and Jordan, 1996] [Bengio and Frasconi, 1996] (referred to as Input-Output HMM in [Bengio and Frasconi, 1996]) model a temporal sequence of input-output data by dynamic switching between local static models. The expert sequence

$(z^t)_{t\in\mathbb{N}}$ is assumed to be generated by a first-order Markov chain. The inputs and outputs corrupted by noise are measured, but states are hidden and must be estimated from measurements.

Following [Fifer et al., 2014] [Hotson et al., 2016] [Srinivasan et al., 2007], it was here decided to explicitly model the temporal dependencies between internal states using a first-order Markov chain. The MSLM proposed herein is a variant of MME which relies on Hidden Markov Models (HMMs) for dynamic gating of the experts' estimates. It relies on the underlying hypothesis that neural signals are generated by internal (cognitive) states. Thus, the distribution of the extracted neural features $\mathbf{x}^t$ is conditioned on unobserved discrete states z^t, $P(\mathbf{x}^t|z^t)$. By contrast, the previously reported MMEs and IOHMMs [Bengio and Frasconi, 1996] [Meila and Jordan, 1996] assumed that the input variable $\mathbf{x}^t$ conditions the probability of switching from one state to another, $P(z^{t+1}|z^t, \mathbf{x}^t)$.

3.2.1.1 HMM-based dynamic gating

Hidden Markov Models (HMMs) are a powerful tool for the modelling of stochastic time-varying processes [Rabiner and Juang, 1986]. Two discrete-time stochastic processes are involved in discrete-time HMMs. The first process is an unobservable ("hidden") dynamic sequence $(z^t)_{t\in\mathbb{N}}$. By contrast, the second sequence $(\mathbf{x}^t)_{t\in\mathbb{N}}$ is observable, and is referred to as the "observation sequence". The MSLM state decoder models the state $(z^t)_{t\in\mathbb{N}}$ and feature sequences $(\mathbf{x}^t)_{t\in\mathbb{N}}$ by a HMM, where $z^t \in \mathbb{N}$ is the hidden variable and $\mathbf{x}^t \in \mathbb{R}^m$ is the observed variable.

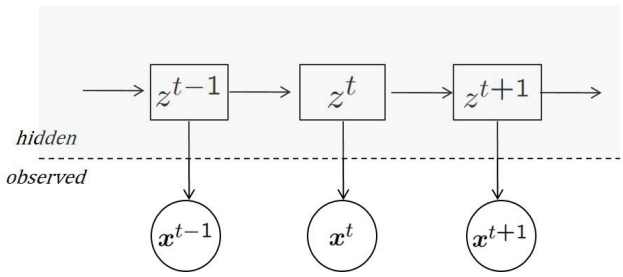

Figure 3.4: Hidden Markov Model. The sequence $(z^t)_{t\in\mathbb{Z}}$ is hidden. The sequence $(\mathbf{x}^t)_{t\in\mathbb{Z}}$ is observed and permits to infer an estimate $(\hat{z}^t)_{t\in\mathbb{Z}}$ of the hidden sequence.

HMM's hidden variable z^t can take K distinct discrete values ("states"), $z^t \in \mathbb{Z}$ with $Z = \{z_1, z_2, ..., z_K\} \subset \mathbb{Z}$. For the sake of simplicity, $\{z_1, z_2, ..., z_K\}$ is here taken to be $\{1, 2, ..., K\}$. In the typical case of a first-order HMM, the value of z^{t+1} depends exclusively on the current value z^t: $P(z^{t+1} = k|z^{1:t}) = P(z^{t+1} = k|z^t)$, $t \in \mathbb{N}$. The hidden state sequence is then fully characterized by the transition matrix $\mathbf{A} = (a_{ij})$, $a_{ij} = P(z^{t+1} = j|z^t = i)$, and by the probability π associated with the value of the first hidden state: $\pi_i = P(z^1 = z_i)$. The probability $P_i(d)$ to remain in state i for exactly d consecutive time steps is therefore explicitly modelled as $P_i(d) = P(z^1 = z_i, z^2 = z_i, \ldots, z^d = z_i, z^{d+1} = z_j) = (a_{ii})^{d-1}(1 - a_{ij})$, where

$i \neq j$. The probability distribution of the observed variable $\mathbf{x}^t \in \mathbb{R}^m$ is exclusively conditioned on the current hidden state z^t: $P(\mathbf{x}^t|z^{1:t}) = P(\mathbf{x}^t|z^t)$, $t \in \mathbb{N}$, where the conditional probability $P(\mathbf{x}^t|z^t)$ is referred to as "state emission density". Various parametric (Gaussian, Student) and non-parametric state emission distributions have been integrated in HMMs [Rabiner, 1989]. Although conventional HMMs thus directly model the emission distributions $P(\mathbf{x}^t|z^t = j),\ j, \ldots, K$, variants based on discriminative modelling have been proposed by several teams (e.g., using Neural Networks [Renals et al., 1994] [Bourlard and Morgan, 1998] [Ordóñez et al., 2013] or Support Vector Machines [Valstar and Pantic, 2007] [Ordóñez et al., 2013]). These variants consist in applying a discriminative model on the data, thereby obtaining the conditional probability $P(z^t = j|\mathbf{x^t})$. Class priors $P(z^t = j)$ and Bayes rule are then combined to compute HMM emission probabilities $P(\mathbf{x}^t|z^t = j) \propto \frac{P(z^t=j|\mathbf{x^t})}{P(z^t=j)}$ [Bourlard and Morgan, 1998] [Valstar and Pantic, 2007] [Ordóñez et al., 2013]. Thus, the MSLM can embed both discriminative and generative state detection into its HMM-based gate.

Let $\{b_1, \ldots, b_K\}$ gather the parameters necessary to characterize the distributions $P(\mathbf{x}^t|z^t = j), j = 1, \ldots, K$. Then the HMM is fully described by the parameter set $\lambda = \left\{\mathbf{A},\ \{b_j\}_{j=1}^K,\ \pi\right\}$. HMM training is performed using supervised or unsupervised approaches, and typically relies on Maximum-Likelihood estimation [Ghahramani, 2001].

3.2.1.2 MSLM online gating

As online estimation of users' intentions is required for the utilization of BCI systems, the MSLM's input-dependent mixing coefficients are exclusively conditioned on past and current neural observations: $g_k^t = P(z^t = k|\mathbf{x}^{1:t})$. Mixing coefficients are first decomposed using Bayes rule:

$$P(z^t = k|\mathbf{x}^{1:t}) = \frac{P(z^t = k,\ \mathbf{x}^{1:t})}{P(\mathbf{x}^{1:t})}. \tag{3.5}$$

The Viterbi algorithm or the forward-backward algorithm are often used to infer the hidden state sequence $\mathbf{z} = (z^t)_{t=1}^T$ associated with an observation sequence $\mathbf{X} = (\mathbf{x}^t)_{t=1}^T$. Both Viterbi and forward-backward algorithms cannot be applied for online processing because they rely on a back-tracking procedure from the last estimated state $\hat{z}^T = \underset{k}{\operatorname{argmax}}\, P(z^t = k \mid \mathbf{x}^{1:T})$ [Rabiner, 1989]. By contrast, the forward algorithm [Rabiner, 1989] permits to compute estimates $P(z^t = k|\mathbf{x}^{1:t})$ which are only conditioned on past and current observations. The denominator $P(\mathbf{x}^{1:t})$ and numerator $P(z^t = k,\ \mathbf{x}^{1:t})$ of (3.5) are thus computed using the HMM's forward algorithm for online application of the MSLM.

3.2.1.3 Soft and hard gating

Both discrete-valued [Flamary and Rakotomamonjy, 2012] [Suway et al., 2013] [Bundy et al., 2016] and continuous-valued mixing coefficients [Chen et al., 2014b] were used

in the switching regression models previously implemented for neural signal decoding. When the mixing coefficients are discrete-valued, i.e. $g_k^t \in \{0,1\}\, \forall t, k = 1, \ldots, K$, the model which is associated with the most likely class at time t is applied on the neural features $\mathbf{x}^t$ ("hard" combination). This Winner-Takes-All strategy (or "hard" switching) has been used in the majority of switching kinematic decoders [Flamary and Rakotomamonjy, 2012] [Suway et al., 2013] [Bundy et al., 2016]. By contrast, the utilization of continuously-valued mixing coefficients, i.e. $g_k^t \in [0,1]$, has rarely been reported in BCI studies. It was for example used in [Chen et al., 2014b] to weight IC models for asynchronous finger trajectory reconstruction. In contrast with hard model combination which results in abrupt transitions between decoding regimes, this probabilistic, "soft" combination results in smooth transitions between experts.

Although abrupt transitions may not be absolutely deterrent because users may, to a certain extent, learn to correct high-frequency errors [Stavisky et al., 2015] [Marathe and Taylor, 2015], we expect them to be disturbing to BCI users and therefore undesirable. Without loss of generality, the MSLM thus makes use of a soft, probabilistic strategy to combine models.

3.2.2 Experts

Any expert $f_k : \mathbb{R}^m \to \mathbb{R}^n$ can be integrated into the ME switching structure. The extraction of kinematic parameters from ECoG features during IC periods has mainly been performed by means of linear models, in particular in the case of wrist [Schalk et al., 2007] [Hammer et al., 2016] [Bundy et al., 2016] and finger [Kubánek et al., 2009] [Liang and Bougrain, 2009] [Flamary and Rakotomamonjy, 2012] trajectory estimation. Linear experts were therefore integrated into the proposed switching model. If a Winner-takes-all strategy is used to combine its experts' estimates, the MSLM can be seen as a piecewise linear model:

$$\mathbf{y}^t = \sum_{k=1}^{K} \delta_{z^t,k} \mathbf{B}_k \mathbf{x}^t + \boldsymbol{\epsilon}^t \qquad k = 1, \ldots, K, \tag{3.6}$$

where $\mathbf{B}_k \in \mathbb{R}^{n \times m}$ and $\boldsymbol{\epsilon}^t \in \mathbb{R}^n$ is the observation noise.

The use of other types of nonlinear models, e.g. Generalized Linear Models [Eliseyev and Aksenova, 2014], has been explored for (possibly asynchronous) ECoG decoding. When compared to generic non-linear models, the MSLM makes use of the prior knowledge that discrete internal states are the sources of the model non-linearity. It benefits from a simple application procedure, a straightforward interpretation, and can be easily extended by adding a new state (e.g., progressive introduction of new limbs during training). Moreover, as GLMs apply a non-linear link function on an intermediary trajectory estimate, i.e. state and trajectory decoding are not independent, they can be seen as a sophisticated post-processing approach to the integration of NC support into regression models.

3.3 MSLM training

Let $\Theta = \{\Theta^g, \Theta^e\}$ gather the HMM-based gating network $\Theta^g = \left\{\mathbf{A}, \{b_j\}_{j=1}^K, \pi\right\}$ and the experts' $\Theta^e = \{\mathbf{B}_k\}_{k=1}^K$ parameters, where, depending on context, $\mathbf{B}_k$ refers either to the full set of parameters associated with expert i (i.e., parameters of the linear model and of the noise distribution) or only to the parameters of the linear model. Prior to any MSLM application, estimation of both expert and gating networks' parameters is required. When the expert responsibility is known, training is supervised with respect to the expert sequence. Training is otherwise unsupervised with respect to the expert sequence.

The Maximum-Likelihood (ML) criterion is routinely used for both EM [Yuksel et al., 2012] and HMM training [Ghahramani, 2001]. ML training procedures were derived for supervised and unsupervised training of the MSLM.

3.3.1 Supervised training

Fully supervised training of the MSLM relies on a training data set $\{\mathbf{X}, \mathbf{Y}, \mathbf{z}\}$, where $\mathbf{X} \in \mathbb{R}^{T\times m}$, $\mathbf{Y} \in \mathbb{R}^{T\times n}$, and $\mathbf{z} \in \mathbb{N}^{T\times K}$ gather the observed sequences $(\mathbf{x}^t)_{t=1}^T$, $(\mathbf{y}^t)_{t=1}^T$ and $(z^t)_{t=1}^T$ respectively.

3.3.1.1 Independent training for gate and experts

Let us consider a MSLM composed of K linear experts gated by a HMM-based sequential decoder. Each linear expert i is parametrized by $\mathbf{B}_i$. Let $\mathbf{A} = (a_{ij})$ be the transition matrix $a_{ij} = P(z^{t+1} = j|z^t = i), i, j = 1, ..., K$ and π be the initial state distribution $\pi_i = P(z^1 = z_i),\ i = 1, ..., K$ associated with the first-order Markovian hidden state variable z^t. Finally, let $\{b_1, \ldots, b_K\}$ gather the parameters necessary to characterize the distributions $P(\mathbf{x}^t|z^t = j), j = 1, \ldots, K$, e.g. mean and variance in the case of Gaussian emissions. Let $\{\mathbf{X}, \mathbf{Y}, \mathbf{z}\}$ be a complete training data set available for Maximum-Likelihood estimation of the MSLM parameters $\Theta = \left\{\{\mathbf{B}_i\}_{i=1}^K, \pi, \mathbf{A}, \{b_i\}_{i=1}^K\right\}$.

Assuming the temporal independence $P(\mathbf{x}^{1:T}, \mathbf{y}^{1:T}|z^{1:T}, \mathbf{\Theta}) = \prod_{t=1}^{T} P\left(\mathbf{x}^t, \mathbf{y}^t|z^{1:T}, \mathbf{\Theta}\right)$, the data complete log-likelihood $L_c(\mathbf{\Theta}, \mathbf{X}, \mathbf{Y}, \mathbf{z}) = \ln P(\mathbf{X}, \mathbf{Y}, \mathbf{z}|\mathbf{\Theta})$ can be expressed as (see **Appendix A**)

$$\begin{aligned} L_c(\Theta, \mathbf{X}, \mathbf{Y}, \mathbf{z}) = &\sum_{i=1}^{K} \delta_{z^1,i} \ln(\pi_i) + \sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{t=1}^{T-1} \omega_{i,j}^t \ln(a_{ij}) \\ &+ \sum_{i=1}^{K}\sum_{t=1}^{T} \delta_{z^t,i} \ln(P(\mathbf{y}^t|\mathbf{x}^t, \mathbf{B}_i)) + \sum_{i=1}^{K}\sum_{t=1}^{T} \delta_{z^t,i} \ln P(\mathbf{x}^t|b_i), \end{aligned} \quad (3.7)$$

where $\delta_{z^t,k} = 1$ if $z^t = k$ and $\delta_{z^t,k} = 0$ otherwise, and $\omega_{i,j}^t = 1$ if $z^t = i$ and $z^{t+1} = j$, and $\omega_{i,j}^t = 0$ otherwise.

Maximizing (3.7) amounts to solving the following maximization problems (see **Appendix A**):

$$\underset{\mathbf{B}_i}{\text{maximize}} \quad \sum_{t=1}^{T} \delta_{z^t,i} \ln P(\mathbf{y}^t|\mathbf{x}^t, \mathbf{B}_i) \qquad i = 1, \ldots, K \tag{3.8}$$

$$\begin{aligned} \underset{\pi}{\text{maximize}} \quad & \sum_{j=1}^{K} \delta_{z^1,j} \ln \pi_j \\ \text{subject to} \quad & \sum_{j=1}^{K} \pi_j = 1 \end{aligned} \tag{3.9}$$

$$\begin{aligned} \underset{\mathbf{A}}{\text{maximize}} \quad & \sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{t=1}^{T-1} \omega_{ij}^t \ln a_{ij} \\ \text{subject to} \quad & \sum_{j=1}^{K} a_{ij} = 1 \quad \forall i \in \{1, \ldots, K\}. \end{aligned} \tag{3.10}$$

$$\underset{\mathbf{b_i}}{\text{maximize}} \quad \sum_{t=1}^{T} \delta_{z^t,i} \ln P(\mathbf{x}^t|b_i) \qquad i = 1, \ldots, K. \tag{3.11}$$

The maximization problems (3.8), (3.9), (3.10), and (3.11) correspond to ML estimation of the expert, initial state, transition and emission parameters, respectively.

3.3.1.2 Expert parameter identification

ML estimates of the experts' parameters $\hat{\mathbf{B}}_k, k = 1, \ldots, K$ are found by solving the problems (3.8). As expressed in (3.8), the training of expert k is based on sample observations $\{\mathbf{X}_k, \mathbf{Y}_k\}$, where $\mathbf{X}_k$ and $\mathbf{Y}_k$ gather training samples observed at times t such that $z^t = k$. The resolution of (3.8) depends on the expert noise distribution.

Let for example assume that the observation noise ϵ^t is Gaussian, i.e. $P(\epsilon^t) = P(\mathbf{y}^t - \mathbf{x}^t\mathbf{B}_k) \sim \mathcal{N}(\mathbf{0}, \sigma\mathbf{I})$, $k = 1, \ldots, K$. It can then be shown that the maximization problem (3.8) amounts to the minimization of the squared error $\|\mathbf{Y}_k - \mathbf{X}_k\mathbf{B}_k\|_2$. The corresponding system is generally undetermined in BCI studies, which input variables are regularly highly correlated and/or high-dimensional, i.e. $m >> T$. The minimum of $\|\mathbf{Y}_k - \mathbf{X}_k\mathbf{B}_k\|_2$ is reached at several points, i.e. the solution is not unique. One solution consists in computing the Minimal Length Least Squares (MLLS) estimator $\hat{\mathbf{B}}_{MLLS} = (\mathbf{X}_k^T\mathbf{X}_k)^\dagger\mathbf{X}_k^T\mathbf{Y}_k$, where $\mathbf{X}^\dagger$ refers to Moore–Penrose pseudoinverse. The MLLS estimator may be unstable when high-dimensional features are considered. Several approximate solutions can be used to deal with high dimensional variables, e.g. penalized training (ridge, LASSO), Principal Component Regression or Partial Least Squares. The estimators which use has been considered for the MSLMs implemented in the present doctoral work are presented in **Appendix B**. The solutions which were finally utilized for efficient training of the implemented MSLMs are exposed in **Chapter 6**.

3.3.1.3 Gate parameter identification

Training of the HMM-based gating network is performed on the data set $\{\mathbf{X}, \mathbf{z}\}$.

(i) Transition matrix The estimate $\hat{\mathbf{A}}$ of the transition matrix is found by counting transition frequencies in the training sequence $(z^t)_{t=1}^T$ [Dietterich, 2009]:

$$\hat{a}_{ij} = \frac{\sum_{t=1}^{T} \omega_{ij}^t}{\sum_{k=1}^{K} \sum_{t=1}^{T} \omega_{ik}^t}. \tag{3.12}$$

(ii) Initial probabilities The initial probability associated with the constrained optimization problem (3.9) is $\hat{\pi}_i = \delta_{z^1,i} \quad \forall i \in \{1, \ldots, K\}$.

(iii) Emission probabilities The MSLM gate either relies on generic emission models, or it exploits discriminative classifiers and Bayes rule for indirect emission modelling (see **section 3.2.1**). The solution to (3.11) depends on the distribution chosen to model $P(\mathbf{x}^t|z^t = i)$. In the case of Gaussian distributions for example, analytical solutions to (3.11) are obtained by applying well-known ML estimators on each data set $\mathbf{X}_k$, e.g. $\mu_k = \frac{1}{\text{card}(t \ s.t. z^t=i)} \sum_{t=1,s.t.z^t=i}^{T} \mathbf{x}^t$ for the mean of a Gaussian distribution. Details on the resolution of (3.11) for the distributions which were here considered for the MSLM implementation are given in **Chapter 6**.

3.3.2 Unsupervised training

Supervised training is performed when the expert sequence $(z^t)_{t=1}^T$ is not hidden in the training data set. This may not always be the case. For example, several experts may be devoted to trajectory encoding during IC states. In that case, no prior information is available about the responsibility of each expert. In this more general case where the expert sequence $(z^t)_{t=1}^T$ is hidden in the training data set, ME training is an unsupervised learning problem [Weigend et al., 1995].

The estimator $\hat{\Theta}_{ML}(\mathbf{X}, \mathbf{Y}) = \underset{\Theta}{\text{argmax}}\ L(\Theta, \mathbf{X}, \mathbf{Y})$, where $L(.)$ refers to the incomplete data log-likelihood, is considered. The use of gradient ascend methods has been reported for direct likelihood maximization of both HMM [Cappé et al., 2006] and ME [Jacobs et al., 1991] [Moerland, 1997] parameters. These methods don't utilize the fact that knowledge on the HMM or ME switching latent variable permits to decouple the parameters' contribution on data likelihood, and therefore involve complex derivation procedures [Moerland, 1997]. On the other hand, data augmentation-based methods, e.g. the well-known Expectation-Maximization (EM) algorithm [Dempster et al., 1977], are particularly suited for parameter estimation in the case of missing data.

The EM algorithm facilitates ML or Maximum A Posteriori (MAP) parameter estimation [Dempster et al., 1977] when the training observations are conditioned on an unobserved (latent, missing) variable, here the switching state z^t. It exploits the fact that the maximization of the complete data log-likelihood

$L_c(\Theta, \mathbf{X}, \mathbf{Y}, \mathbf{z}) = \ln P(\mathbf{X}, \mathbf{Y}, \mathbf{z}|\Theta)$ is simpler than the maximization of $L(\Theta, \mathbf{X}, \mathbf{Y})$ because the impact of each parameter is decoupled. The unknown latent variable vector $\mathbf{z}$, $\mathbf{z} = (z^1, \ldots, z^T)'$ is artificially re-introduced in the maximization problem, and permits to consider the complete data log-likelihood $L_c(\Theta, \mathbf{X}, \mathbf{Y}, \mathbf{z})$. An alternating estimation procedure is used to infer both the parameters Θ and the vector $\mathbf{z}$ from the training data. The two alternating steps are referred to as the E- and M-step, for Expectation- and Maximization-step, respectively.

Application of the EM algorithm is regularly reported for both HMM [Rabiner, 1989] [Cappé et al., 2006] and ME [Yuksel et al., 2012] unsupervised training. An EM-based training was thus derived for the MSLM unsupervised training. An insight on the application of the EM algorithm for MSLM training is given in **section 3.3.2.1**. MSLM-specific E- and M-step are exposed in **sections 3.3.2.2** and **3.3.2.3**, respectively.

3.3.2.1 The Expectation-Maximization algorithm

Let Θ gather the unknown MSLM parameters. Let $\{\mathbf{X}, \mathbf{Y}\}$ be the incomplete training data set available for Maximum-Likelihood identification of the parameters Θ. After initialization of the MSLM parameters, the Expectation-step (E-step) and the Maximization-step (M-step) are performed iteratively. Successive iterations of the E- and M-step lead to the convergence of the data set log-likelihood to a local optimum [Dempster et al., 1977]. Let us consider the i^{th} iteration of the EM algorithm.

E-step The unknown latent variable vector $\mathbf{z}$ depends on the parameter Θ. During the i^{th} E-step, the value of Θ is fixed and is assumed to be equal to its current estimate $\hat{\Theta}^i$. The expectation of the data log likelihood with respect to the unknown latent variable vector $\mathbf{z}$ is computed for this estimate $\hat{\Theta}^i$. Let first define the function $Q : \left\{\mathcal{D}^{\Theta}, \mathcal{D}^{\hat{\Theta}}, \mathbb{R}^m, \mathbb{R}^n\right\} \to \mathbb{R}$ as follows:

$$Q(\Theta^+, \hat{\Theta}, \mathbf{X}, \mathbf{Y}) = E\left[L_c(\Theta^+, \mathbf{X}, \mathbf{Y}, \mathbf{z})|\Theta = \hat{\Theta}, \mathbf{X}, \mathbf{Y}\right]. \tag{3.13}$$

The E-step considers the function Q evaluated at $\hat{\Theta}^i$ and $\mathbf{X}, \mathbf{Y}$:

$$\Theta^+ \mapsto Q(\Theta^+, \hat{\Theta}^i, \mathbf{X}, \mathbf{Y}) = E\left[L_c(\Theta^+, \mathbf{X}, \mathbf{Y}, \mathbf{z})|\Theta = \hat{\Theta}^i, \mathbf{X}, \mathbf{Y}\right], \tag{3.14}$$

i.e. the expectation of $L_c(\Theta, \mathbf{X}, \mathbf{Y}, \mathbf{z})$ with respect to $\mathbf{z}$, given the current parameter estimate $\hat{\Theta}^i$:

$$E\left[L_c(\Theta^+, \mathbf{X}, \mathbf{Y}, \mathbf{z})|\Theta = \hat{\Theta}^i, \mathbf{X}, \mathbf{Y}\right] = \sum_{\mathbf{z}} L_c(\Theta^+, \mathbf{X}, \mathbf{Y}, \mathbf{z})\ P(\mathbf{z}|\Theta = \hat{\Theta}^i, \mathbf{X}, \mathbf{Y}) \tag{3.15}$$

M-step The M-step consists of maximizing $Q(\Theta^+, \hat{\Theta}^i, \mathbf{X}, \mathbf{Y})$ with respect to Θ^+, i.e. of optimizing the parameter value when the latent state is replaced by its expectation. The following updated ML estimate $\hat{\Theta}^{i+1}$ is obtained:

$$\hat{\Theta}^{i+1} = \underset{\Theta^+}{\operatorname{argmax}}\ Q(\Theta^+, \hat{\Theta}^i, \mathbf{X}, \mathbf{Y}) \tag{3.16}$$

Depending on the function Q, analytical solutions or iterative approaches (gradient methods) are used to compute $Q(\Theta^+, \hat{\Theta}^i, \mathbf{X}, \mathbf{Y})$ with respect to Θ^+. The utilization of iterative maximization approaches results in the presence of an "inner-loop" within each M-step, and significantly increases the computational cost of the EM algorithm.

The E- and M-steps were here derived for MSLM unsupervised training.

3.3.2.2 E-step for MSLMs

Execution of the MSLM-specific E-step consists of expressing the function $Q(\Theta^+, \hat{\Theta}, \mathbf{X}, \mathbf{Y}) = E\left[L_c(\Theta^+, \mathbf{X}, \mathbf{Y}, \mathbf{Z})|\Theta = \hat{\Theta}, \mathbf{X}, \mathbf{Y}\right]$ (see equation (3.13)). Taking the expectation of the complete data log-likelihood presented in equation (3.7), and thus exploiting the hypotheses presented for the MSLM supervised training, we get:

$$\begin{aligned} E\left[L_c(\Theta^+, \mathbf{X}, \mathbf{Y}, \mathbf{z})|\Theta = \hat{\Theta}, \mathbf{X}, \mathbf{Y}\right] &= \sum_{i=1}^{K} \gamma_i^1 \ln(\pi_i^+) + \sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{t=1}^{T-1} \xi_{i,j}^t \ln(a_{ij}^+) \\ &+ \sum_{i=1}^{K}\sum_{t=1}^{T} \gamma_i^t \ln P(\mathbf{y}^t|\mathbf{x}^t, \mathbf{B}_i^+) + \sum_{i=1}^{K}\sum_{t=1}^{T} \gamma_i^t \ln P(\mathbf{x}^t|b_i^+), \end{aligned} \tag{3.17}$$

where

$$\gamma_i^1 = E\left[\delta_{z^1,i}|\Theta = \hat{\Theta}, \mathbf{X}, \mathbf{Y}\right] = P(z^1 = i|\Theta = \hat{\Theta}, \mathbf{x}^{1:T}, \mathbf{y}^{1:T}), \tag{3.18}$$

$$\gamma_i^t = E\left[\delta_{z^t,i}|\Theta = \hat{\Theta}, \mathbf{X}, \mathbf{Y}\right] = P(z^t = i|\Theta = \hat{\Theta}, \mathbf{x}^{1:T}, \mathbf{y}^{1:T}), \tag{3.19}$$

$$\xi_{i,j}^t = E\left[\omega_{i,j}^t|\Theta = \hat{\Theta}, \mathbf{X}, \mathbf{Y}\right] = P(z^t = i, z^{t+1} = j|\Theta = \hat{\Theta}, \mathbf{x}^{1:T}, \mathbf{y}^{1:T}). \tag{3.20}$$

Estimation of the state and transition probabilities The state γ_i^t and transition probabilities $\xi_{i,j}^t$ can be expressed as follows (see derivation of the formula in **Appendix A**):

$$\begin{aligned} \gamma_i^t &= P(z^t = i|\mathbf{x}^{1:T}, \mathbf{y}^{1:T}, \Theta) \\ &= \frac{P(\mathbf{y}^{t+1:T}, \mathbf{x}^{t+1:T}|z^t = i, \Theta) P(z^t = i, \mathbf{x}^{1:t}, \mathbf{y}^{1:t}|\Theta)}{P(\mathbf{x}^{1:T}, \mathbf{y}^{1:T}|\Theta)}, \end{aligned}$$

and

$$\begin{aligned}\xi_{i,j}^t &= P(z^t = i, z^{t+1} = j|\mathbf{x}^{1:T}, \mathbf{y}^{1:T}, \Theta) \\ &= \frac{P(\mathbf{x}^{t+2:T}, \mathbf{y}^{t+2:T}|z^{t+1} = j, \Theta)P(\mathbf{x}^{t+1}, \mathbf{y}^{t+1}|z^{t+1} = j, \Theta)}{P(\mathbf{x}^{1:T}, \mathbf{y}^{1:T}|\Theta)} \\ &\quad \times P(z^{t+1} = j|z^t = i, \Theta)P(z^t = i, \mathbf{x}^{1:t}, \mathbf{y}^{1:t}|\Theta),\end{aligned}$$

where

$$\begin{aligned}&P(z^{t+1} = j|z^t = i, \Theta) = a_{ij}, \\ &P(\mathbf{x}^t, \mathbf{y}^t|z^t = j, \Theta) = P(\mathbf{y}^t|\mathbf{x}^t, \mathbf{B}_j)P(\mathbf{x}^t|b_j).\end{aligned}$$

Finally, the intermediary probabilities $P(z^t = k, \mathbf{x}^{1:t}, \mathbf{y}^{1:t}|\Theta)$, $P(\mathbf{x}^{t+1:T}, \mathbf{y}^{t+1:T}|z^t = j, \Theta)$ and $P(\mathbf{x}^{1:T}, \mathbf{y}^{1:T}|\Theta)$ are computed using extensions of the forward and backward algorithms originally developed for HMMs [Rabiner, 1989]. The HMM forward and backward algorithms permit to compute the probabilities $P(z^t = i, \mathbf{x}^{1:t}|\Theta)$ and $P(\mathbf{x}^{t+1:T}|z^t = i, \Theta)$ (see their derivation in [Bengio and Frasconi, 1996]) . MSLM-specific forward and backward algorithms are derived in the present doctoral work. They extend the HMM-specific ones by taking into account the likelihood of each continuous expert when estimating the probability associated with each possible hidden state.

MSLM forward algorithm The MSLM forward algorithm permits to compute the probabilities $P(z^t = k, \mathbf{x}^{1:t}, \mathbf{y}^{1:t}|\Theta), t = 1, \ldots, T$. The recurrence is initialized by $P(z^1 = j, \mathbf{x}^1, \mathbf{y}^1|\Theta) = P(\mathbf{y}^1|\mathbf{x}^1, z^1 = j, \Theta)P(\mathbf{x}^1|z^1 = j, \Theta)P(z^1 = j, \Theta)$. The following recurrence formula is then applied (see derivation of the formula in **Appendix A**):

$$\begin{aligned}P(z^{t+1} = j, \mathbf{x}^{1:t+1}, \mathbf{y}^{1:t+1}|\Theta) &= P(z^{t+1} = j, \mathbf{x}^{1:t}, \mathbf{y}^{1:t}, \mathbf{x}^{t+1}, \mathbf{y}^{t+1}|\Theta) \\ &= P(\mathbf{x}^{t+1}, \mathbf{y}^{t+1}|z^{t+1} = j, \Theta) \\ &\quad \times \sum_i P(z^{t+1} = j|z^t = i, \Theta)P(z^t = i, \mathbf{x}^{1:t}, \mathbf{y}^{1:t}|\Theta),\end{aligned}$$

where $P(\mathbf{x}^t, \mathbf{y}^t|z^t = j, \Theta) = P(\mathbf{y}^t|\mathbf{x}^t, \mathbf{B}_j)P(\mathbf{x}^t|b_j)$. Finally, a termination step yields $P(\mathbf{x}^{1:T}, \mathbf{y}^{1:T}|\Theta) = \sum_{i=1}^{N} P(z^T = i, \mathbf{x}^{1:T}, \mathbf{y}^{1:T}|\Theta)$.

MSLM backward algorithm Similarly, a modified backward algorithm permits to compute $P(\mathbf{x}^{t+1:T}, \mathbf{y}^{t+1:T}|z^t = k, \Theta)$. Following the generic HMM-specific backward algorithm [Fink, 2014], the recurrence is initialized with $P(\mathbf{x}^T, \mathbf{y}^T|z^T = j, \Theta) = (1, \ldots, 1)'$. The following iterations are then completed (see derivation of the formula in **Appendix A**):

$$P(\mathbf{x}^{t+1:T}, \mathbf{y}^{t+1:T}|z^t = i, \Theta) = \sum_j P(\mathbf{x}^{t+1}, \mathbf{y}^{t+1}|z^{t+1} = j\Theta) \times P(\mathbf{x}^{t+2:T}, \mathbf{y}^{t+2:T}|z^{t+1} = j, \Theta)P(z^{t+1} = j|z^t = i, \Theta).$$

Again, the likelihood of each continuous expert is taken into account during state estimation.

3.3.2.3 M-step for MSLMs

The introduction of the latent switching variable z^t permits to decouple parameters in (A.12). Optimization problems similar to the ones associated with supervised MSLM training are obtained, namely:

$$\underset{\mathbf{B}_i}{\text{maximize}} \quad \sum_{t=1}^{T} \gamma_i^t \ln P(\mathbf{y}^t|\mathbf{x}^t, \mathbf{B}_i) \qquad i = 1, \ldots, K \tag{3.21}$$

$$\begin{aligned} \underset{\pi}{\text{maximize}} \quad & \sum_{j=1}^{K} \gamma_j^1 \ln \pi_j \\ \text{subject to} \quad & \sum_{j=1}^{K} \pi_j = 1 \end{aligned} \tag{3.22}$$

$$\begin{aligned} \underset{\mathbf{A}}{\text{maximize}} \quad & \sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{t=1}^{T-1} \xi_{i,j}^t \ln a_{ij} \\ \text{subject to} \quad & \sum_{j=1}^{K} a_{ij} = 1 \quad \forall i \in \{1, \ldots, K\}. \end{aligned} \tag{3.23}$$

$$\underset{\mathbf{b_i}}{\text{maximize}} \quad \sum_{t=1}^{T} \gamma_i^t \ln P(\mathbf{x}^t|b_i) \qquad i = 1, \ldots, K. \tag{3.24}$$

where (3.21), (3.22), (3.23), and (3.24) correspond to ML estimation of the expert, initial state, transition and emission parameters, respectively. In contrast with the MSLM supervised maximization problems, the impact of observation t on a parameter set associated with state i is weighted by the corresponding state probability $\gamma_j^t = P(z^t = i|\mathbf{x}^{1:T}, \mathbf{y}^{1:T}, \Theta)$.

Experts Identification of the experts' parameters consists in maximizing (3.21). Specific training procedures are required because training samples are weighted by state probabilities rather than pooled by state labels. Again, they depend on the noise distribution. The approximate solution used to solve this maximization problem for the MSLMs implemented in the present doctoral work (case of Gaussian noise with high-dimensional features) is presented in **Chapter 6**.

Gate Similarly, weighted training is required to identify the transition, initial state and emission probabilities.

(i) Transition matrix The application of Lagrange Multipliers permits to obtain the updated estimate of the transition matrix [Bilmes, 1998]:

$$\hat{a}_{ij} = \frac{\sum_{t=1}^{T} \xi_{ij}^{t}}{\sum_{k=1}^{K} \sum_{t=1}^{T} \xi_{ik}^{t}} \qquad i = 1, \ldots, K, j = 1, \ldots, K. \tag{3.25}$$

Intuitively, this expression corresponds to a weighted frequency of each possible transition.

(ii) Initial state probability The constrained optimization problem (3.22) is again solved by applying Lagrange Multipliers, which yield [Bilmes, 1998]:

$$\pi_{z^1} = \gamma_i^1 \qquad i = 1, \ldots, K.$$

(iii) Emission probability The solution of the maximization problem (3.24) depends on the distribution used to model feature emission. The solution is for example well known in the case of Gaussian emission distributions (i.e., weighted Bayes classifier) [Rabiner, 1989]. Implementation details for the MSLMs considered in the present doctoral work are given in **Chapter 6**.

3.4 Conclusion

In this chapter, a decoder has been introduced for the task of accurate, asynchronous mono- and multi-limb ECoG kinematic decoding. This decoder, referred to as Markov Switching Linear Model (MSLM), presents three key features expected to be associated with efficient asynchronous kinematic decoding. First, it has been developed as a switching model. Thus, as opposed to post-processed models, the MSLM combines several models. It is therefore able to adjust to task-dependent changes in the model between ECoG signals and limb kinematics. This property is expected to be useful to both introduce No-Control (NC) support into kinematic decoders and to improve Intentional Control (IC) decoding accuracy during complex movements. If one active limb only is associated with each model, the MSLM intrinsically limits parallel limb activations. Second, it has been designed as a discriminative rather than generative switching model. The discriminative framework is expected to facilitate state and kinematic extraction from high-dimensional neural features. The Mixture of Experts (ME) framework has been used to describe the discriminative switching models reportedly applied for the integration of NC support into synchronous ECoG decoders. Finally, in contrast with the discriminative switching models proposed in the literature, the MSLM performs dynamic state estimation. Dynamic state detection is liable to improve state detection, and thus to further reduce spurious activations during NC periods. It relies on an extension of

the ME framework to the case of dynamical state detection. A supervised approach has been proposed for the training of MSLMs. It is applicable when the values taken by the switching variable are known in the training data set. An EM-based training algorithm, which can be applied when the switching variable is hidden in the training data set, has been derived for unsupervised training of the MSLM. The implementation of the supervised and unsupervised MSLM training approaches depends on the distributions chosen to model the neural feature distribution and the experts' measurement noise. Implementation details for the MSLMs considered in the present doctoral work are presented in **Chapter 6**.

Chapter 4

Data set presentation

Contents

The performance of the proposed Markov Switching Linear Model (MSLM) was evaluated on two publicly available ECoG data sets, namely a preclinical and a multi-limb (multi-fingers) clinical data set. These data sets permitted to complete a first offline investigation on the ability of the MSLM to perform accurate asynchronous multi-DoF (Degrees of Freedom) or multi-limb ECoG decoding, and paved the way for the MSLM closed-loop evaluation which will be accomplished during CLINATEC's forthcoming clinical trial.

4.1 Asynchronous preclinical data set

The preclinical data set was composed of subdural and epidural ECoG data sets acquired and distributed by the Laboratory for Adaptive Intelligence, RIKEN Brain Science Institut, Saitama, Japan. The data sets are publicly available at `http://neurotycho.org/`. Full description of the experimental set-ups can be found in [Chao et al., 2010] and [Shimoda et al., 2012]. The cortical activity of 4 Non-Human Primates (Monkeys A, K, B and C) was recorded by chronic ECoG arrays during a food reaching task.

4.1.1 Behavioural task

The Non-Human Primates (NHPs) were sitting in front of an experimenter during the data acquisition sessions. The experimenter presented food to the NHPs at random intervals. The NHPs performed reaching movements to grab the food and to

bring it to their mouth (**Figure 4.1**). An example of reaching trajectory is shown in **Figure 4.2**.

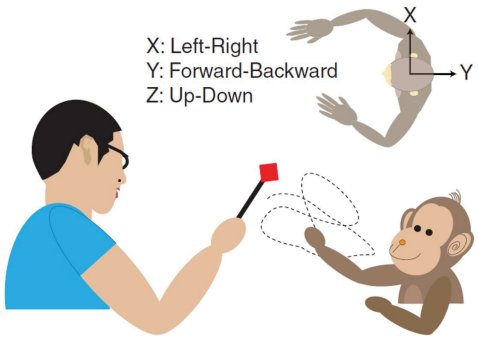

Figure 4.1: Behavioral task (reproduced from [Chao et al., 2010]). The NHPs performed reaching movements to grab food presented at random intervals.

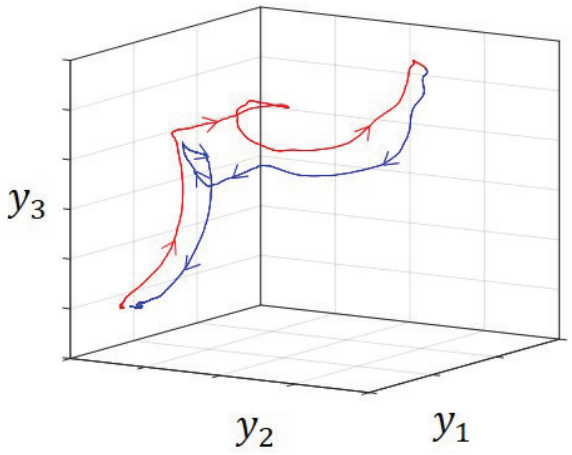

Figure 4.2: Reaching movement example. After having grabbed the food presented by the experimenter and put it in its mouth, the NHP brought its arm back to an idle position.

4.1.2 Signal acquisition

Monkeys A and K were implanted with a 32- and 64- channel subdural ECoG array, respectively (**Figure 4.3**). Both Monkeys B and C were epidurally implanted with a 64-channel ECoG array (**Figure 4.3**). The customized arrays embedded 2.1mm-diameter electrodes with 3.5mm inter-electrode distances.

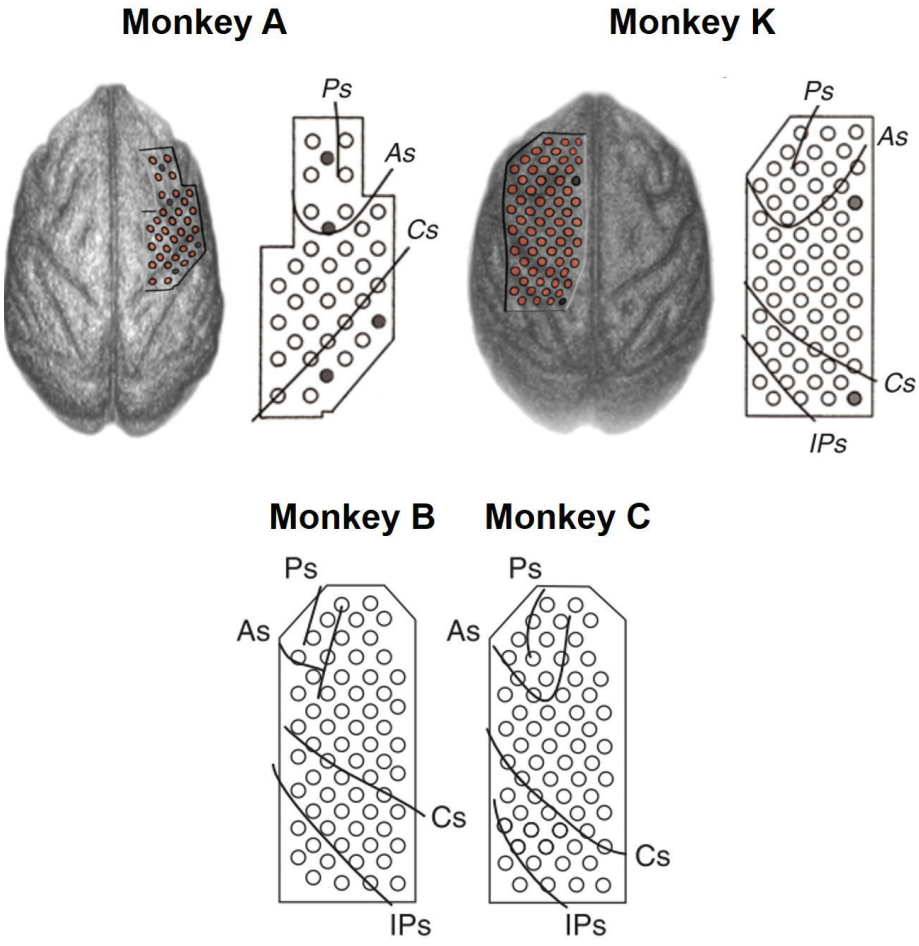

Figure 4.3: ECoG array implantation for Monkeys A, K, B and C (reproduced from [Chao et al., 2010] [Shimoda et al., 2012]). Solid grey circles indicate reference electrodes. Ps: principal sulcus; As: arcuate sulcus; Cs: central sulcus; IPs: intraparietal sulcus.

Table 4.1: State sequence characteristics, preclinical data set. The intra-session mean number of movements and mean duration of NC and IC periods was averaged over 5, 3, 6 and 10 sessions for Monkeys A, K, B and C, respectively.

Data set	NHP	Movement number	NC period duration (s)	IC period duration (s)
subdural	*A*	104 ± 24	5.2 ± 1.7	5.7 ± 2.0
	K	69 ± 8	5.7 ± 2.0	3.6 ± 0.1
epidural	*B*	70.3 ± 14.3	7.7 ± 3.1	7.5 ± 1.2
	C	127.4 ± 20.6	4.4 ± 0.9	3.7 ± 0.5

ECoG signals were acquired at a sampling rate of 1kHz. A motion tracking system tracked the subjects' wrist coordinates at a sampling frequency of 120Hz. Body-centred 3D wrist trajectories downsampled at 20Hz were extracted from the outputs of the motion tracking system.

4.1.3 Data set characteristics

4 sessions were excluded from the epidural data set because of difficulties to assess NC/IC labels. The final data set was thus composed of 8 subdural sessions and 16 epidural sessions. The corresponding tracked trajectories exhibit essential differences in terms of temporal characteristics of NC/IC state succession (i), variability of the wrist position during relaxed states (ii) and trajectory complexity (iii) (see trajectory examples in **Figure 4.4**).

(i) Duration of NC and IC periods Manual labelling of NC and IC states was used to compute the average duration of NC and IC periods and the number of reaching movement executed during each acquisition session (**Table 4.1**). NC and IC classes were found to be relatively well balanced for most sessions.

(ii) Neutral position variability The variability of monkeys' wrist position during NC periods was established for each axis. Results are displayed in **Table 4.2**. The y_1- and y_2-axes are associated with unstable NC wrist positions, i.e. NHPs relaxed their wrist on variable points on the horizontal plan. Wrist positions along the y_3-axis (vertical), on the other hand, are relatively stable during NC periods.

(iii) Trajectory complexity The complexity of the IC trajectories was measured as followed. First, the trajectory derivative was computed by convolving it with the derivative of a Gaussian kernel. The derivatives' zero-crossings were then detected (see **Figure 4.5**). Their average number was obtained by dividing the number of detected zero-crossings by the number of IC movements. 3 sessions from the subdural data set were found to have an average complexity superior to 2.

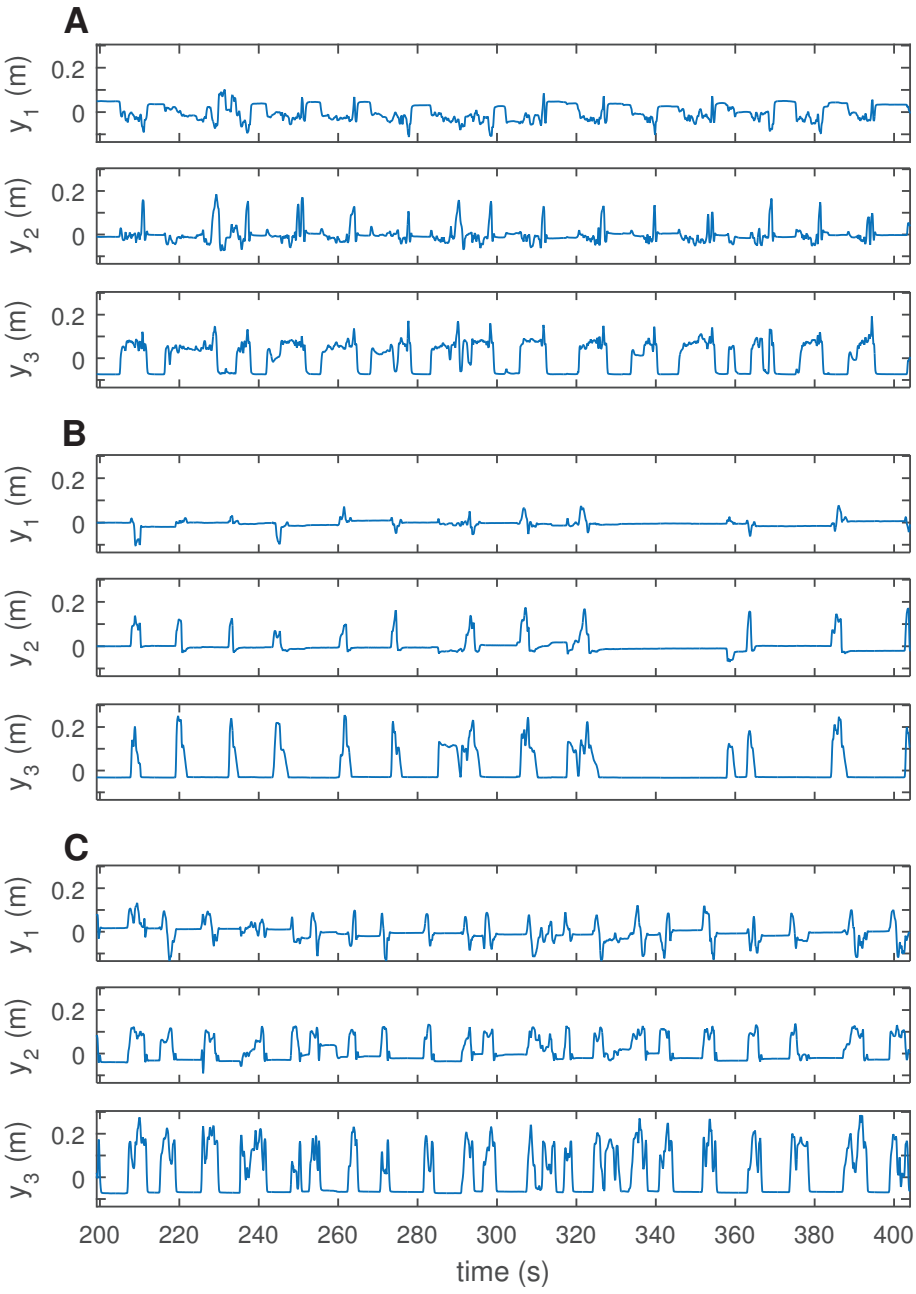

Figure 4.4: Trajectory examples. Trajectories along each axis were centred around 0. An identical scale was used for all NHPs and all axes. **A.** Monkey A, session 1. **B.** Monkey K, session 2. **C.** Monkey C, session 6.

Table 4.2: NC position variability during NC states. The intra-session standard deviation of the wrist position was averaged over 5, 3, 6 and 10 sessions for Monkey A, K, B and C, respectively.

Data set	NHP	y_1-axis (mm)	y_2-axis (mm)	y_3-axis (mm)
subdural	*A*	25.9 ± 12.3	29.6 ± 14.2	8.9 ± 2.8
	K	15.8 ± 3.1	21.4 ± 10.9	2.8 ± 0.6
epidural	*B*	12.2 ± 4.6	16.3 ± 7.5	28.2 ± 10.6
	C	38.7 ± 4.2	36.4 ± 8.7	11.1 ± 4.5

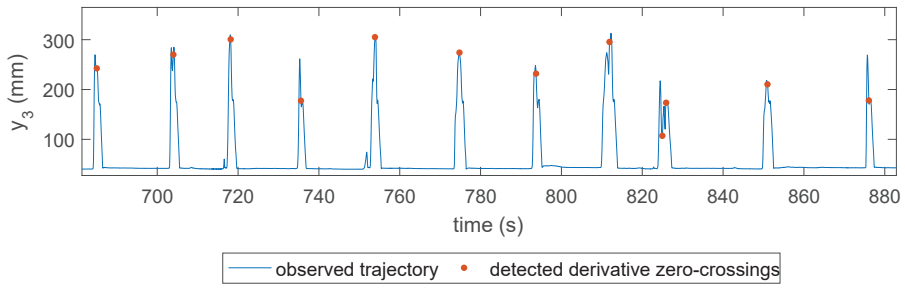

Figure 4.5: Measuring the movement complexity, example (subdural data set, Monkey K). Red dots indicate detected zero-crossings of the trajectory's smoothed derivative.

The preclinical data set was analysed session per session (24 sessions, 4 monkeys). Each recording was split into training and test set: the first 70% of each session were used for model training and validation, and the last 30% were used for testing.

4.2 Asynchronous multi-limb (multi-finger) clinical data set

The ECoG clinical data set was acquired while human subjects were executing sequential finger movements. It gathers data made publicly available for the IVth BCI competition [Schalk et al., 2007] [Miller and Schalk, 2008] [Tangermann et al., 2012] (`http://www.bbci.de/competition/iv/`) and data distributed by Stanford University (`https://purl.stanford.edu/zk881ps0522`). The BCI competition IV data set comprises data measured in three subjects. The Stanford data set is composed of data acquired in nine subjects. All patients participated in a purely voluntary manner, after providing written consent, under experimental protocols approved by the Institutional Review Board of the University of Washington (#12193). All patient data was anonymized according to IRB protocol, in accordance with HIPAA mandate. These data originally appeared in the manuscript "Human Motor Cortical Activity Is Selectively Phase-Entrained on Underlying Rhythms" published in PLoS Computational Biology in 2012 [Miller et al., 2012]. As a result, the full data set was composed of data acquired from 12 different subjects. A similar protocol was utilized to acquire the IVth BCI competition and the Stanford data sets.

4.2.1 Behavioural task

All subjects participated in one 10min-long acquisition session. Subjects were instructed to perform finger extensions and flexions in response to visual cues. Each cue was presented during 2 seconds, and indicated which finger had to be moved (sequential movements). The subjects were asked to execute successive extensions and flexions of the requested finger as long as the cue was on display. A blank screen was displayed for 2 seconds after disappearance of the visual cue. Subjects were instructed to relax during this inter-trial period. 30 cues were devoted to each finger. The resulting 150 cues were randomly presented to the subjects.

4.2.2 Data acquisition

The 12 subjects were epileptic patients undergoing an ECoG-array implantation for the localization of their epileptic focus. ECoG arrays (Ad-Tech, Racine, WI) were implanted under the subjects' dura mater (subdural implantation) [Miller and Schalk, 2008] [Miller et al., 2012]. They gathered 38 to 64 4mm-diameter electrodes spaced out with a 1cm-inter-electrode distance. The location of the ECoG array was chosen with respect to the monitoring purpose of the implantation. No information has been given as to the precise localization of the arrays implanted into the subjects of

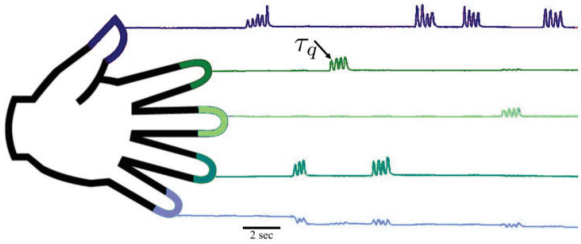

Figure 4.6: Behavioral task (reproduced from [Miller and Schalk, 2008]).

the BCI competition IV data set [Miller and Schalk, 2008]. The arrays of Stanford's data base covered various brain areas, namely frontal, parietal, occipital and/or temporal areas (see details in the data base description file), i.e. both motor and non-motor activity was measured by the electrodes. ECoG signals were acquired at a sampling rate of 1kHz, and band-pass filtered between 0.15 and 200Hz.

Subjects were asked to move the fingers contralateral to the array location. A data glove tracked the finger coordinates at a 25Hz sampling frequency. The tracked coordinates were normalized and centred so that each finger trajectory exhibited similar spatial amplitudes within an acquisition session. A trajectory example is presented in **Figure 4.7**. Fingers are ordered as follows: thumb, index finger, middle finger, ring finger and little finger.

4.2.3 Data set characteristics

One session of the Stanford data set was discarded because of the absence of NC states (subject "jp"), and two more sessions were excluded because they were shorter and therefore unsuited for the training and test of complex models (subjects "mv" and "wm"). Details on the number of channels available in the remaining data sets are shown in **Table 4.3**. The 9 remaining finger trajectories were analysed from two points of view: duration of the NC and IC states (i) and stability of finger position during NC states (ii).

(i) Duration of NC and IC$_i$ periods Both IC and NC states were theoretically 2s-long, as IC cues were displayed for 2 seconds and were followed by 2s-long blank screens. Manual segmentation of the finger trajectories was used to measure the empirical duration of NC and IC states (**Table 4.4**). NC and IC classes, where the IC class is composed of the IC$_i$ class, $i = 1, \ldots, 5$, were relatively well balanced.

(ii) Neutral position variability The stability of the NC positions was measured for each finger i by averaging finger positions during NC periods and during IC$_j$, $j \neq i, j \in \{1, \ldots, 5\}$ periods (**Table 4.5**).

Similarly to the preclinical data set, the clinical data set was analysed session

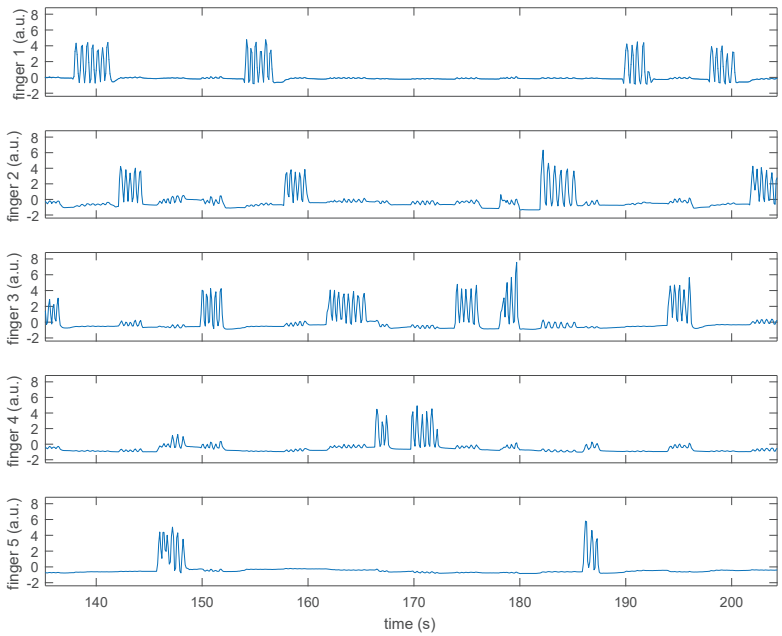

Figure 4.7: Finger trajectories example (subject 1). The trajectory of each finger was centred and normalized. Movement amplitudes are thus indicated with an arbitrary unit (a.u.). Finger 1: thumb, Finger 2: index finger, Finger 3: middle finger, Finger 4: ring finger, Finger 5: little finger.

Table 4.3: Number of ECoG electrodes, clinical data set.

Data source	Subjects	Electrode number
BCI competition IV	1	62
	2	48
	3	64
Stanford data set	bp	46
	cc	63
	ht	64
	jc	47
	wc	64
	zt	61

Table 4.4: State sequence characteristics, clinical data set. The number of movements and the intra-session average duration of NC and IC periods was measured for each session of the clinical data set.

Data set	Subject	NC period duration (s)	IC period duration (s)	Movement number
BCI IV	1	1.10 ± 0.66	2.95 ± 0.94	148
	2	1.85 ± 0.62	2.22 ± 0.73	147
	3	1.12 ± 0.50	3.26 ± 2.42	137
Stanford	bp	1.93 ± 0.78	2.21 ± 0.64	147
	cc	1.32 ± 0.82	2.79 ± 1.05	148
	ht	1.60 ± 0.79	2.62 ± 0.86	144
	jc	1.74 ± 0.67	2.27 ± 0.31	132
	wc	1.29 ± 0.94	2.85 ± 1.40	147
	zt	1.21 ± 0.87	2.93 ± 0.81	147

Table 4.5: Variability of normalized finger i position during NC and $\text{IC}_j, j \neq i, j \in \{1, \ldots, 5\}$ states, clinical data set. The intra-session standard deviation of the finger positions was measured for each session of the clinical data set. Deviations are indicated in arbitrary units because normalized finger trajectories are considered.

Data set	Subject	Finger 1	Finger 2	Finger 3	Finger 4	Finger 5
BCI IV	1	0.22	0.43	0.33	0.47	0.33
	2	0.43	0.54	0.36	0.59	0.49
	3	0.35	0.58	0.70	0.75	0.33
Stanford	bp	0.44	0.54	0.34	0.57	0.49
	cc	0.33	0.58	0.72	0.76	0.34
	ht	0.33	0.56	0.41	0.49	0.56
	jc	0.19	0.27	0.17	0.42	0.43
	wc	0.28	0.39	0.33	0.31	0.67
	zt	0.21	0.43	0.36	0.48	0.34

per session (9 subjects, one session per subject). Likewise, each recording was split into training and test set (70% and 30% of each session, respectively).

4.3 Conclusion

The data sets used to assess the ability of the Markov Switching Linear Model (MSLM) to perform accurate asynchronous multi-DoF or multi-limb ECoG decoding were presented in the present chapter. Both data sets are publicly available. The first data set is composed by preclinical ECoG data acquired while monkeys were executing asynchronous unimanual reaching movements. The second data set gathers clinical ECoG data collected while human subjects were performing sequential multi-fingers flexions and extensions. Next chapter consists of a presentation of the methodology used to measure the MSLM performance when applied on the two considered data sets.

Chapter 5

Evaluation and comparison

Contents

The Markov Switching Linear Model (MSLM) was compared to a post-processed decoder and to a switching state-space filter. These two decoders were chosen so as to represent two alternative strategies previously exploited for asynchronous and/or multi-limb kinematic decoding. Performance indicators specific to discrete and continuous variables were selected to measure the respective ability of each decoder to handle (possibly multi-limb) asynchronous and accurate decoding, respectively. A statistical procedure was chosen to assess the statistical significance of the observed performance differences.

5.1 Comparative approaches for supervised MSLMs

The performance of the MSLM was first compared to the decoding performance of a Markovian post-processed Wiener Filter (MpWF). This comparison permitted to assess the interest of the first key feature of the MSLM, namely switching modelling, when compared to the post-processing approach utilized in earlier studies. It was then compared to another switching model, namely the Switching Kalman Filter (SKF). This comparison allowed for the evaluation of the second key feature of the MSLM, i.e. the discriminative framework, as opposed to the generative framework utilized in a few BCI studies. Both the MpWF and the SKF performed dynamic state detection. Variants of the MSLM and of the MpWF with static detection were finally

considered to evaluate the interest of the MSLM's Markovian hypothesis. They are referred to as SLM (for Switching Linear Model) and pWF (for post-processed Wiener Filter), respectively.

5.1.1 Markovian post-processed Wiener Filter

MSLM decoders combine several models to reckon with context-dependent modifications in the dependence between trajectory kinematics and neural features.

By contrast, the use of a single continuous decoder has been reported for asynchronous [Wang et al., 2013b] [Velliste et al., 2014] and/or multilimb [Wang et al., 2011] decoding in earlier BCI studies. The output of the single decoder was post-processed so as to integrate NC and/or sequential multi-limb NC support into regression models [Wang et al., 2011] or into state-space filters [Wang et al., 2013b] [Velliste et al., 2014].

Formally, a first static or dynamic model $f : \mathbb{R}^m \to \mathbb{R}^n$ output an intermediary estimate $\hat{\mathbf{y}}^t_{tmp} = f(\mathbf{x}^t)$. A post-processing model $v : \mathbb{R}^n \to \mathbb{R}^n$ was then applied on $\hat{\mathbf{y}}^t_{tmp}$, and yielded the final estimate $\hat{\mathbf{y}}^t = v(\hat{\mathbf{y}}^t_{tmp})$ (**Figure 5.1**). In some studies, the post-processing operator v was additionally fed with the input signal (see **Figure 5.1**, blue arrows (ii)), e.g. in [Wang et al., 2013b] [Velliste et al., 2014]. State detection was therefore performed using neural signals rather than intermediary position or velocity estimates.

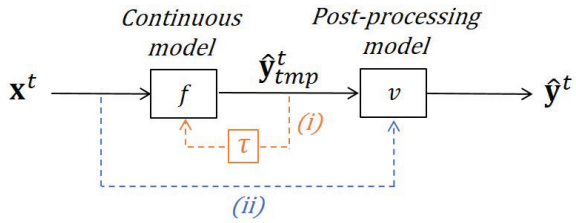

Figure 5.1: Post-processed kinematic decoder, general data flow. A regression model or a state-space filter (additional orange arrows (i)) is fed with the neural features, and, in the case of state-space filters, with statistical characteristics of the previous estimate. A post-processing operator is applied on the output of this decoder.

The respective efficiency of MSLMs and of post-processed models depends on the existence or absence of context-driven modifications of the dependence between neural signals and kinematic parameters. If such modifications do exist, then the training of a single continuous model may yield a suboptimal decoder. Kinematic decoding accuracy may be degraded, for example during IC states. Moreover, if the post-processing operator is not conditioned on neural signals, i.e. if it is exclusively fed with the intermediary trajectory estimates $\hat{\mathbf{y}}^t_{tmp}$, then state detection accuracy depends on the quality of the trajectory estimates. Consequently, state detection is also liable to be corrupted by the application of a suboptimal regression model.

The respective relevance of switching and post-processed models for asynchronous and multi-limb ECoG decoding is unclear. The comparison between switching models and post-processed models has rarely been drawn. In [Williams et al., 2013], a simple post-processing operator fed with velocity estimates and a switching model which gate was fed with ECoG neural signals were compared for the integration of NC support into the kinematic estimates issued by a linear regression model. The authors concluded to the superiority of the second decoding strategy. This study suggested the interest of switching models. The use of post-processing strategies has nonetheless been reported for asynchronous [Wang et al., 2013b] and/or multi-limb [Wang et al., 2011] ECoG decoding in several BCI studies.

To clarify the relevance of switching modelling, a Markovian post-processed Wiener Filter (MpWF) was implemented for comparison purposes. The performance of the MpWF was evaluated for two decoding tasks, namely mono-limb and multi-limb (multi-finger) asynchronous trajectory reconstruction. The continuous model was fitted on both NC and (potentially multiple) IC samples, and the post-processing operator was exclusively fed with kinematic estimates (see **Figure 5.2**). This configuration was chosen to investigate the interest of the MSLM switching hypothesis for both kinematic decoding during (possibly multiple) IC states and for state detection. Similarly to the MSLM which embeds a dynamic gate, a Markovian hypothesis was used by the post-processing operator, i.e. the post-processed operator performed dynamic state detection from the kinematic estimates yielded by the single continuous model. This Markovian post-processing strategy permitted to discard the effect of dynamic or static state detection when investigating the comparative efficiency of the post-processing and switching approaches. The comparative quality of dynamic and static post-processed-based state detection was nevertheless established using a variant of the considered post-processed WF with static rather than dynamic state detection (see **section 5.1.3**).

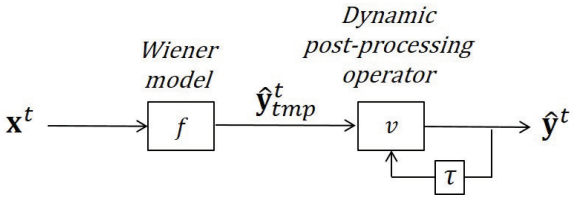

Figure 5.2: Markovian post-processed Wiener Filter decoder, data flow. The post-processed operator performs dynamic state detection from the kinematic estimates yielded by a single Wiener filter.

5.1.2 Switching Kalman Filter

The second competing decoder was chosen to assess whether switching regression models outperform or are outperformed by switching state-space filters. The MSLM

combines regression models. Regression models are based on a discriminative-like interpretation of the relationship between neural signals and kinematic parameters, i.e. they model the probabilities $P(\mathbf{y}^t|\mathbf{x}^t, z^t = i, \Theta)$. By contrast, the switching state-space models considered in BCI studies are generative, i.e. they model the probability $P(\mathbf{x}^t, \mathbf{y}^t, z^t = i, \Theta)$ [Wu et al., 2004] [Srinivasan et al., 2007].

An analogue of the MSLM in the framework of switching state-space models was here considered, namely the Switching Kalman Filter (SKF). The SKF (see **Appendix C**) has been applied in several BCI studies [Wu et al., 2004] [Wu et al., 2003a], and its use has been advocated for the integration of NC support into state-space models [Srinivasan et al., 2007]. Similarly to the MSLM, it relies on the assumption that its observation and transition models are conditioned on a Markovian hidden state [Murphy, 1998]. Trajectory estimation is based on a probabilistic combination of state-specific models [Murphy, 1998]. Because of this hypothesis about neural signal generation, the value of the mixing coefficients depends indirectly on the observed neural signals. In contrast with the MSLM, SKF state estimation is based on the consistency of both its emission and transition models with the observed neural signals (see **Figure 5.3** and **section C.2.3** in **Appendix C**).

5.1.3 Static decoders

The interest of the Markovian hypothesis used to perform dynamic state detection was investigated by comparing the MSLM and the MpWF with the SLM and pWF, i.e. variants of the MSLM and MpWF with static rather than dynamic state detection. The corresponding data flow are shown in **Figures 5.4** and **5.5**.

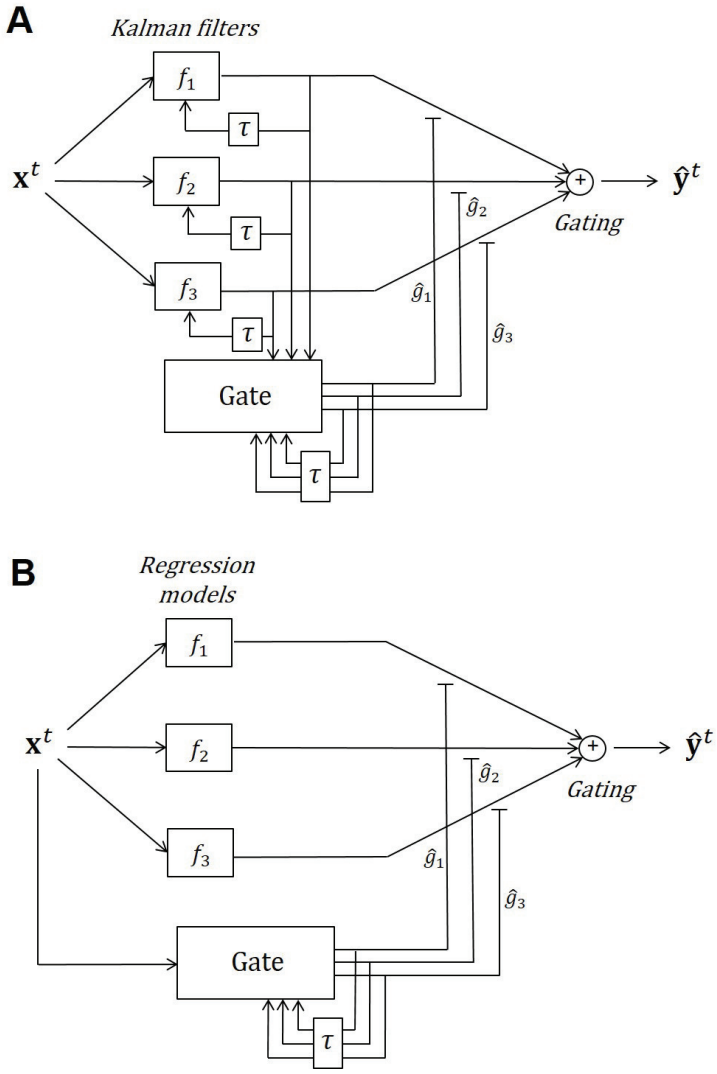

Figure 5.3: SKF (**A**) and MSLM (**B**) data flow. **A.** The weights used to combine the KF experts depend on the adequacy of each KF with the current and past observed neural signals. **B.** The weights used to combine the linear experts directly depend on the value of current and past observed neural signals.

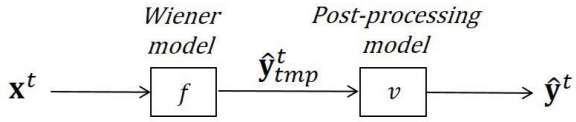

Figure 5.4: Post-processed Wiener Filter decoder, data flow. The post-processed operator performs static state detection from the kinematic estimates yielded by a single Wiener filter.

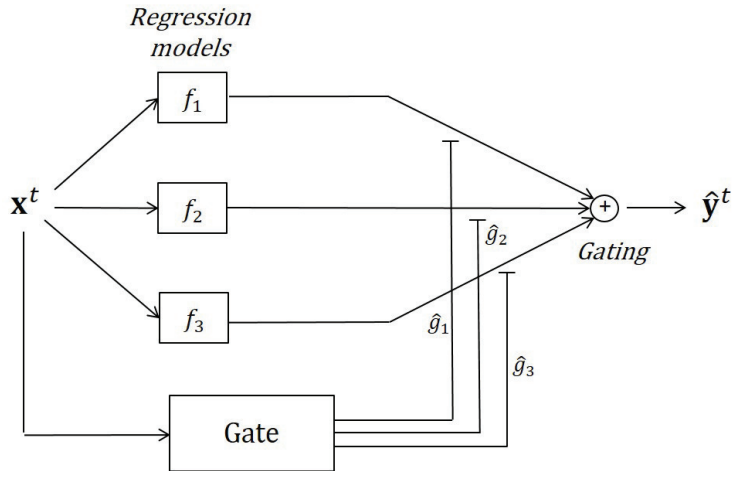

Figure 5.5: Switching Linear Model decoder, data flow. The gate performs static state detection. The SLM thus corresponds to a generic Mixture of Experts model.

5.2 Performance indicators

Performance assessment was performed on test sets independent from the training data sets. Indicators associated with binary decoding were first computed to monitor the ability of each decoder to support NC states when performing asynchronous mono-limb or multi-limb decoding. When multiple IC experts were considered, the multiple IC states $\text{IC}_j, j = 1, \ldots, K$ were pooled into a general IC state, and an additional indicator was introduced to measure the ability of the state decoder to distinguish between the multiple $\text{IC}_j, j = 1, \ldots, K$ states when detecting an IC state. The accuracy of the kinematic estimates during IC periods was finally established via a set of indicators suited for continuously-valued variables.

5.2.1 NC support for asynchronous mono- and multi-limb decoding

Two groups of performance indicators were used to measure the quality of the NC support associated with asynchronous mono- and multi-limb decoding. A first indicator set characterized NC/IC classification accuracy (i.e., number of correctly classified samples). In the case of multi-limb decoding, IC_i, $i = 1, \ldots, K$ states were pooled into a global IC state. A second indicator set focused on the number of false activations/deactivations and on their durations.

Confusion matrix-based performance indicators The indicators used to assess the performance of binary decoders [Mason et al., 2006] mainly rely on the confusion matrix. The confusion matrix gathers the number of NC states which are correctly (True Negatives, TN) or wrongly (False Positives, FP) labelled, and the number of IC states which are correctly (True Positives, TP) or wrongly (False Negatives, FN) labelled by the decoder. Several indicators are derived from the confusion matrix. The classification error $ERR = \frac{FP+FN}{TP+TN+FP+FN}$ and the accuracy $ACC = \frac{TP+TN}{TP+TN+FP+FN}$ are relevant measures of the global classification quality when classes are well balanced [Mason et al., 2006] [Thomas et al., 2013]. The True Positive Rate $TPR = \frac{TP}{TP+FN}$ (sensitivity) and the False Positive Rate $FPR = \frac{FP}{FP+TN}$ additionally permit to reflect the classifier ability to correctly decode IC and NC states, respectively.

The NC and IC classes were relatively well balanced in the preclinical and clinical data sets (see **Chapter 4**). The accuracy ACC, the TPR and the FPR were therefore computed. In the case of mono-limb decoding, the Area Under the Curve (AUC) [Davis and Goadrich, 2006] was additionally computed to reflect the decoding performances which would be obtained by using a different prior probability for each state, i.e. by using a probability threshold different from 0.5 to assign labels to observation samples.

Block-wise indicators: frequency and duration of false detections The TPR, FPR, ACC and AUC are sample-based indicators. They don't take into account

the dynamic of misclassified samples, even if it is liable that several consecutive misclassified samples are less disturbing to users that the same number of isolated misclassified samples. State decoding performance was thus considered in terms of false activations/deactivations' duration and frequency, where a "false activations" (respectively, a "false deactivation") refers to a block of consecutive NC samples misclassified as IC samples (respectively, a block of IC samples mistaken for NC samples). A block of misclassified samples was counted as one false activation or deactivation. Its duration was computed as the number of samples divided by the decision rate. These block-wise criteria are expected to reflect the users' evaluation of the quality of NC support during asynchronous closed-loop control sessions.

5.2.2 Multiple IC state discrimination for sequential multi-limb asynchronous decoding

When multiple IC states are considered, it may additionally be needed to assess the decoder ability to discriminate between the K different IC states, IC_i, $i = 1, \ldots, K$. In the case of multi-limb movements for example, it is desirable to check whether the activation of one limb can be distinguished from the movement of the other limbs. To this aim, the ratio of correctly classified TP samples was established where appropriate. A $K \times K$ confusion matrix was computed on TP samples, and the sum of its diagonal terms was divided by the total number of IC samples (see, e.g., [Gouy-Pailler et al., 2009]).

5.2.3 Performance indicators for continuous dependent variable

The accuracy of continuous variable estimates is typically assessed via the Pearson Correlation Coefficient (PCC) and the Normalized Root-Mean-Squared Error (NRMSE) [Spuler et al., 2015]. The PCC measures the amount of linear dependence between observed y and predicted $\hat{y}$ variables: $\text{PCC}(y, \hat{y}) = \frac{cov(y,\hat{y})}{\sigma_y \sigma_{\hat{y}}}$, where $cov(y, \hat{y})$ indicates the covariance between y and $\hat{y}$ and σ_y refers to the standard deviation of y. The NRMSE measures the ℓ^2-error between the vectors of observations $\mathbf{y} = (y^1, y^2, \ldots, y^T)'$ and $\hat{\mathbf{y}} = (\hat{y}^1, \hat{y}^2, \ldots, \hat{y}^T)'$, where $(.)'$ denotes the transpose: $\text{NRMSE} = \frac{\|\mathbf{y}-\hat{\mathbf{y}}\|_2}{\|\mathbf{y}-\bar{\mathbf{y}}\|_2}$, where $\|\mathbf{y}\|_2 = \sqrt{\sum_{t=1}^{T}(y^t)^2}$ and $\bar{\mathbf{y}} = \frac{1}{T}\sum_{t=1}^{T} y^t$.

The Normalized Mean Absolute Error $\text{NMAE} = \frac{\|\mathbf{y}-\hat{\mathbf{y}}\|_1}{\|\mathbf{y}-\bar{\mathbf{y}}\|_1}$, where $\|.\|_1$ is the ℓ^1-norm ($\|\mathbf{y}\|_1 = \sum_{t=1}^{T} |y^t|$), was additionally computed. It issues a measure of the ℓ^1-error between $\mathbf{y}$ and $\hat{\mathbf{y}}$ and is less sensitive to outliers than the NRMSE [Hyndman and Koehler, 2006].

5.2.4 Trajectory synchronization

Temporal uncertainties are often observed in BCI estimated trajectories, for example because of the temporal instability of neural signals [Tranquillo, 2013] or of the limited temporal resolution of the BCI decoders. Generic distance-based performance indicators are highly sensitive to temporal uncertainties, e.g. shift or time warping

[Cassisi et al., 2012] [Wang et al., 2013d]. Slight positive and negative delays between estimated and target time series, for example, substantially degrade distance-based performance indicators, especially the ℓ^2-distance between both time series. Such performance indicators have consistently be shown to be suboptimal when they are used to estimate the similarity between time series, and when the resulting similarity measures are utilized to train a time series classifier [Wang et al., 2013d]. Although large delays between intended and estimated trajectories, e.g. 300ms pure delay [Willett et al., 2013], are known to significantly degrade BCI control performances [Willett et al., 2013] [Marathe and Taylor, 2015], it is unclear whether slight temporal uncertainties are more or less disturbing to users than spatial errors, which have been shown to degrade control quality [Marathe and Taylor, 2015]. Performance indicators insensitive to temporal uncertainties were thus added to the set of generic indicators presented in **section 5.2.3**.

Various indicators have been proposed to measure the similarity between time series [Wang et al., 2013d], e.g. the Dynamic-Time-Warping or the Longest Common Subsequence similarity measures [Cassisi et al., 2012]. In particular, Dynamic-Time-Warping (DTW) criteria have been used alonside traditional ("*lock-step*" [Wang et al., 2013d]) indicators to assess the performance of several BCI decoders in [Eliseyev and Aksenova, 2014]. Let us consider two times series $\mathbf{x} \in \mathbb{R}^T$ and $\mathbf{y} \in \mathbb{R}^T$ of similar length T. Dynamic Time Warping finds the $\tilde{T}$-long path $\tilde{\mathbf{P}} \in \mathbb{N}^{\tilde{T}\times 2}$ which minimizes the cumulative distance $\sum_{k=1}^{\tilde{T}} d(x_{\tilde{p}_{k,1}}, y_{\tilde{p}_{k,2}})$, where d is a "local cost measure" [Muller, 2007], e.g. the ℓ^2- or ℓ^1-distance. The admissible paths can be limited by constraining the maximum delay between two paired samples, e.g. using the Sakoe-Chiba band [Cassisi et al., 2012]. An illustration of the errors reflected by lock-step and DTW-based indicators is shown in **Figure 5.6**.

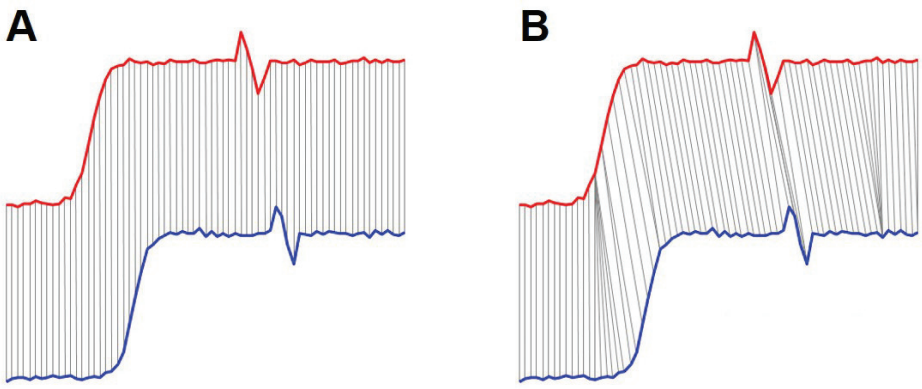

Figure 5.6: Generic lock-step- (**A**) and DTW-matching (**B**).

Synchronized signals are defined on a longer, distorted temporal scale defined by the DTW optimal path (see **Figure 5.7.A**). Although the cumulative distance $DTW(\mathbf{x}, \mathbf{y}) = \min_{\tilde{\mathbf{P}}} \sum_{k=1}^{\tilde{T}} d(x_{\tilde{p}_{k,1}}, y_{\tilde{p}_{k,2}})$ associated with the optimal DTW path can be used as a criterion, it does not reflect the impact of desynchronization on tradi-

tional performance indicators, e.g. the PCC, NRMSE or NMAE. A method is here proposed for desynchronization-robust direct computation of generic performance indicators between two trajectories $\mathbf{x} \in \mathbb{R}^T$ and $\mathbf{y} \in \mathbb{R}^T$. The indicators of interest are computed after having obtained signals synchronized in the original time scale of one of the trajectories, taken as the reference signal. This approach makes direct comparisons of performance indicators between raw and synchronized trajectories possible.

First, synchronized signals $\tilde{\mathbf{x}} \in \mathbb{R}^{\tilde{T}}$ and $\tilde{\mathbf{y}} \in \mathbb{R}^{\tilde{T}}$ are obtained along the DTW-distorted time scaled defined by the DTW-optimal path $\tilde{\mathbf{P}} = \text{argmin}_{\tilde{\mathbf{P}}} \sum_{k=1}^{\tilde{T}} d(x_{\tilde{p}_{k,1}}, y_{\tilde{p}_{k,2}})$. The DTW path cannot be directly interpreted as a generic function $f : \mathbb{N}^T \to \mathbb{N}^T$ because several outputs can be associated with an input value, i.e. the DTW path can be seen as a multi-valued map [Kuratowski, 2014] (**Figure 5.7.B**). For this reason, the DTW path $\tilde{\mathbf{P}} \in \mathbb{N}^{\tilde{T}\times 2}$ was approximated by a function-like path $\mathbf{P} \in \mathbb{N}^{T\times 2}$. This path is obtained by computing the average of the signal $\mathbf{y}$ over the (possibly multiple) instants associated with an original instant (see **Figure 5.7.C**). A synchronization example is shown in **Figure 5.8**. The ℓ^2-distance was used to synchronize the estimated trajectory before computing the PCC and NRMSE because these two performance indicators are based on the ℓ^2-distance. Similarly, ℓ^1-based synchronization was performed to assess the NMAE between synchronized signals. Maximum lags of 100ms or 200ms (1 and 2 samples, respectively) were utilized to constrain the considered DTW paths (see **Chapter 7**).

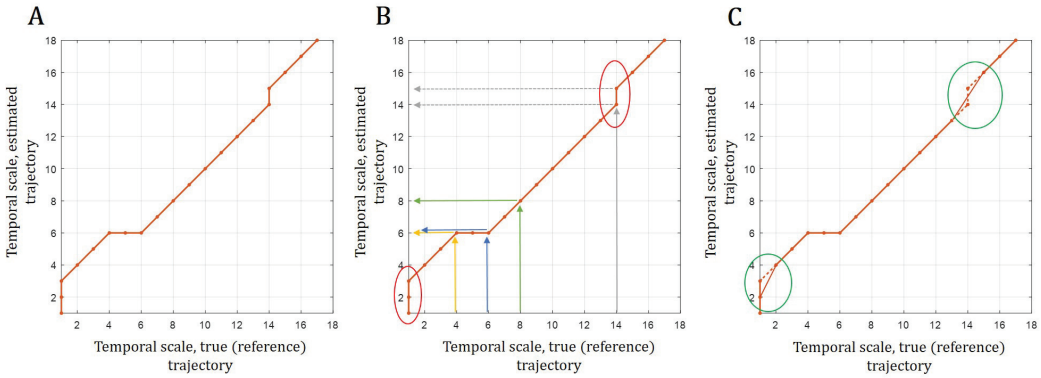

Figure 5.7: **A.** Example of a DTW path, clinical data set. DTW was applied on a true and estimated finger trajectory. The initial part of the DTW path (corresponding to the first 18 samples of the reference (true) trajectory) is here displayed. Signals are synchronized when they are evaluated on the distorted time scales represented by the DTW path. **B.** Correspondence between instants of the time scale 1 and 2. Most instants defined on the first (reference) time scale are associated with a single instant of the second time scale (yellow, blue and green arrows). In some cases, however, several instants of the second time scale correspond to the same original instant (red circles). **C.** Approximation of the DTW path by a function-like path.

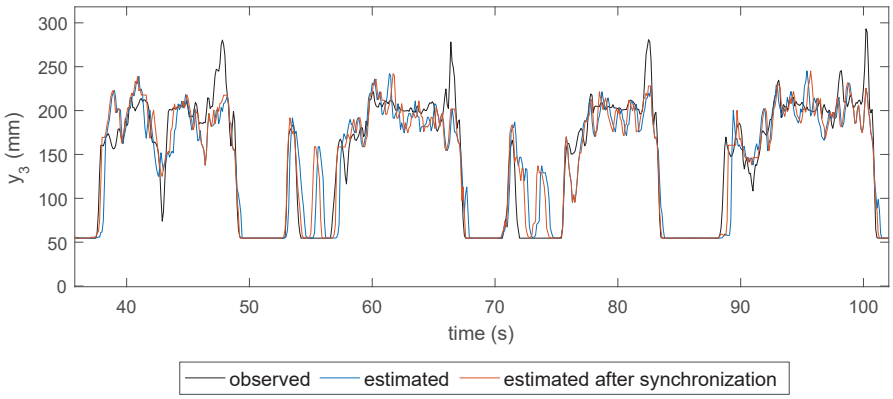

Figure 5.8: Synchronization example, MSLM decoder, subdural data set. The proposed synchronization method was used to synchronize the MSLM trajectory estimate (blue line) with the observed trajectory (black line). The ℓ^1-distance and a maximum lag of 300ms (3 samples) were used to compute the synchronized MSLM estimate (red line).

5.3 Statistical tests

A statistical procedure for the comparison of several decoders on multiple data sets is here necessary. Following the recommandations of [Demšar, 2006], the significance of performance differences was established using the Friedman test with the significance level $\alpha = 0.05$.

The Friedman test permits to analyse repeated measures performed over the same samples, here for the analysis of the performance of $D \geq 2$ different decoders applied on the same acquisition sessions. It ranks the decoders on each acquisition session according to their respective performance [Demšar, 2006]. It then tests whether the decoders' respective average ranks are significantly different. To this aim, the test statistic $t_{Friedman} = \frac{12N}{D(D+1)} \left[\sum_j (\frac{1}{N} \sum_i r_i^j)^2 - \frac{D(D+1)^2}{4} \right]$, where r_i^j is the rank of the decoder j for the acquisition session i, is computed. Here, $D = 3$ is the number of compared decoders and N is the number of acquisition sessions ($N = 8$, $N = 16$ and $N = 9$ for the subdural, epidural and finger data sets, respectively). The $t_{Friedman}$ statistic follows a χ_F^2 law with $D - 1$ degrees of freedom for large enough N and D (typically, $N > 10, D > 5$) [Demšar, 2006]. Adjusted critical values have been proposed for small N and D, and have been used for the subdural data set ($N = 7$). The implementation made publicly available by [Cardillo, 2009] was used.

Various procedure have been proposed for post-hoc testing, e.g. the Nemenyi Post-hoc test [Demšar, 2006] [Pereira et al., 2015], Bonferroni-Dunn, or step-wise procedures such as Holm's, Hochberg's or Hommel's procedures. In [Benavoli et al., 2015], it was advocated to avoid post-hoc tests based on the mean-ranks issues by the Friedman test. Post-hoc pairwise decoder comparisons were thus performed using the sign test with Bonferroni correction [Benavoli et al., 2015].

5.4 Conclusion

In this chapter, it is proposed to compare the Markov Switching Linear Model (MSLM) with a Wiener Filter with Markovian post-processing (MpWF) and with a Switching Kalman Filter (SKF), which combines K Kalman filters. The MpWF and SKF permit to investigate the relevance of key features of the MSLM, namely switching and discriminative modelling. It is finally proposed to specifically evaluate the impact on state detection of the third MSLM's key feature, i.e. dynamic gating, by considering a Switching Linear Model (SLM) and a post-processed Wiener Filter (pWF). The SLM and pWF are variants of the MpWF and the MSLM with static state detection.

A set of performance indicators has been chosen to assess the ability of the decoders to perform asynchronous mono-limb and sequential multi-limb trajectory decoding and to accurately reconstruct kinematic parameters during IC periods. The quality of NC support is traditionally measured by means of confusion-matrix-based performance indicators. It is here proposed to complement these indicators by block-wise performance indicators. Block-wise indicators take into account the dynamic of false activations. They are therefore expected to reflect the users' evaluation of

the quality of NC support during asynchronous closed-loop control sessions. The performance indicators associated with kinematic reconstruction during Intentional Control (IC) periods generally measure both temporal and spatial accuracy. Because desynchronizations between observed and estimated trajectories are likely to impact performance indicators, it is proposed to compute generic performance indicators between both raw (desynchronized) and synchronized versions of the observed and estimated trajectories. An approach has been designed to compute synchronized trajectories which are defined on the original time scale of the observed trajectory. Finally, a statistical procedure has been chosen to assess the significance of the differences between the decoders' respective performances.

Chapter 6

Implementation

Contents

Transducers embedding a Markov Switching Linear Model (MSLM) were implemented for two decoding tasks, namely asynchronous wrist trajectory reconstruction (preclinical data set) and multi-limb (multi-finger) trajectory reconstruction (clinical data set). Wiener Filters with Markovian post-processing (MpWF) and Switching Kalman Filter decoders (SKF) were additionally implemented for comparison purposes (see **section 5.1**). Application of the MSLM, MpWF and SKF decoders was preceded by the extraction of neural and kinematic features.

6.1 Feature extraction and pre-processing

Features were extracted from the subjects' ECoG neural signals and from their wrist's or fingers' kinematics.

6.1.1 Kinematic parameter extraction

Kinematic encoding in ECoG features has mainly been disclosed for direction, position, velocity and speed (see **section 1.2.3.2**). Although speed decoding has yielded promising results [Bundy et al., 2016] [Hammer et al., 2016], speed alone cannot enable users to control a BCI effector. While ECoG directional tuning has been explored in several studies [Schalk et al., 2007] [Anderson et al., 2012] [Nurse et al., 2015], kinematic reconstruction from ECoG signals has mainly been investigated for position and/or velocity [Chao et al., 2010] [Shimoda et al., 2012] [Marathe and Taylor, 2013] [Pistohl et al., 2008] [Spüler et al., 2016] [Liang and Bougrain, 2009] [Chen et al., 2014b] [Flamary and Rakotomamonjy, 2012]. In particular, the studies previously led on the preclinical [Chao et al., 2010] [Shimoda

et al., 2012] and clinical data sets [Liang and Bougrain, 2009] [Chen et al., 2014b] [Flamary and Rakotomamonjy, 2012] considered in the present dissertation focused on position reconstruction with Wiener filtering or variants. Position decoding was therefore utilized to assess the performance of the MSLM.

A state variable composed of both position and velocity was used to implement the SKF, as it had been reported as optimal for Kalman filtering of ECoG signals in an earlier study [Pistohl et al., 2008].

Preclinical data set The NHPs' wrist position $\mathbf{y}^t \in \mathbb{R}^3$ was issued by the motion tracking system. Velocity was derived from position using a forward-difference approximation [Eberly, 2014]. NC and IC states were manually labelled.

Clinical data set Finger positions $\mathbf{y}^t \in \mathbb{R}^5$ were acquired by the data glove worn by the subjects. Similarly to the preclinical data set, velocity was derived from position using a forward-difference approximation. NC and $\mathrm{IC}_i, i = 1, \ldots, \mathrm{IC}_5$ states were manually labelled.

6.1.2 Neural signal feature extraction

A high-dimensional time-frequency-space representation was chosen to exploit the position- and velocity-tuned features disclosed in ECoG signals.

Time-frequency features were extracted for each channel from Δt-long ECoG sliding epochs ($\Delta t = 1s$, sliding step: 100ms) following [Eliseyev and Aksenova, 2014]. A Complex Continuous Wavelet Transform (CCWT) was applied on neural features. The complex Morlet wavelet (**Figure 6.1**), which had been found efficient for EEG and ECoG neural signal analysis in earlier studies [Lemm et al., 2004] [Eliseyev and Aksenova, 2014], was chosen in the present study. The ECoG frequency content was analysed between 1 and 250Hz to exploit the frequency bands in which kinematic tuning had been disclosed (see **section 1.2.3.2**). Redundant sampling of this frequency domain was achieved via 38 daughter wavelets chosen with a logarithmic scale. A logarithmic transform was applied on the CCWT transformed signals' absolute value. The average value of the log-transformed signals was computed in 100ms sliding windows (100ms step), resulting in a 10-points description of ECoG 1s-long time bins for each frequency band and each channel. Low frequency components have been shown to be particularly informative for kinematic reconstruction from ECoG signals (see **section 1.2.3.2**). They were extracted by means of a Savitzky-Golay filter [Schafer, 2011] (window length of 200ms, order 2), and added to the CCWT-based frequency features. Thus the ECoG epoch $[t - \Delta t : t]$ was described by the temporal-frequency-spatial feature vector $\mathbf{x}^t \in \mathbb{R}^m$, where $m = n_c \times 10 \times (38 + 1)$. Regarding the preclinical data set, $n_c = 32$ for Monkey A and $n_c = 64$ for Monkeys K, B and C. Because of differences in the number of electrodes implanted in each one of the subjects (see **section 4.2.3**), $46 \leqslant n_c \leqslant 64$ for the clinical data set. Finally, the pre-processing approach proposed in [Eliseyev and Aksenova, 2014] has been

used to detect potential artefacts in the neural data (e.g., chewing artefacts in the epidural data set), and to replace them with neutral values.

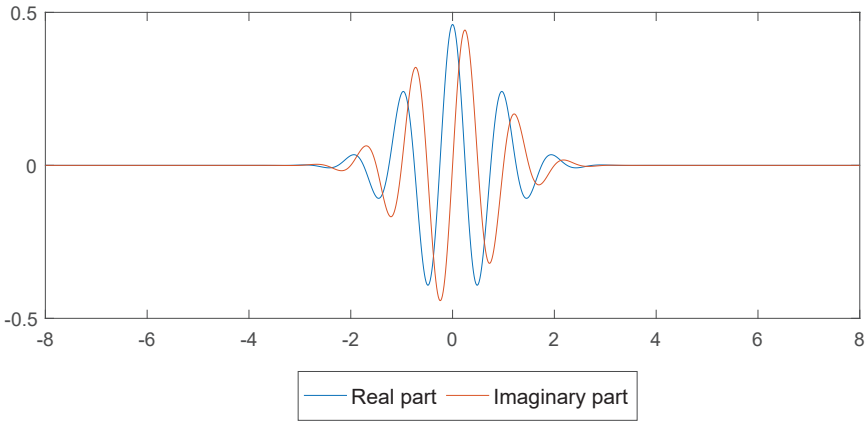

Figure 6.1: Real and imaginary parts of a Morlet wavelet.

6.2 Decoder implementation

The three decoders (MSLM, MpWF and SKF) were trained on the first 70% of each session of the preclinical and clinical data set.

6.2.1 Preclinical data set

MSLMs with $K = 2$ experts were first implemented to explore the integration of NC support into mono-limb kinematic decoders. As NC and IC were labelled in the training data set, a supervised training approach was used. A MpWF and a SKF with $K = 2$ experts were trained for comparison purposes. The training of MSLMs with $K > 2$ experts was additionally performed for complex reaching movements. As the labels associated with each IC experts were unknown, unsupervised training of the MSLMs was carried out.

6.2.1.1 NC/IC MSLM (supervised training)

MSLM decoders with $K = 2$ experts, namely a NC and an IC expert, were implemented for the integration of NC support into a continuous mono-limb trajectory decoder. Supervised Maximum Likelihood training of the MSLM decoders was performed using complete training data sets $\{\mathbf{X}, \mathbf{Y}, \mathbf{z}\}$.

Experts The NC expert yielded the NC neutral position value $\bar{\mathbf{y}}_{NC}$, which was estimated as the average value of $\mathbf{y}^t$ computed over NC states. The IC expert was dedicated to wrist position decoding during IC periods. ML training of the IC expert

parametrized by $\mathbf{B}_{IC}$ consisted in maximizing $\sum_{t=1}^{T} \delta_{z^t,IC} \ln P(\mathbf{y}^t|\mathbf{x}^t, \mathbf{B}_{IC})$, where T refers to the number of training samples, $\delta_{z^t,IC} = 1$ if the observation t corresponds to an IC period, and $\delta_{z^t,IC} = 0$ otherwise. A Gaussian noise was associated with the IC expert. The corresponding Ordinary Least Squares (OLS) solution to this maximization problem is unstable because of the high dimension of the input variable $\mathbf{x}^t$. Several approximate solutions have been proposed to identify high-dimensional linear models, in particular Principal Component Regression, Partial Least Squares and LASSO training. A pilot study was therefore performed to select a relevant approximate ML estimator of the IC expert (see **Appendix B**). Partial Least Squares (PLS) [Höskuldsson, 1988] regression was found to provide accurate IC experts, and was therefore chosen to identify the IC experts. The optimization of the subspace dimension relied on a 6-fold cross-validation procedure completed on the training data set, namely on the application of Wold's R criterion on the cross-validated PRESS statistic [Li et al., 2002].

Gate Fully supervised training of the MSLM involves the identification of HMM-based dynamic state decoder (gate) (see **section 3.3.1**). In the considered case of mono-limb asynchronous decoding, the gate has to distinguish between 2 states, namely NC and IC states.

The state-specific emission distributions associated with the HMM-based gate of the MSLM can be modelled by means of generic distributions or by using an alternative approach based on discriminative modelling: $P(\mathbf{x}^t|z^t = j) \propto \frac{P(z^t=j|\mathbf{x}^t)}{P(z^t=j)}$, where $P(z^t = j|\mathbf{x}^t)$ is yielded by a discriminative state decoder (e.g., using Neural Networks [Renals et al., 1994] [Bourlard and Morgan, 1998] [Ordóñez et al., 2013] or Support Vector Machines [Valstar and Pantic, 2007] [Ordóñez et al., 2013], see **section 3.2.1**). Thus, both generative and discriminative state decoders can be embedded into the dynamic HMM-based gate of the MSLM. Both types of model have been applied for state detection in BCI studies (see **section 2.4.2**). Their respective performances has often been reported as dependent on the problem at hand. In particular, decoding efficiency may depend on the distribution of the neural features within each state. If for example this distribution cannot be satisfyingly modelled with a generic distribution like a Gaussian or Mixture of Gaussians distribution, using a discriminative approach may permit to improve state detection. A preliminary study was therefore completed to assess the relevance of different modelling approaches and select the strategy which was optimal for NC/IC discrimination on the preclinical data set (see **Appendix B**). For the sake of simplicity, this study was carried out on static decoders. The selected modelling approach was subsequently embedded into the HMM-based dynamic gate of the MSLM.

The set of investigated classifiers included Linear Discriminant Analysis (LDA) and Support-Vector-Machine (SVM), which are regularly applied to distinguish between neural states in motor BCIs [Fifer et al., 2014] [Kapeller et al., 2015] [Hotson et al., 2016] [Hortal et al., 2015]. It additionally comprised nonlinear

counterparts of the LDA and of the SVM, namely Quadratic Discriminant Analysis (QDA) and a nonlinear SVM. Although it has been comparatively unfrequently applied in BCI studies, high performance of Logistic Regression-based (LR) state classification has been reported in recent EEG- and ECoG-based studies [Bashashati et al., 2015] [Bundy et al., 2016]. The relevance of LR was therefore additionally assessed for the considered binary classification task. LDA and QDA are generative classifiers relying on Gaussian distributions (see **Chapter 2**). They are thus static analogues of generic HMMs with Gaussian emission distribution. By contrast, LR and SVMs are discriminative classifiers (see **Chapter 2**). They can be seen as static analogues of HMMs exploiting a discriminative approach to model the emission probabilities. Both linear (LDA, SVM and LR) and nonlinear (QDA and nonlinear SVM) classifiers were considered because linear state detection is liable to be suboptimal if neural signals are not linearly separable.

The neural input feature variable was here high dimensional. As classifier performances have been shown to depend on the characteristics of the input variable [Bashashati et al., 2015], in particular on its dimensionality [Bhattacharyya et al., 2011], the classifiers were tested after reduction of the feature dimension. Different projection-based dimensionality reduction procedures were compared. Both unsupervised and supervised projections were considered, namely PCA- and PLS-based dimensionality reduction. Computation of the PLS-based projector was found by fitting a PLS model between the high-dimensional input variable $\mathbf{x}^t$ and the state variable $z^t \in \{0, 1\}$. A detailed presentation of the preliminary study can be found in **Appendix B**.

PLS-based dimensionality reduction followed by the application of a logistic model was found to be efficient for NC/IC detection (see **Tables B.6** and **B.7**). It was consequently integrated into the HMM-based gating network for dynamic state detection. 6-fold cross-validation was used to find the optimal dimension of the latent subspace yielded by a PLS regression between $\mathbf{x}^t$ and z^t. Optimization was completed by applying Wold's R criterion on the 6-fold cross-validated PRESS statistic [Li et al., 2002]. The logit model parameters were fitted using the Iteratively Reweighted Least Squares algorithm [Bishop, 2006].

A Switching Linear Model with static state detection was additionally inferred from the MSLM gate and experts. It was obtained by discarding the Markovian hypothesis used by the MSLM's gate, i.e. by exploiting the static LR embedded into the MSLM's HMM-based gate to combine the experts identified during the training of the MSLM.

6.2.1.2 Post-processed Wiener Filters

A Markovian post-processed Wiener Filter was fitted via PLS regression on the training data set $\{\mathbf{X}, \mathbf{Y}\}$. The optimal number of PLS factors was estimated by applying Wold's R criterion on the 6-fold cross-validated PRESS statistic. Trajectory post-processing was used to integrate NC support into the decoder. A logistic

regression was trained to infer NC and IC states from the kinematic estimates $\hat{\mathbf{y}}^t$. Similarly to the MSLM, a Markovian hypothesis was used to limit spurious detections during both NC and IC states. The corresponding transition matrix was found using the procedure presented in **Chapter 3** in the case of the MSLM. The state probabilities $P(z^t|\hat{\mathbf{y}}^{1:t})$ yielded by the dynamic LR state decoder were used to weight the Wiener estimates and the neutral values associated with NC states. The dynamic logistic regression was identified on the training data set $\left\{\hat{\mathbf{Y}}, \mathbf{z}\right\}$.

The parameters of a Wiener Filter with static post-processing (pWF) were additionally inferred from the parameters of the MpWF. The pWF combined the LR and Wiener models identified during the training of the MpWF, i.e. the outputs of the LR embedded into the dynamic post-processing operator of the MpWF were directly used to weight the Wiener estimates and the neutral values associated with NC states.

6.2.1.3 SKF

A SKF (see **Appendix C**) was implemented for dynamic combination of $K = 2$ Kalman filters, one specialized in NC periods and the other in IC periods. The SKF state variable (i.e., response variable) $\mathbf{y}^t_{SKF}$ was chosen as the monkey's wrist position and velocity because it had been reported as optimal for ECoG decoding [Pistohl et al., 2008]. The neural features $\mathbf{x}^t$ fed to MSLM decoders embedded a 1s-long history of ECoG signal, i.e. they consisted of frequency-space features observed in ten 100ms-long time bins. By contrast, the neural features $\mathbf{x}^t_{SKF}$ considered by SKFs were only observed within an 100ms-long time bin, because SKFs rely on an explicit approach to perform dynamic modelling. Similarly to the MSLM state decoder, the dimension of the neural features was reduced before application of the SKF. PCA- and PLS-based dimensionality reduction procedures were compared in a pilot study presented in **Appendix C**, which additionally permitted to find the optimal lag between the neural feature time bin and the kinematic parameters of the monkeys' wrist. PLS-based dimensionality reduction of the neural features associated with the last 100ms before the instant t considered for trajectory estimation (10^{th} time bin of the neural features $\mathbf{x}^t$) was found to correspond to the best KF decoding accuracy, and was therefore used to limit the SKF computational cost. PLS regression between $\mathbf{x}^t_{SKF}$ and $\mathbf{y}^t_{SKF}$ was thus used to identify an informative low-dimensional subspace. While neural features were normalized after feature extraction, a constant term was nevertheless added to the neural features to account for modifications of the baseline activity during NC and IC states. Subspace dimension was chosen by 6-fold cross-validation on the training data set. OLS estimates of the transition $\mathbf{A}_k$ and emission $\mathbf{C}_k$ matrices were computed on the training data sets $\{\mathbf{X}_{SKF,k}, \mathbf{Y}_{SKF,k}\}$. ML estimates of the variance matrices $\mathbf{\Gamma}_k$ and $\mathbf{\Sigma}_k$ were found following [Aggarwal et al., 2013].

6.2.1.4 MSLM with multiple IC experts (unsupervised training)

Some reaching movements of the preclinical data set exhibit complex behaviours, where complexity is measured by the number of zero-crossing of the trajectory first derivative (see **Chapter 4**). Subdural sessions with an average complexity superior to 2 were selected for an additional analysis (three sessions of the Monkey A), namely for the training of a MSLM with 2 IC experts. Unsupervised Maximum Likelihood training of the MSLM decoders was performed using the incomplete training data sets $\{\mathbf{X}, \mathbf{Y}\}$. Both NC and $\text{IC}_i, i = 1, 2$ states were considered as hidden during training. The decision of using unconstrained NC/IC state labels was taken because of labelling uncertainties, for example around transitions or for very short null-velocity periods during manually identified movements.

Structure A MSLM decoder with $K = 3$ experts was implemented to improve kinematic modelling during reaching movements in the context of asynchronous mono-limb kinematic decoding. One expert was specialized in NC periods, and the $K - 1 = 2$ other experts were dedicated to complex wrist kinematic decoding during IC periods. The experts were assumed to be associated with a Gaussian noise. Following the observation that penalized LR was optimal for multi-class classification (see **Appendix B**, clinical data set), a 3-class LR was here considered to compute the emission probabilities of the MSLM HMM-based gate.

Initialization The EM-algorithm converges to a local solution [Roche, 2011]. Parameter initialization is therefore crucial. While clustering algorithms are frequently used for parameter initialization, generic clustering algorithms such as Mixtures of Gaussian and k-means are generally not suitable for high-dimensional data [Parsons et al., 2004]. High-dimensional clustering involves complex and potentially time-consuming procedures [Parsons et al., 2004] [Bouveyron and Brunet, 2013]. For this reason, initialization was here performed by dividing movements into K-1 = 2 parts (see **Figure 6.2**), and by using the corresponding state sequence to perform supervised training of the MSLM gate and experts. This initialization approach permitted to exploit the *a priori* knowledge that models are likely devoted to different phases of the movement. Similarly to the supervised NC/IC MSLM presented above, an approximate solution to the initial ML training of the MSLM was considered because of the high-dimension of the neural subspace. PLSR was used to fit the experts, and LASSO training was used to identify the parameters of the LR which permitted to model the emission probabilities. Both gate and expert training depend on the value of a hyperparameter, namely the amount λ of penalization for LASSO training of the gate and the number F of PLS factors for the experts. Both hyperparameters were selected via 6-fold cross-validation performed during the gate and expert initialization. The amount λ of penalization was selected so as to minimize the cross-validated error. Similarly, the number F of expert factors was found by applying Wold's R criterion on the cross-validated PRESS statistic. The considered hyperparameters were not refitted during the M-step iterations.

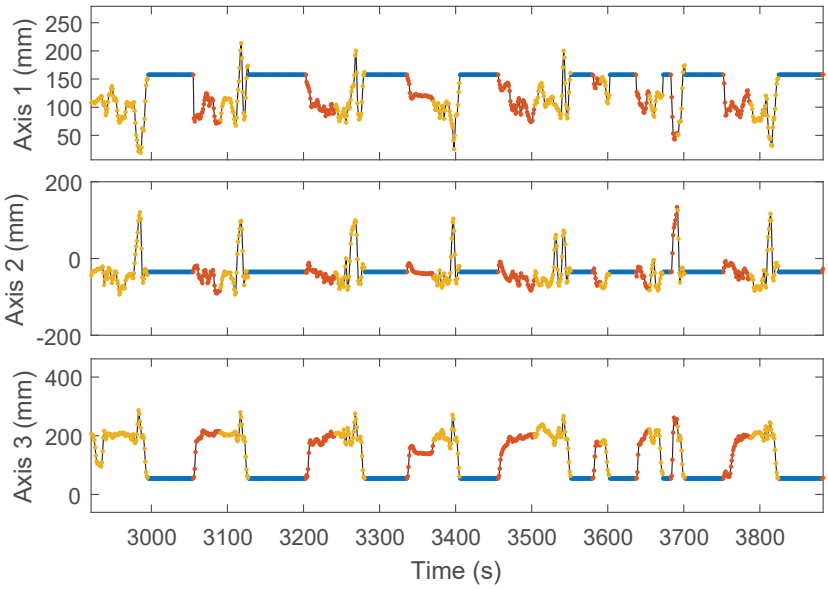

Figure 6.2: Initialization of the EM algorithm, example (subdural data set). Each color refers to a different state.

E-step and M-step The E-step and M-step presented in **Chapter 3** were iteratively performed to train the MSLM. Specific optimization methods are generally necessary to solve the M-step maximization problems because the impact of the observations t on a parameter associated with state i is weighted by the corresponding state probability $\gamma_i^t = P(z^t = i|\mathbf{x}^{1:T}, \mathbf{y}^{1:T}, \Theta)$ (see **section 3.3.2.3**).

Weighted expert training While the Weighted Linear Least Squares (WLLS) estimator $\hat{\mathbf{B}}_k = (\mathbf{X}'\mathbf{\Gamma}_k\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Gamma}_k\mathbf{Y}$, where the weight matrix $\mathbf{\Gamma}_k \in \mathbb{R}^{T\times T}$ is a diagonal matrix with the diagonal elements $\gamma_k^t, t = 1, \ldots, T$ [Moerland, 1997], is regularly used to train Mixtures of Experts, it is not defined when $m \leq T$. For this reason, a weighted variant of the PLS-based approximate solution presented in **Appendix B** was here utilized, namely weighted PLS. A PLS-based approximate ML estimator of expert k was found by fitting a PLSR between $\mathbf{\Gamma}_k^{\frac{1}{2}}\mathbf{X}$ and $\mathbf{\Gamma}_k^{\frac{1}{2}}\mathbf{Y}$ [Souza and Araújo, 2014]. Similarly to the supervised training case, the NC expert yielded the NC neutral position value $\bar{\mathbf{y}}_{NC}$.

Weighted gate training Similarly, weighted LR training was required to update the gate parameters at each M-step. Procedures for weighted training of LR models are for example used for EM-based training of Mixtures of Experts with a LR gate (e.g., [Waterhouse, 1997] [Moerland, 2000]). As generic estimators of LR models are liable to be unstable [Bishop, 2006], and may not exist when high-dimensional features are considered, a LASSO weighted training procedure was utilized to update the parameters of the LR embedded into the HMM's gate. Similarly to the comparative study presented in **Appendix B**, the efficient training approach [Friedman et al., 2010] implemented in the open-source glmnet toolbox [Qian et al., 2013] was used to perform the LR weighted LASSO training. It should be noted that the utilization of such iterative optimization approaches results in the presence of an "inner-loop" within each M-step, and significantly increases the computational cost of the EM algorithm (hence the LDA-based gating network presented in [Xu et al., 1994] or the approximate approach proposed in [Moerland, 1997]).

Convergence A stopping criterion was used to terminate the EM training when the data log-likelihood was not improved by additional iterations.

6.2.2 Clinical data set

MSLMs with $K = 6$ experts were implemented to explore the ability of the MSLM to handle asynchronous sequential multi-limb kinematic decoding. Similarly to the study led on the preclinical data set with binary MSLM, a supervised training approach was used because NC and $\text{IC}_i, i = 1, \ldots, 6$ states were labelled in the training data set. A MpWF and a SKF with $K = 6$ experts were trained for comparison purposes.

6.2.2.1 MSLM

A MSLM decoder with $K = 6$ was implemented for asynchronous finger trajectory decoding. The IC expert IC_i was associated with periods during which finger i was moving, $i = 1, \ldots, \text{IC}_5$.

Experts Similarly to the preclinical data set, the NC expert yielded the NC neutral position value $\bar{\mathbf{y}}_{NC}$, estimated as the average value of $\mathbf{y}^t$ computed over NC states. The i^{th} IC state referred to periods during which movements were executed by the i^{th} finger. A linear expert was trained for each IC state, $\text{IC}_1, \ldots, \text{IC}_5$. LASSO training was found optimal to identify the parameters of the experts in the preliminary study presented in **Appendix B**. 6-fold cross-validation was used to choose the amount λ of ℓ^1-penalization. Similarly to the preclinical data set, the parameters of the SLM were inferred from the MSLM parameters.

Gate A preliminary study was again completed to select the best state decoder for the considered gating task. Multi-class decoders were considered, as opposed to the binary decoders considered for the preclinical data set. LR fitted on the whole feature set using a LASSO procedure was found to surpass the other multi-class decoders (see **Appendix B**). It was therefore integrated into the MSLM recurrent gate.

Supervised training of the MSLM decoders was performed using complete training data sets $\{\mathbf{X}, \mathbf{Y}, \mathbf{z}\}$. As a prior was set on the gate and expert parameters, it corresponded to Maximum A Posteriori rather to Maximum Likelihood training. Similarly to the preclinical data set, the parameters of the SLM decoders were inferred from the MSLM parameters.

6.2.2.2 Post-processed WF

A Wiener filter was fitted via PLS regression on the training data set $\{\mathbf{X}, \mathbf{Y}\}$. Wold's R criterion was applied on the 6-fold cross-validated PRESS statistic to estimate the optimal number of PLS factors.

Trajectory post-processing was used to integrate NC support into the decoder. A logistic regression was trained to infer NC and IC states from the kinematic estimates $\hat{\mathbf{y}}^t$. Similarly to the MSLM, a Markovian hypothesis was used to limit erroneous state detections. The corresponding transition matrix was found by using the procedure presented in the case of the MSLM. The state probabilities $P(z^t = i|\hat{\mathbf{y}}^{1:t}), i = 1, \ldots, 6$ yielded by the dynamic LR state decoder were used to weight the Wiener estimates associated with each finger and the neural values corresponding to NC states. The dynamic logistic regression was identified on the training data set $\left\{\hat{\mathbf{Y}}, \mathbf{z}\right\}$. Similarly to the preclinical data set, the parameters of the pWF decoders were inferred from the MpWF parameters.

6.2.2.3 SKF

A SKF was implemented for dynamic combination of $K = 6$ Kalman filters, one specialized in NC periods and the 5 other ones in IC periods. Similarly to the preclinical data set, the SKF state variable $\mathbf{y}_{SKF}^t$ (i.e., response variable) was chosen as the finger position and velocity. The dimension of the neural features $\mathbf{x}_{SKF}^t$ corresponding to the last 100ms before the instant t was reduced before application of the SKF (see preliminary study in **Appendix C**). PLS regression between $\mathbf{x}_{SKF}^t$ and the $\mathbf{y}_{SKF}^t$ was used to identify an informative low-dimensional subspace. The subspace's dimension was chosen by 6-fold cross-validation on the training data set. A constant term was integrated into the SKF's neural features to account for modifications of the baseline activity during NC and $\text{IC}_i, i = 1, \ldots, 5$ states. OLS estimates of the transition $\mathbf{A}_k$ and emission $\mathbf{C}_k$ matrices were computed on the training data sets $\{\mathbf{X}_{k,SKF}, \mathbf{Y}_{k,SKF}\}$. ML estimates of the variance matrices $\mathbf{\Gamma}_k$ and $\mathbf{\Sigma}_k$ were estimated following [Aggarwal et al., 2013].

6.3 Conclusion

This chapter includes a description of the procedure used to extract features from both ECoG signals and limb trajectories. Details on the implementation of the Markov Switching Linear Model (MSLM) have been exposed for the two evaluation data sets, namely the publicly available preclinical and clinical data sets. Precisions on the unsupervised procedure used to identify several IC experts have been given for the considered expert and gate structures. The procedure completed to implement Markovian post-processed Wiener Filter (MpWF) and Switching Kalman Filter (SKF) decoders on the preclinical and clinical data set has been detailed. An insight on the approaches used to optimize the MSLM's gate and experts as well as to perform efficient dimensionality reduction before applying the SKF has additionally been presented. In particular, the respective relevance of generative and discriminative strategies for the modelling of the MSLM's gate emission distribution has been investigated. Next chapter consists in the presentation of the decoding performance of the MSLM for the considered pre-clinical and clinical data sets, along with of a comparison of its decoding efficiency with the ones obtained with the alternative decoders.

Chapter 7

Results

Contents

The performance of the Markov Switching Linear Model (MSLM) was assessed for two tasks, namely asynchronous mono-limb and asynchronous multi-limb decoding of preclinical and clinical ECoG data, respectively.

7.1 Preclinical data set

A first study focused on the utilization of the MSLM, Wiener Filter with Markovian post-processing (MpWF) and Switching Kalman Filter (SKF) decoders for the introduction of NC support into kinematic decoders. The MSLM decoding performance was compared to the MpWF and SKF respective performances on the test subsets of the preclinical data set's acquisition sessions.

A second study focused on complex movements. It was attempted to improve kinematic estimates during IC states by combining several IC experts. As three data sets only exhibited complex movements, this study was completed on a reduced subdural data set.

7.1.1 Integration of NC support into kinematic decoders

The MSLM decoding performance was compared to SKF and MpWF respective performances on the test subset of the preclinical data set's 24 acquisition sessions (2 and 2 monkeys, 8 and 16 sessions for the subdural and epidural data sets, respectively). The state detection performances of variants of the MSLM and

MpWF with static state detection (referred to as SLM and pWF, respectively) were additionally considered to assess the interest of the Markovian hypothesis to limit spurious activations. Post-hoc pairwise decoder comparisons [Benavoli et al., 2015] associated with the Friedman test with the significance level 0.05 [Demšar, 2006] were performed using the sign test with Bonferroni correction (see **section 5.3**).

7.1.1.1 State decoding performance

Table 7.1 shows the respective performances of the MSLM, SLM, MpWF, pWF and SKF decoders in terms of NC/IC classification accuracy. Both raw and synchronized state estimates were considered (maximum delay between synchronized samples $\tau_{max} = 200$ ms, ℓ^1-based synchronization). $\tau_{max} = 200$ ms was chosen because it had been used in an earlier study completed on the epidural data set [Eliseyev and Aksenova, 2014]. The corresponding p-values are gathered in **Table 7.2** for the decoders with dynamic state detection (MSLM, MpWF and SKF) and in **Table 7.3** for the MSLM and the alternative decoders with static state detection (SLM and pWF).

The AUCs corresponding to the MSLM were found to be significantly better than the AUCs obtained with all alternative dynamic decoders (i.e., the MpWF and the SKF) (see **Table 7.2**). The MSLM additionally permitted to significantly improve the ACC when compared to the SKF on the subdural and epidural data sets. The small number of sessions composing the subdural and epidural data sets may have limited the power of the statistical test used to assess the significance of performance differences. While the observed differences were not found to be significant, the MSLM was additionally associated with generally better TPR, FPR and ACC than the MpWF (median improvements of 1.2%, 26.6% and 1.8% for the subdural data set, respectively, and of 3.1%, 14.4%, 2.5% for the epidural data set, respectively). A significant improvement of the ACC by the MSLM when compared to the MpWF was also observed on synchronized state sequences of the subdural data set. When compared to the SKF, the MSLM improved the TPR, FPR and ACC by 1.3%, 43.2% and 3.4% on the subdural data set (median value over 8 sessions), respectively, and by 19.8%, 29.8% and 17.0% on the epidural data set (median value over 16 sessions), respectively.

No difference in the confusion-matrix-based performance indicators was observed between the MSLM and the SLM, i.e. its variant with static state detection (see **Table 7.3**).

Table 7.4 shows false activations/deactivations in terms of frequency and duration. P-values are gathered in **Table 7.5** and **Table 7.6**.

The decoders with dynamic state detection generally corresponded to false activations and deactivations with similar (not-significantly different) characteristics, except for the SKF which may have suffered from the lower state detection ability it exhibited in **Table 7.1**. The frequency of the false activations and false deactivations associated with the MpWF (SKF) was nevertheless higher that the one obtained

			TPR	FPR	ACC	AUC
Subdural (8 sessions)	raw	MSLM	0.90	0.06	0.91	0.97
		SLM	0.91	0.07	0.91	0.97
		MpWF	0.88	0.10	0.89	0.96
		pWF	0.87	0.10	0.89	0.95
		SKF	0.88	0.10	0.88	0.94
	synchronized	MSLM	0.92	0.03	0.94	0.98
		SLM	0.93	0.06	0.94	0.98
		MpWF	0.92	0.07	0.92	0.98
		pWF	0.92	0.09	0.93	0.97
		SKF	0.93	0.08	0.91	0.96
Epidural (16 sessions)	raw	MSLM	0.77	0.14	0.79	0.89
		SLM	0.76	0.15	0.79	0.89
		MpWF	0.72	0.17	0.80	0.88
		pWF	0.69	0.16	0.78	0.86
		SKF	0.67	0.21	0.66	0.76
	synchronized	MSLM	0.80	0.11	0.83	0.91
		SLM	0.80	0.11	0.82	0.91
		MpWF	0.77	0.13	0.83	0.90
		pWF	0.74	0.14	0.81	0.89
		SKF	0.71	0.18	0.70	0.79

Table 7.1: Preclinical data set, state classification performance, raw and synchronized state estimates ($\tau_{max} = 200$ ms, ℓ^1-based synchronization). Median False Positive Rate (FPR), True Positive Rate (TPR), accuracy (ACC) and the Area Under the Curve (AUC) are displayed for the MSLM, SLM, MpWF, pWF and SKF decoders. They were computed over 8 and 16 acquisition sessions for the subdural and epidural ECoG data sets, respectively.

			TPR	FPR	ACC	AUC
Subdural (8 sessions)	raw	MSLM/MpWF	0.07	0.07	0.07	0.008*
		MSLM/SKF	1	0.07	0.008*	0.008*
	synchronized	MSLM/MpWF	0.73	0.07	0.008*	0.008*
		MSLM/SKF	0.29	0.07	0.008*	0.008*
Epidural (16 sessions)	raw	MSLM/MpWF	0.45	0.21	0.02	< 0.001*
		MSLM/SKF	0.61	0.08	< 0.001*	< 0.001*
	synchronized	MSLM/MpWF	0.45	0.45	0.08	0.004*
		MSLM/SKF	0.45	0.08	< 0.001*	< 0.001*

Table 7.2: Preclinical data set, p-values for state classification performance, MSLM, MpWF and SKF decoders, raw and synchronized state estimates ($\tau_{max} = 200$ ms, ℓ^1-based synchronization). The significance of the differences between the decoders' respective performances was assessed using the Friedman test with the significance level $\alpha = 0.05$. Post-hoc comparisons were performed where appropriate using the sign test with Bonferroni correction, i.e. $\alpha_{Bonferroni} = 0.0167$. Significant differences are indicated by a star (*).

			TPR	FPR	ACC	AUC
Subdural (8 sessions)	raw	MSLM/SLM	0.45	0.29	0.07	0.29
		MSLM/pWF	0.29	0.29	0.008*	0.008*
	synchronized	MSLM/SLM	0.29	1	0.73	0.07
		MSLM/pWF	0.73	0.07	0.008*	0.008*
Epidural (16 sessions)	raw	MSLM/SLM	0.80	0.80	1	0.21
		MSLM/pWF	0.80	0.45	0.45	0.004*
	synchronized	MSLM/SLM	1	0.30	1	0.21
		MSLM/pWF	0.80	0.45	0.21	0.004*

Table 7.3: Preclinical data set, p-values for state classification performance, MSLM, SLM and pWF decoders, raw and synchronized state estimates ($\tau_{max} = 200$ ms, ℓ^1-based synchronization). The significance of the differences between the decoders' respective performances was assessed using the Friedman test with the significance level $\alpha = 0.05$. Post-hoc comparisons were performed where appropriate using the sign test with Bonferroni correction, i.e. $\alpha_{Bonferroni} = 0.0167$. Significant differences are indicated by a star (*).

with the MSLM by a median of 26.3% and 6.9% (62.3% and 32.4%), respectively, on the subdural data set, and by a median of 12.1% and 5.3% (43.8% and 16.5%), respectively, on the epidural data set. The improvement of the false activation frequencies was found to be significant for both data sets when the MSLM and the SKF were considered (**Table 7.5**).

The decoders with static gating were generally associated with more frequent false activations and deactivations than the MSLM. In particular, the MSLM's dynamic gate resulted in significantly fewer false activations/deactivations than the SLM on the epidural data set (median improvement of the false activations and deactivations of 32.7% and 16.7%, respectively, on the subdural data set, and of 31.9% and 26.0%, respectively, on the epidural data set). It additionally significantly surpassed the pWF in terms of false activation frequencies on both the subdural and epidural data sets (median improvement of the false activations and deactivations of 50.0% and 21.8%, respectively, on the subdural data set, and of 40.0% and 26.6%, respectively, on the epidural data set). The duration of the false activations and deactivations were, on the other hand, higher for the MSLM than for the SLM (median increase: 30.1% and 19.2% respectively, on the subdural data set, and 26.3% and 28.9% respectively, on the epidural data set).

		False activations		False deactivations	
		Frequency (1/s)	Duration (s)	Frequency (1/s)	Duration (s)
subdural	MSLM	0.05	0.30	0.08	0.31
	SLM	0.10	0.26	0.09	0.24
	MpWF	0.09	0.36	0.07	0.27
	pWF	0.15	0.28	0.10	0.27
	SKF	0.14	0.28	0.08	0.32
epidural	MSLM	0.14	0.35	0.19	0.39
	SLM	0.20	0.28	0.26	0.30
	MpWF	0.14	0.37	0.22	0.40
	pWF	0.20	0.29	0.27	0.32
	SKF	0.23	0.35	0.25	0.36

Table 7.4: Preclinical data set, false activation/deactivation median frequency and duration for MSLM, SLM, MpWF, pWF and SKF. False activations (respectively, false deactivations) are blocks of consecutive NC samples misclassified as IC samples (respectively, a block of IC samples mistaken for NC samples).

		False activations		False deactivations	
		Frequency	Duration	Frequency	Duration
Subdural (8 sessions)	MSLM/MpWF	0.07	0.008*	0.73	0.29
	MSLM/SKF	0.07	0.29	0.29	0.29
Epidural (16 sessions)	MSLM/MpWF	0.45	0.08	0.61	0.45
	MSLM/SKF	< 0.001*	0.21	0.45	0.45

Table 7.5: Preclinical data set, p-values for the false activation/deactivation frequency and duration associated with the MSLM, MpWF and SKF decoders. The significance of the differences between the decoders' respective performances was assessed using the Friedman test with the significance level $\alpha = 0.05$. Post-hoc comparisons were performed where appropriate using the sign test with Bonferroni correction, i.e. $\alpha_{Bonferroni} = 0.0167$. Significant differences are indicated by a star (*).

		False activations		False deactivations	
		Frequency	Duration	Frequency	Duration
Subdural (8 sessions)	MSLM/SLM	0.07	0.008*	0.07	0.008*
	MSLM/pWF	0.008*	0.07	0.22	0.29
Epidural (16 sessions)	MSLM/SLM	< 0.001*	< 0.001*	< 0.001*	< 0.001*
	MSLM/pWF	< 0.001*	< 0.001*	0.021	< 0.001*

Table 7.6: Preclinical data set, p-values for the false activation/deactivation frequency and duration associated with the MSLM, SLM and pWF decoders. The significance of the differences between the decoders' respective performances was assessed using the Friedman test with the significance level $\alpha = 0.05$. Post-hoc comparisons were performed where appropriate using the sign test with Bonferroni correction, i.e. $\alpha_{Bonferroni} = 0.0167$. Significant differences are indicated by a star (*).

7.1.1.2 IC decoding performance, known states

The relevance of the continuous IC kinematic models embedded in the hybrid decoders was then assessed. As state labels are traditionally hidden during decoder application, the performance indicators measured over IC states reflect both the quality of the IC continuous decoder and of the state decoder (a false negative for example results in a high ℓ^2-error). To decouple the impact of the state and continuous decoders, the MSLM, MpWF and SKF were applied on complete test data sets $\{\mathbf{X}, \mathbf{z}\}$, i.e. the state sequence was not hidden during application. Continuous performance indicators were then computed using exclusively the true IC samples. **Table 7.7** gathers the IC continuous models' decoding performance.

While no significant differences were observed between the MSLM's and MpWF's IC decoding ability on the subdural data set, the MSLM corresponded to significantly better NRMSE and NMAE than the MpWF on the epidural data set. A preliminary synchronization of the considered trajectories additionally disclosed a significantly higher PCC for the MSLM than for the MpWF on the epidural data set.

The MSLM and the SKF generally exhibited similar IC decoding abilities on both the subdural and epidural data sets. When raw (un-synchronized) subdural trajectories were considered, however, the MSLM significantly outperformed the SKF in terms of NRMSE and NMAE on the subdural data set.

An illustration of the decoded trajectories is shown in **Figure 7.1**.

			PCC	NRMSE	NMAE
subdural	raw	MSLM	0.60	0.77	0.77
		MpWF	0.61	0.79	0.74
		SKF	0.57	0.82	0.79
	synchronized	MSLM	0.66	0.71	0.67
		MpWF	0.67	0.74	0.66
		SKF	0.64	0.77	0.72
epidural	raw	MSLM	0.27	0.99	0.96
		MpWF	0.24	1.03	0.99
		SKF	0.33	0.98	0.94
	synchronized	MSLM	0.41	0.92	0.87
		MpWF	0.36	0.98	0.92
		SKF	0.42	0.92	0.85

Table 7.7: Preclinical data set, known state sequence. Decoding performance during IC states, with and without synchronization. The PCC, RMSE and NMAE associated with the MSLM, MpWF and SKF decoders were averaged over the 3 available axes within each session. Their median values were then computed over 8 and 16 sessions for the subdural and epidural data sets, respectively. Synchronization was completed with $\tau_{max} = 200$ms (ℓ^2-based synchronization for the PCC and NRMSE, ℓ^1-based synchronization for the NMAE).

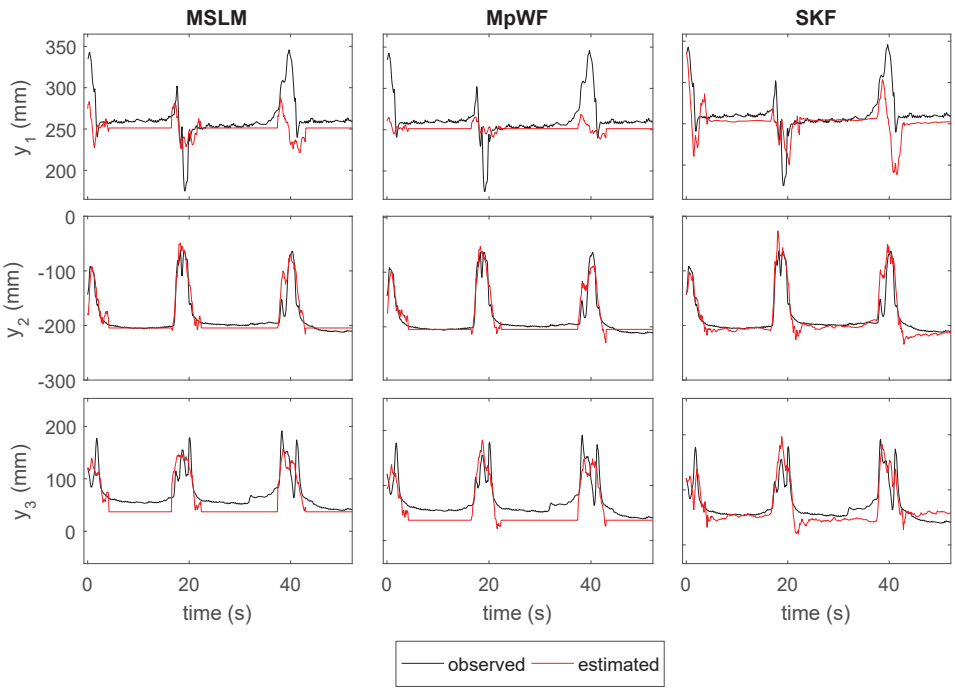

Figure 7.1: Preclinical data set, known state sequence. Example of observed and estimated trajectories (epidural data set, Monkey C). The projections of the monkey's wrist trajectory onto the horizontal axes (y_1 and y_2) and the vertical axis (y_3) are indicated in solid black lines. Red trajectories represent the estimates yielded by the MSLM, SKF and MpWF decoders.

			PCC	NMRSE	NMAE
subdural	raw	MSLM/MpWF	0.29	1	0.73
		MSLM/SKF	0.07	0.07	0.29
	synchronized	MSLM/MpWF	0.29	0.29	1
		MSLM/SKF	0.29	0.008*	0.008*
epidural	raw	MSLM/MpWF	1	< 0.001*	< 0.001*
		MSLM/SKF	0.08	0.80	0.80
	synchronized	MSLM/MpWF	< 0.001*	< 0.001*	< 0.001*
		MSLM/SKF	0.80	1	0.08

Table 7.8: Preclinical data set, known state sequence. P-values for IC decoding performance, raw and synchronized signals ($\tau_{max} = 200$ms, ℓ^2-based synchronization for the PCC and NRMSE, ℓ^1-based synchronization for the NMAE). The significance of the differences between the decoders' respective performances was assessed using the Friedman test with the significance level $\alpha = 0.05$. Post-hoc comparisons were performed using the sign test with Bonferroni correction, i.e. $\alpha_{Bonferroni} = 0.0167$. Significant differences are indicated by a star (*).

7.1.1.3 IC decoding performance, hidden states

The performance of the three decoders was then assessed over IC periods when state labels were hidden in the test data set. Results are gathered in **Table 7.9**, and corresponding p-values are displayed in **Table 7.10**. An example of decoded trajectories is presented in **Figure 7.2**.

The MSLM, MpWF and SKF generally presented similar decoding performances on the subdural data set. The MSLM nevertheless corresponded to PCC, NRMSE and NMAE higher by a median of 10%, 7% and 10% than the SKF's PCC, NMRSE and NMAE, respectively. This trend was confirmed by the significant difference between the MSLM's and SKF's NRMSE and NMAE when temporal errors were partly discarded, i.e. when synchronized trajectories were compared.

On the epidural data set, however, the MSLM significantly surpassed the MpWF in terms of NRMSE and NMAE. Synchronized performance indicators additionally permitted to disclose a significant improvement of the PCC when the MSLM rather than the MpWF was applied on the epidural data set. While the MSLM corresponded to NRMSE and NMAE higher by a median of 4% and 5% than the SKF, these differences were not found to be significant.

An additional comparison was completed to assess the interest of training the MSLM on IC samples exclusively in the case of hidden state sequences. An alternative MSLM was built by replacing the IC experts trained on IC samples by models trained on both NC and IC samples, i.e. by the linear models applied by the MpWF. The ability of each switching model to correctly deal with IC periods was compared (see **Table 7.11** for the median results, and **Table 7.12** for the corresponding p-values).

			PCC	NRMSE	NMAE
subdural	raw	MSLM	0.60	0.78	0.76
		MpWF	0.61	0.80	0.72
		SKF	0.56	0.86	0.81
	synchronized	MSLM	0.66	0.72	0.67
		MpWF	0.66	0.75	0.66
		SKF	0.61	0.81	0.74
epidural	raw	MSLM	0.23	1.04	1.00
		MpWF	0.23	1.07	1.03
		SKF	0.21	1.04	1.02
	synchronized	MSLM	0.32	0.98	0.93
		MpWF	0.30	1.04	0.98
		SKF	0.28	1.00	0.97

Table 7.9: Preclinical data set, hidden state sequence. Decoding performance during IC states, with and without synchronization. The PCC, RMSE and NMAE associated with the MSLM, MpWF and SKF decoders were averaged over the 3 available axes for each session. Their median values were then computed over 8 and 16 sessions for the subdural and epidural data sets, respectively. Synchronization was completed with $\tau_{max} = 200$ms (ℓ^2-based synchronization for the PCC and NRMSE, ℓ^1-based synchronization for the NMAE).

			PCC	NMRSE	NMAE
Subdural	raw	MSLM/MpWF	0.29	0.73	0.73
		MSLM/SKF	0.07	0.07	0.07
	synchronized	MSLM/MpWF	1	0.29	0.73
		MSLM/SKF	0.07	0.008*	0.008*
Epidural	raw	MSLM/MpWF	0.08	0.004*	< 0.001*
		MSLM/SKF	0.21	0.45	0.02
	synchronized	MSLM/MpWF	< 0.001*	< 0.001*	< 0.001*
		MSLM/SKF	0.21	0.45	0.08

Table 7.10: Preclinical data set, hidden state sequence. P-values for IC decoding performance, raw and synchronized trajectories ($\tau_{max} = 100$ms, ℓ^2-based synchronization for the PCC and NRMSE, ℓ^1-based synchronization for the NMAE). The significance of the differences between the decoders' respective performances was assessed using the Friedman test with the significance level $\alpha = 0.05$. Post-hoc comparisons were performed using the sign test with Bonferroni correction, i.e. $\alpha_{Bonferroni} = 0.0167$. Significant differences are indicated by a star (*).

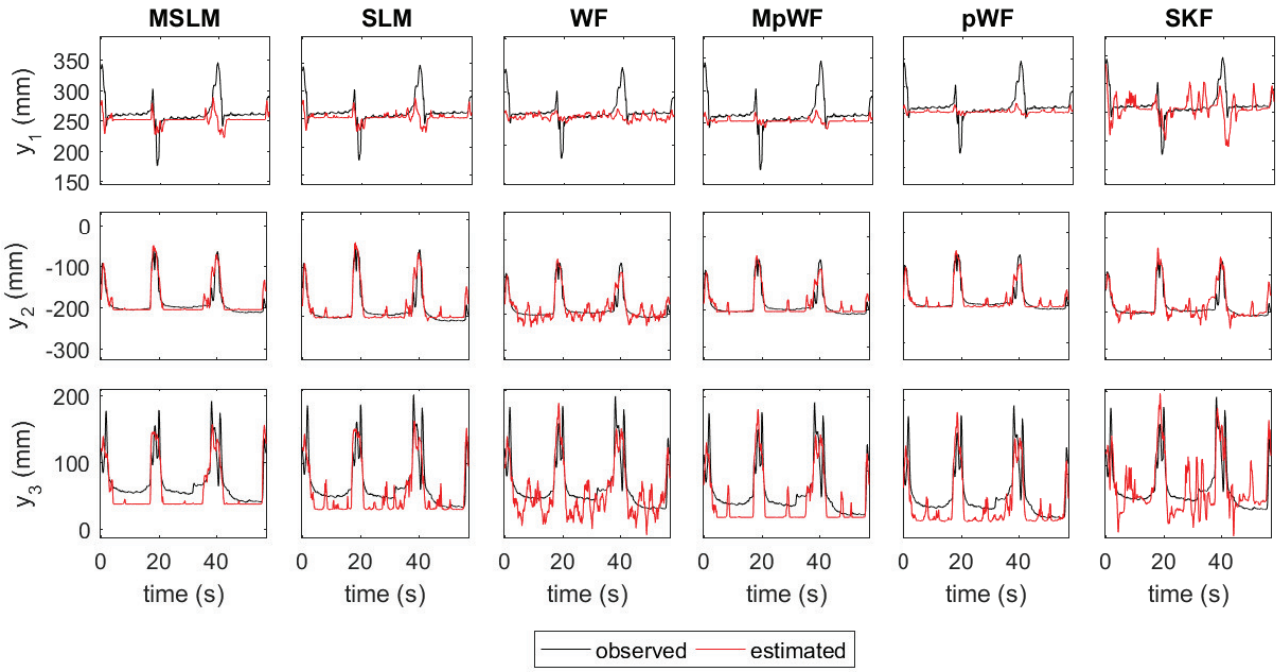

Figure 7.2: Preclinical data set, hidden state sequence. Example of observed and estimated trajectories for the epidural data set (Monkey C). The projections of the monkey's wrist trajectory onto the horizontal axes (y_1 and y_2) and the vertical axis (y_3) are indicated in solid black lines. Red trajectories represent the estimates yielded by the MSLM, SLM, Wiener Filter (WF), MpWF and SKF decoders.

In contrast with the indicators computed on IC samples when state sequences were known (see **Table 7.7** above), differences between IC experts and single models of the MpWF trained on both NC and IC samples were thus monitored on IC samples correctly classified by the MSLM gate. As a result, the single kinematic decoder associated with the MpWF was not penalized if its TPR was lower than the MSLM's TPR.

Consistently with the previously reported results, no difference was observed on the subdural data set between the MSLM's IC experts exclusively fitted on IC samples and the models trained on both NC and IC samples (i.e., the single models utilized by the MpWFs). On the epidural data set, however, the NRMSE and NMAE associated with models trained on both NC and IC samples were significantly higher than the ones corresponding to the MSLM' IC experts, i.e. on models exclusively trained on IC samples. A non significant median improvement of 6% was additionally found for the PCC, which was significantly improved when synchronized trajectories were considered.

		MSLM	PCC	NRMSE	NMAE
subdural	raw	IC expert training	0.60	0.78	0.76
		NC/IC expert training	0.61	0.80	0.73
	synchronized	IC expert training	0.66	0.72	0.67
		NC/IC expert training	0.66	0.75	0.66
epidural	raw	IC expert training	0.23	1.04	1.00
		NC/IC expert training	0.22	1.07	1.03
	synchronized	IC expert training	0.32	0.98	0.93
		NC/IC expert training	0.30	1.02	0.97

Table 7.11: Preclinical data set, hidden state sequence. Decoding performance of MSLMs over IC states with either a IC- or a NC/IC- trained IC expert, hidden state sequence, with and without synchronization. The PCC, RMSE and NMAE associated with both variants of the MSLM were averaged over the 3 available axes for each session. Their median values were then computed over 8 and 16 sessions for the subdural and epidural data sets, respectively. Synchronization was completed with $\tau_{max} = 200$ms (ℓ^2-based synchronization for the PCC and NRMSE, ℓ^1-based synchronization for the NMAE).

MSLM	IC expert training	MSLM	PCC	NMRSE	NMAE
Subdural	raw	IC / NC-IC	0.07	0.73	0.73
	synchronized	IC / NC-IC	0.73	0.07	0.29
Epidural	raw	IC / NC-IC	0.08	< 0.001*	< 0.001*
	synchronized	IC / NC-IC	< 0.001*	< 0.001*	< 0.001*

Table 7.12: Preclinical data set, hidden state sequence. P-values for the decoding performance of MSLMs over IC states with either a IC- or a NC/IC- trained IC expert, hidden state sequence, raw and synchronized trajectories ($\tau_{max} = 200$ms, ℓ^2-based synchronization for the PCC and NRMSE, ℓ^1-based synchronization for the NMAE). The significance of the differences between the decoders' respective performances was assessed using the sign test with the significance level $\alpha = 0.05$. Significant differences are indicated by a star (*).

7.1.1.4 Modality influence

An example of the influence of frequency, temporal and spatial features of the MSLM's IC expert and gate is shown in **Figure 7.3** (subdural data set, Monkey A, session 1). Contributions were assessed as the normalized summation of absolute values of models' coefficients along each modality. The WF corresponding modality influence is presented in **Figure 7.4** (subdural data set, Monkey A, session 1).

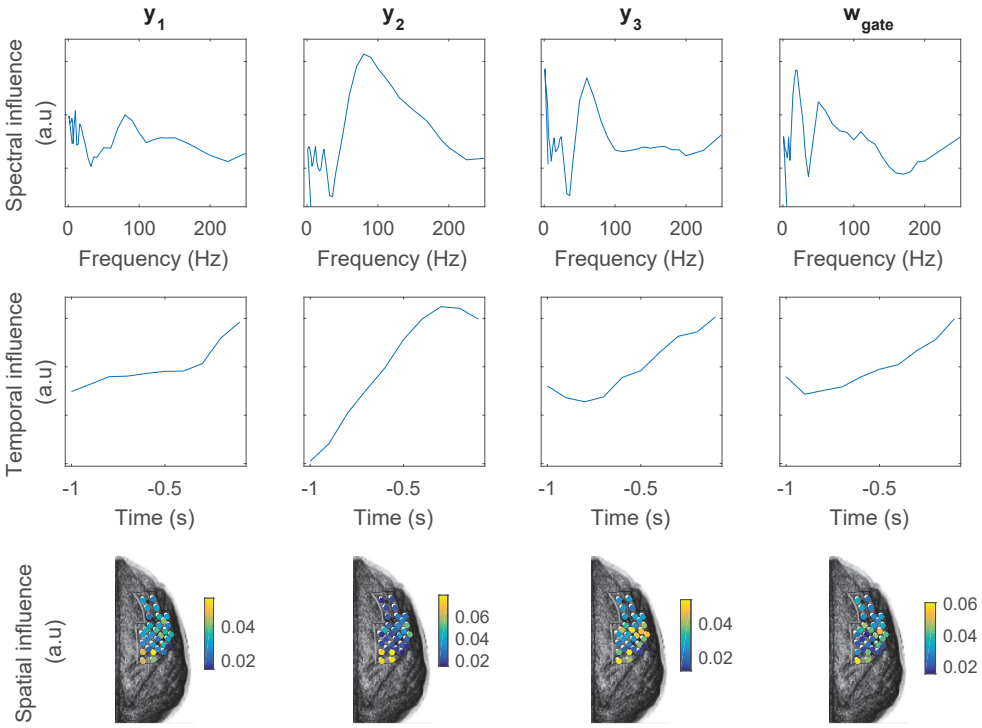

Figure 7.3: Preclinical data set, example of the average contribution of frequency, temporal and spatial modalities to the MSLM's IC expert and to the MSLM's gate (subdural data set, Monkey A, session 1). The vector $\mathbf{w}_{gate}$ parametrizes the PLS-based followed by the logistic regression such that $P(z^t = 1|\mathbf{x}^t) = \frac{1}{1+e^{-\mathbf{w}'_{gate}\mathbf{x}^t}}$. Contributions were assessed as the normalized summation of absolute values of models' coefficients along each modality. "a.u." refers to the utilization of arbitrary units.

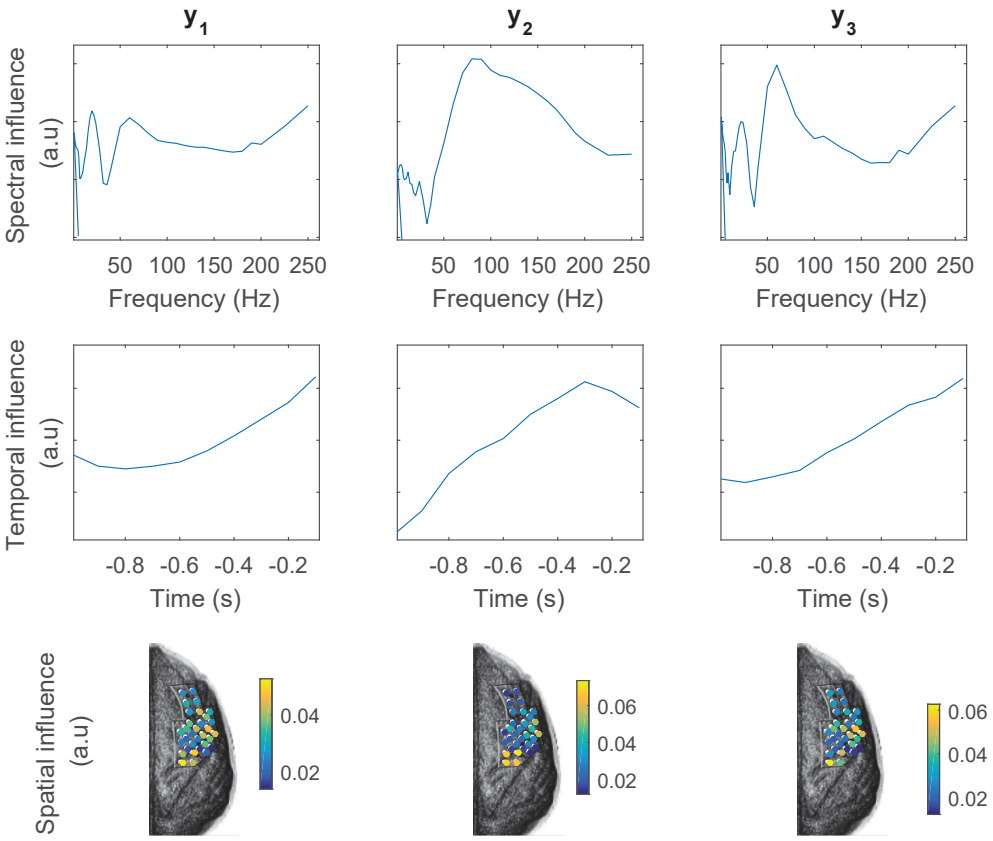

Figure 7.4: Preclinical data set, example of the average contribution of frequency, temporal and spatial modalities to the MpWF continuous decoder (subdural data set, Monkey A, session 1). Contributions were assessed as the normalized summation of absolute values of models' coefficients along each modality. "a.u." refers to the utilization of arbitrary units.

7.1.2 Multiple IC experts for kinematic reconstruction of complex movements

3 sessions of the subdural data set exhibit complex reaching movements (see **Chapter 4**). MSLM with $k = 2$ IC experts were trained to assess whether complex movements are better modelled using several IC experts. Because of this limited number of sessions, 10-fold cross-validation was performed on each of the 3 data sets. The sign test was applied on the 10 values obtained for each data set, and permitted to assess whether MSLMs with 1 or 2 IC experts performed similarly on each data set.

7.1.2.1 State decoding performance

Considering several IC states may result in a degradation of the state detection. A first analysis thus focused on state detection. 10-fold cross-validated state-related performance indicators are gathered in **Table 7.13**.

NC/IC state labels, where IC states gather IC_1 and IC_2 states, were not constrained during training because of labelling uncertainties, for example around transitions or for very short null-velocity periods during movements (see **Chapter 6**). Using several IC states resulted in a dramatic diminution of the FPR for the 3 considered sessions. By contrast, the TPR (and, consequently, the ACC) was diminished when compared with the supervised binary MSLM. As continuous performance indicators computed over IC samples were improved when a 3-state MSLM was used (see next paragraph), the TPR nevertheless seems irrelevant to assess decoding accuracy during IC periods. These results suggest that the MSLM may have advantageously applied the NC rather than IC expert on samples manually labelled as IC, for example at the beginning of IC periods if the IC expert was not able to properly model movement initiations, or for samples corresponding to short null-velocity periods within movements.

7.1.2.2 IC decoding performance

The IC decoding ability of single and multiple IC experts was then assessed. 10-fold cross-validated performance indicators between raw and synchronized observed and estimated trajectories are gathered in **Table 7.14**. Using multiple experts generally permitted to improve the modelling of wrist kinematic during IC states. A significant improvement of the MAE was observed for the three data sets. The PCC was additionally significantly higher for two out of three data sets, and the NRMSE for one data set. An illustration of the trajectory estimates yielded by single and multiple ($k = 2$) IC experts is shown in **Figure 7.5**.

	Session	decoder	TPR	FPR	ACC
raw	1	MSLM, 1 IC state	0.90	0.09	0.10
		MSLM, 2 IC states	0.83	0.05	0.12
	4	MSLM, 1 IC state	0.88	0.13	0.88
		MSLM, 2 IC states	0.76	0.01	0.82
	5	MSLM, 1 IC state	0.85	0.08	0.89
		MSLM, 2 IC states	0.67	0.03	0.88
synchronized	1	MSLM, 1 IC state	0.92	0.05	0.07
		MSLM, 2 IC states	0.87	0.04	0.10
	4	MSLM, 1 IC state	0.91	0.08	0.91
		MSLM, 2 IC states	0.79	0.01	0.85
	5	MSLM, 1 IC state	0.86	0.07	0.91
		MSLM, 2 IC states	0.70	0.02	0.90

Table 7.13: Preclinical data set, sessions exhibiting complex movements. Cross-validated state classification performance, raw and synchronized state estimates ($\tau_{max} = 200$ ms, ℓ^1-based synchronization). 10-fold cross-validated median False Positive Rate (FPR), True Positive Rate (TPR) and the accuracy (ACC) are displayed for the MSLM with one and two IC states.

	session	decoder	PCC	RMSE	MAE
raw	1	1 IC state	0.58	0.83	0.84
		2 IC states	0.62	0.80	0.76
	4	1 IC state	0.56	0.85	0.84
		2 IC states	0.60	0.82	0.78
	5	1 IC state	0.69	0.75	0.70
		2 IC states	0.70	0.73	0.70
synchronized	1	1 IC state	0.64	0.79	0.78
		2 IC states	0.70	0.72	0.66
	4	1 IC state	0.59	0.80	0.77
		2 IC states	0.65	0.75	0.70
	5	1 IC state	0.74	0.70	0.64
		2 IC states	0.76	0.67	0.59

Table 7.14: Preclinical data set, hidden state sequence, sessions exhibiting complex movements. The IC cross-validated decoding performance of MSLMs with 1 or 2 IC states was measured on the three subdural data sets composed of complex movements. The PCC, NRMSE and NMAE were averaged over the 3 considered axes. Their 10-fold cross-validated median value is displayed for MSLMs with 1 or 2 IC states. Performance indicators were computed for both raw and synchronized observed and estimated trajectories. Synchronization was completed with $\tau_{max} = 200$ms (ℓ^2-based synchronization for the PCC and NRMSE, ℓ^1-based synchronization for the NMAE).

	session	MSLM decoder	PCC	NMRSE	NMAE
raw	1	1 IC state / 2 IC states	0.02*	0.1	0.02*
	4	1 IC state / 2 IC states	0.02*	0.02*	0.002*
	5	1 IC state / 2 IC states	0.3	0.8	0.02*
synchronized	1	1 IC state / 2 IC states	0.02*	0.02*	0.02*
	4	1 IC state / 2 IC states	0.02*	0.02*	0.002*
	5	1 IC state / 2 IC states	0.1	0.8	0.02*

Table 7.15: Preclinical data set, sessions exhibiting complex movements, hidden state sequence. P-value for IC cross-validated decoding performance of MSLMs with 1 or 2 IC states($\tau_{max} = 100$ms, ℓ^2-based synchronization for the PCC and NRMSE, ℓ^1-based synchronization for the NMAE). The significance of the differences between the decoders' respective performances was assessed using the sign test with the significance level $\alpha = 0.05$. Significant differences are indicated by a star (*).

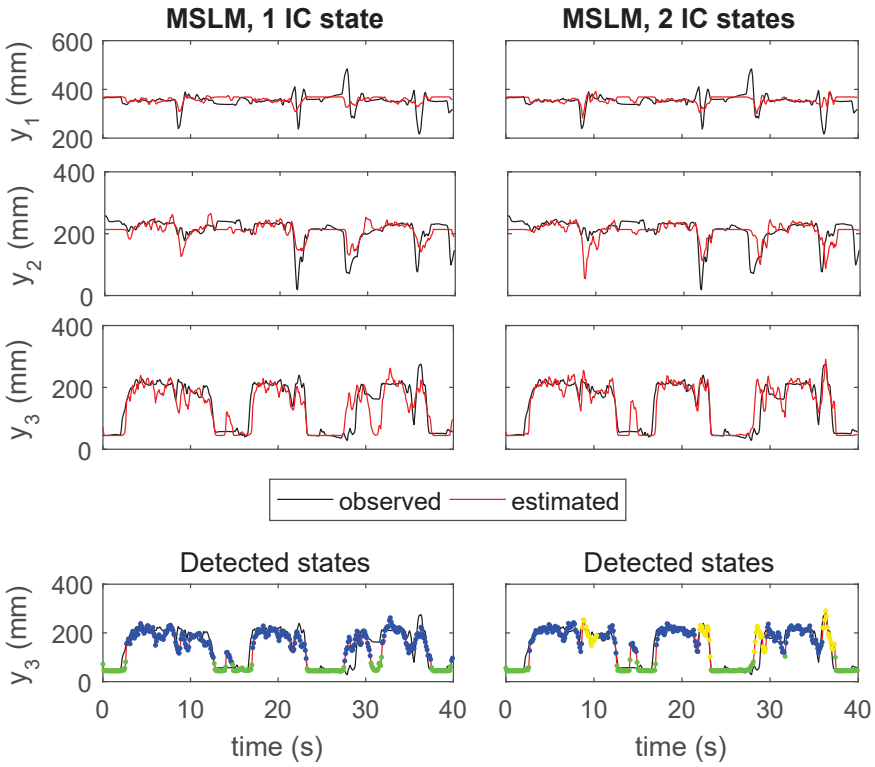

Figure 7.5: Preclinical data set, sessions exhibiting complex movements, example of observed and estimated trajectories (Monkey A). The projections of the monkey's wrist trajectory onto the horizontal axes (y_1 and y_2) and the vertical axis (y_3) are indicated in solid black lines. Red trajectories represent the estimates yielded by MSLM decoders with 1 and 2 IC experts, respectively. Detected NC states are represented with green dots, IC states with blue and yellow dots.

7.2 Clinical data set

The clinical data set was used to assess the respective performances of the MSLM, MpWF and SKF decoders for asynchronous multi-limb decoding. The MSLM decoding efficiency was compared to the SKF and MpWF respective efficiencies on the test data set of the 9 acquisition sessions of the clinical data set (9 subjects, one session per subject). The state detection performance of a SLM and a pWF was additionally investigated to evaluate the interest of the dynamic state detection performed by the MSLM. A maximal lag $\tau_{max} = 100$ ms was chosen for synchronization because the frequency content of the clinical (finger) trajectories was higher than the one associated with preclinical (wrist) trajectories. Using $\tau_{max} = 200$ ms may thus have artificially improved the quality of the kinematic estimates because of a too flexible synchronization.

7.2.0.1 State decoding performance

Table 7.16 shows the respective performances of the decoders in terms of confusion-matrix-based indicators. The corresponding p-values are gathered in **Table 7.17** for the decoders with dynamic state detection (MSLM, MpWF and SKF) and in **Table 7.18** for the MSLM and the alternative decoders with static state detection (SLM and pWF).

The ACCs corresponding to the MSLM were found to be significantly better than the ACCs obtained with the alternative dynamic decoders. MSLMs' ACCs were higher by a median of 5% and 49% than the MpWFs' and SKFs' ACC, respectively. While this improvement was not found to be significant, the FPR associated with the MSLM was higher than the one corresponding to the MpWF by a median of 33%. Although the MSLM's CTPR was significantly better than the one associated with the pWF, it wasn't the case for the SLM and for the MpWF. The Markovian hypothesis used by the MpWF may have permitted to compensate for a possible intrinsic lower decoding ability of the post-processed operator which is suggested by the low CTPR of the pWF.

While the SLM's CTPR was 7% lower than the MSLM's one (for a similar TPR), no significant differences were observed between the MSLM and SLM in terms of ACC and CTPR.

Table 7.19 shows false activations/deactivations in terms of occurrence number and duration. The corresponding p-values are gathered in **Tables 7.20** and **7.21**. The MSLM significantly outperformed the MpWF and SKF in terms of false activation and/or false deactivation frequencies. While the difference was not found significant, the SLM corresponded to false deactivations which were 60% more frequent than the MSLM. This observation seems to support the idea that dynamic state detection may be advantageously used to limit short spurious activations or deactivations of a BCI system.

		TPR	FPR	ACC	CTPR
raw	MSLM	0.91	0.16	0.87	0.70
	SLM	0.87	0.10	0.86	0.65
	MpWF	0.91	0.34	0.82	0.67
	pWF	0.85	0.28	0.80	0.61
	SKF	0.64	0.20	0.57	0.20
synchronized	MSLM	0.94	0.12	0.90	0.70
	SLM	0.90	0.07	0.90	0.66
	MpWF	0.94	0.27	0.86	0.68
	pWF	0.97	0.22	0.84	0.62
	SKF	0.66	0.19	0.60	0.22

Table 7.16: Clinical data set, state classification performance, raw and synchronized signals. Median False Positive Rate (FPR), True Positive Rate (TPR), accuracy (ACC) and Correct TP Ratio (CTPR) are displayed for the MSLM, SLM, MpWF, pWF and SKF decoders on the clinical data set (median value over 9 sessions). Synchronized performance indicators are additionally exposed. True and estimated switching states $\mathbf{y}$ and $\hat{\mathbf{y}}$ were synchronized (see **section 5.2.4**) before computing the TPR, FPR, ACC and CTPR ($\tau_{max} = 100$ms, median value over 9 sessions).

		TPR	FPR	ACC	CTPR
raw	MSLM/MpWF	0.18	0.004*	0.004*	1
	MSLM/SKF	0.004*	0.51	0.004*	0.004*
synchronized	MSLM/MpWF	0.008*	0.004*	0.004*	1
	MSLM/SKF	0.004*	0.51	0.004*	0.004*

Table 7.17: Clinical data set, p-values for state classification performance, MSLM, MpWF and SKF decoders, raw and synchronized signals ($\tau_{max} = 100$ms). The significance of the differences between the decoders' respective performances was assessed using the Friedman test with the significance level $\alpha = 0.05$. Post-hoc comparisons were performed using the sign test with Bonferroni correction, i.e. $\alpha_{Bonferroni} = 0.0167$. Significant differences are indicated by a star (*).

		TPR	FPR	ACC	CTPR
raw	MSLM/SLM	0.18	0.004*	0.51	0.04
	MSLM/pWF	0.18	0.04	0.004*	0.004*
synchronized	MSLM/SLM	0.18	0.004*	0.51	0.04
	MSLM/pWF	0.18	0.04	0.004*	0.04

Table 7.18: Clinical data set, p-values for state classification performance, MSLM, SLM and pWF decoders, raw and synchronized signals (τ_{max} = 100ms). The significance of the differences between the decoders' respective performances was assessed using the Friedman test with the significance level $\alpha = 0.05$. Post-hoc comparisons were performed using the sign test with Bonferroni correction, i.e. $\alpha_{Bonferroni} = 0.0167$. Significant differences are indicated by a star (*).

	False activations		False deactivations	
	Frequency (1/s)	Duration (s)	Frequency (1/s)	Duration (s)
MSLM	0.07	0.26	0.07	0.20
SLM	0.06	0.17	0.18	0.20
MpWF	0.16	0.24	0.10	0.20
pWF	0.19	0.16	0.33	0.16
SKF	0.08	0.24	0.40	0.34

Table 7.19: Clinical data set, false activation/deactivation median frequency and duration for MSLM, MpWF and SKF decoders. False activations (respectively, false deactivations) are blocks of consecutive NC samples misclassified as IC samples (respectively, a block of IC samples mistaken for NC samples.

	False activations		False deactivations	
	Frequency	Duration	Frequency	Duration
MSLM/MpWF	0.004*	1	0.004*	0.04
MSLM/SKF	0.28	1	0.004*	0.004*

Table 7.20: Clinical data set, p-values for the false activation/deactivation frequency and duration associated with the MSLM, MpWF and SKF decoders. The significance of the differences between the decoders' respective performances was assessed using the Friedman test with the significance level $\alpha = 0.05$. Post-hoc comparisons were performed where appropriate using the sign test with Bonferroni correction, i.e. $\alpha_{Bonferroni} = 0.0167$. Significant differences are indicated by a star (*).

	False activations		False deactivations	
	Frequency	Duration	Frequency	Duration
MSLM/SLM	0.69	0.008*	0.18	0.51
MSLM/pWF	0.004*	0.18	0.18	0.04

Table 7.21: Clinical data set, p-values for the false activation/deactivation frequency and duration associated with the MSLM, SLM and pWF decoders. The significance of the differences between the decoders' respective performances was assessed using the Friedman test with the significance level $\alpha = 0.05$. Post-hoc comparisons were performed where appropriate using the sign test with Bonferroni correction, i.e. $\alpha_{Bonferroni} = 0.0167$. Significant differences are indicated by a star (*).

7.2.0.2 IC decoding performance, known states

The relevance of the continuous IC kinematic models embedded in the hybrid decoders was first assessed in the case where switching state values were known in the test data sets. Similarly to the study led on the preclinical data set, the MSLM, MpWF and SKF were applied on complete test data sets $\{\mathbf{X}, \mathbf{z}\}$, i.e. the state sequence was not hidden during application. Continuous performance indicators were then computed over the true IC samples exclusively. **Table 7.22** shows the IC continuous models decoding performance, computed for both desynchronized and synchronized trajectories ($\tau_{max} = 100$ms, ℓ^2-based synchronization for the PCC and NRMSE, ℓ^1-based synchronization for the NMAE). P-values are gathered in **Table 7.23**.

The MSLM IC experts significantly outperformed the MpWF continuous models in terms of NRMSE and NMAE (median improvements of 26% and 22%, respectively). While the PCC was not found to be significantly different because of the presence of an outlier in the small data set, it was improved by a median value of 13%. When synchronized performances indicators were considered, the PCC associated with the MSLM's IC experts became significantly better than the one obtained with the MpWF. Similar results were obtained when the SKF IC experts were compared to the MSLM's ones, i.e. the MSLM IC experts surpassed both the MpWF continuous models and the SKF's experts. Examples of decoded trajectories are shown in **Figures 7.6** and **7.7**.

		PCC	NRMSE	NMAE
raw	MSLM	0.50	0.97	0.86
	MpWF	0.40	1.24	1.09
	SKF	0.31	1.39	1.27
synchronized	MSLM	0.68	0.80	0.65
	MpWF	0.47	1.22	1.05
	SKF	0.43	1.27	1.16

Table 7.22: Clinical data set, known state sequence. Decoding performance during IC states, with and without synchronization. Synchronization was completed with $\tau_{max} = 100$ms (ℓ^2-based synchronization for the PCC and NRMSE, ℓ^1-based synchronization for the NMAE). Within each acquisition session, PCC, NRMSE and NMAE were averaged over the 5 fingers. Median PCC, NRMSE and NMAE are displayed for the MSLM, MpWF and SKF decoders on the clinical data set (median value over 9 sessions).

		PCC	NMRSE	NMAE
raw	MSLM/MpWF	0.04	0.004*	0.004*
	MSLM/SKF	0.04	0.004*	0.004*
synchronized	MSLM/MpWF	0.004*	0.004*	0.004*
	MSLM/SKF	0.004*	0.004*	0.004*

Table 7.23: Clinical data set, known state sequence. P-values for IC decoding performance for raw and synchronized signals ($\tau_{max} = 100$ms, ℓ^2-based synchronization for the PCC and NRMSE, ℓ^1-based synchronization for the NMAE). Within each acquisition session, PCC, NRMSE and NMAE were averaged over the 5 fingers. The significance of the differences between the decoders' respective performances was assessed using the Friedman test with the significance level $\alpha = 0.05$. Post-hoc comparisons were performed using the sign test with Bonferroni correction, i.e. $\alpha_{Bonferroni} = 0.0167$. Significant differences are indicated by a star (*).

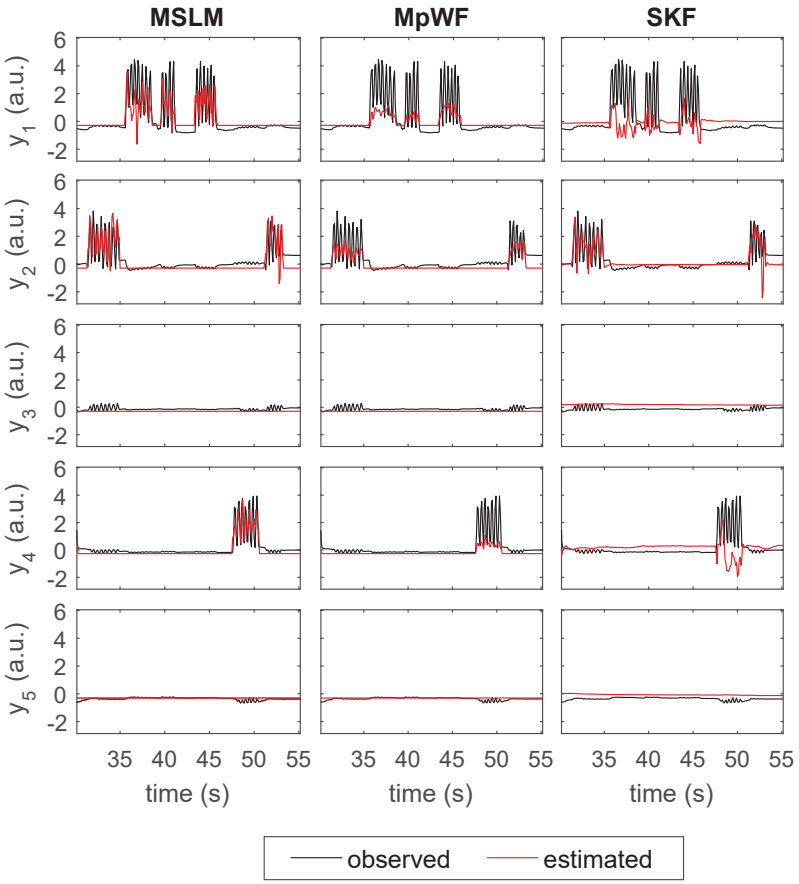

Figure 7.6: Clinical data set, known state sequence. Example of observed and estimated finger trajectories (subject zt). The tracked finger positions are indicated in solid black lines. Red trajectories represent the estimates yielded by the MSLM, MpWF and SKF decoders. "a.u." refers to "arbitrary unit".

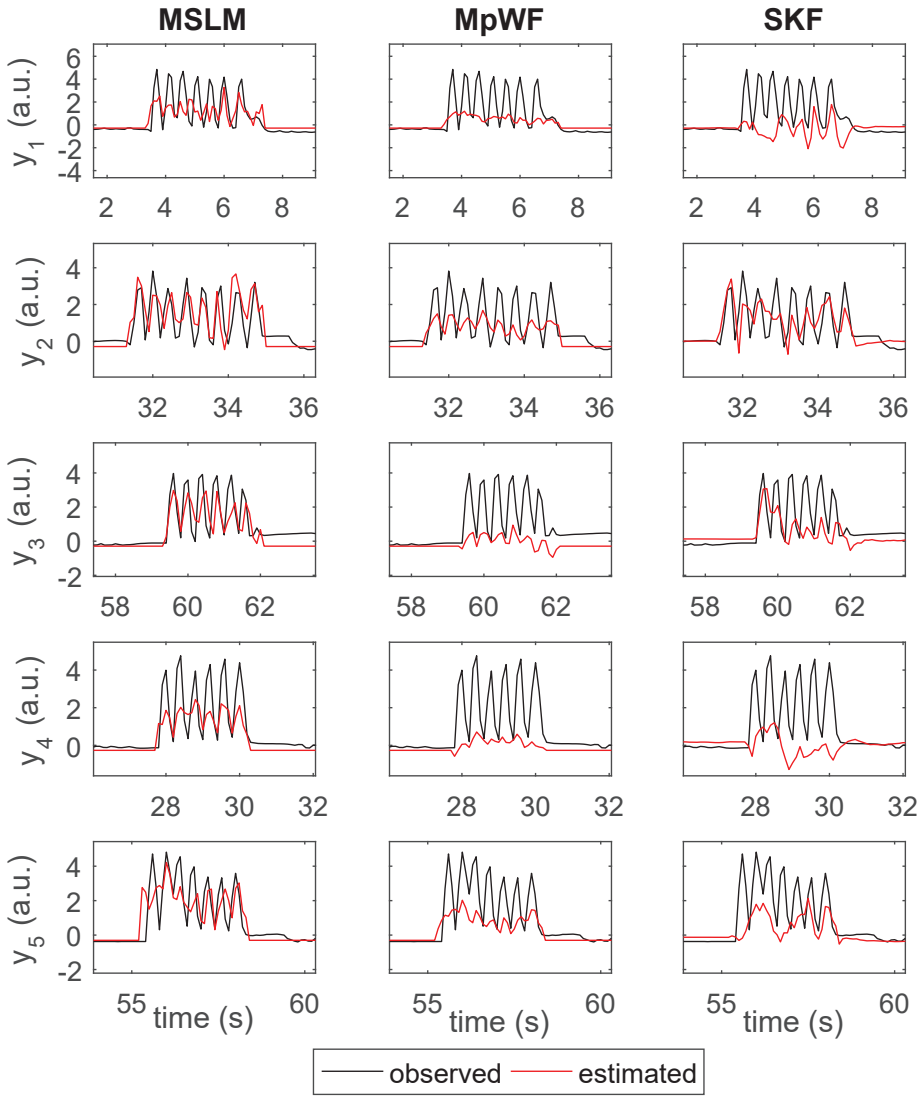

Figure 7.7: Clinical data set, known state sequence, focus on the IC states. Example of observed and estimated finger trajectories (subject zt). The tracked finger positions are indicated in solid black lines. Red trajectories represent the estimates yielded by the MSLM, MpWF and SKF decoders. "a.u." refers to "arbitrary unit".

7.2.0.3 IC decoding performance, hidden states

IC decoding accuracy was then measured in the case where the values taken by the switching state were hidden in the test data set. Results are shown in **Table 7.24**. The corresponding p-values are displayed in **Table 7.25**.

The MSLM significantly outperformed the WF in terms of NRMSE and NMAE (minus 9% and 12%, respectively). The MSLM was additionally associated with a PCC higher by a median of 7%. The SKF was significantly surpassed by the MSLM for all performance indicators for both raw (un-synchronized) and synchronized signals. An example of decoded trajectories is presented on **Figure 7.8**.

		PCC	NRMSE	NMAE
raw	MSLM	0.29	1.21	1.04
	MpWF	0.27	1.34	1.17
	SKF	0.12	1.62	1.46
synchronized	MSLM	0.42	1.11	0.90
	MpWF	0.31	1.29	1.10
	SKF	0.12	1.60	1.44

Table 7.24: Clinical data set, hidden state sequence. Decoding performance during IC states, with and without synchronization. Synchronization was completed with $\tau_{max} = 100$ms (ℓ^2-based synchronization for the PCC and NRMSE, ℓ^1-based synchronization for the NMAE). Within each acquisition session, PCC, NRMSE and NMAE were averaged over the 5 fingers. Median PCC, NRMSE and NMAE are displayed for the MSLM, MpWF and SKF decoders on the clinical data set (median value over 9 sessions).

		PCC	NMRSE	NMAE
raw	MSLM/MpWF	0.51	0.004*	0.004*
	MSLM/SKF	0.004*	0.004*	0.004*
synchronized	MSLM/MpWF	0.04	0.004*	0.004*
	MSLM/SKF	0.004*	0.004*	0.004*

Table 7.25: Clinical data set, hidden state sequence. P-values for IC decoding performance for raw and synchronized signals ($\tau_{max} = 100$ms, ℓ^2-based synchronization for the PCC and NRMSE, ℓ^1-based synchronization for the NMAE). Within each acquisition session, PCC, NRMSE and NMAE were averaged over the 5 fingers. The significance of the differences between the decoders' respective performances was assessed using the Friedman test with the significance level $\alpha = 0.05$. Post-hoc comparisons were performed using the sign test with Bonferroni correction, i.e. $\alpha_{Bonferroni} = 0.0167$. Significant differences are indicated by a star (*).

Similarly to the preclinical data set, an additional comparison was completed to decouple the impact of the state and kinematic decoders on the decoding accuracy over hidden IC states. An alternative MSLM was built by replacing IC experts trained on IC_i samples by models trained on both NC and IC samples, i.e. by the linear models applied by the MpWF. The ability of each switching model to correctly deal with IC periods was compared (see **Table 7.26** for the median results, and **Table 7.27** for the corresponding p-values). In contrast with the indicators computed on true, known IC samples, the decoding performance is here compared on IC_i samples correctly classified as IC by the MSLM gate. As a result, the model fitted on both NC and IC (i.e., embedded in the MpWF) is not penalized by IC observations erroneously classified as NC by both the MSLM and MpWF when it is compared to the MSLM's IC experts.

Using MSLM experts trained on IC samples exclusively permitted to significantly improve all performance indicators (median improvements of 15%, 11% and 12% for the PCC, NRMSE and NMAE, respectively). Significant median improvements of 32%, 16% and 21% were obtained when synchronized signals were compared (τ_{max} = 100ms, i.e. 1 sample), suggesting that the MSLM exhibits slight temporal uncertainties.

		PCC	NRMSE	NMAE
raw	MSLM-IC	0.29	1.21	1.04
	MSLM-full	0.26	1.35	1.18
synchronized	MSLM-IC	0.42	1.11	0.90
	MSLM-full	0.29	1.32	1.13

Table 7.26: Clinical data set, hidden state sequence. MSLM IC decoding performance with experts fitted on IC samples only ("MSLM-IC") or on both NC and IC samples ("MSLM-full"). Within each acquisition session, PCC, NRMSE and NMAE were averaged over the 5 fingers. Median PCC, NRMSE and NMAE are displayed for the MSLM-IC and MSLM-full decoders on the clinical data set (median value over 9 sessions).

		PCC	NMRSE	NMAE
raw	MSLM-IC/MSLM-full	0.04*	0.004*	0.004*
synchronized	MSLM-IC/MSLM-full	0.004*	0.004*	0.004*

Table 7.27: Clinical data set, p-values for IC decoding performance for raw and synchronized signals. Within each acquisition session, PCC, NRMSE and NMAE were averaged over the 5 fingers. The significance of the differences between the decoders' respective performances was assessed using the sign test with the significance level $\alpha = 0.05$. Significant differences are indicated by a star (*).

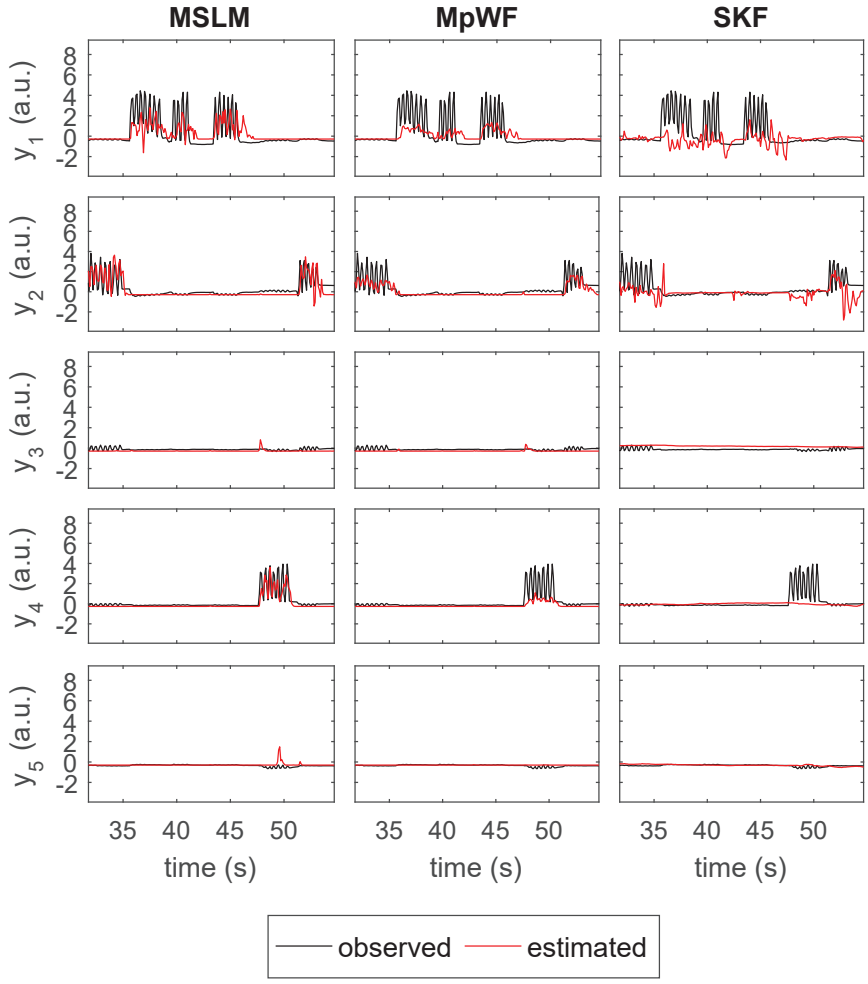

Figure 7.8: Clinical data set, hidden state sequence. Example of observed and estimated finger trajectories (subject zt). The tracked finger positions are indicated in solid black lines. Red trajectories represent the estimates yielded by the MSLM, MpWF and SKF decoders. "a.u." refers to "arbitrary unit".

7.2.0.4 Modality influence

Figures 7.10 and **7.9** illustrate the influence of frequency, temporal and spatial features of the IC experts and the gating network, respectively (subject "zt", 1 session). **Figure 7.11** corresponds to the Wiener filter embedded in the MpWF. Contributions were assessed as the normalized summation of absolute values of models' coefficients along each modality.

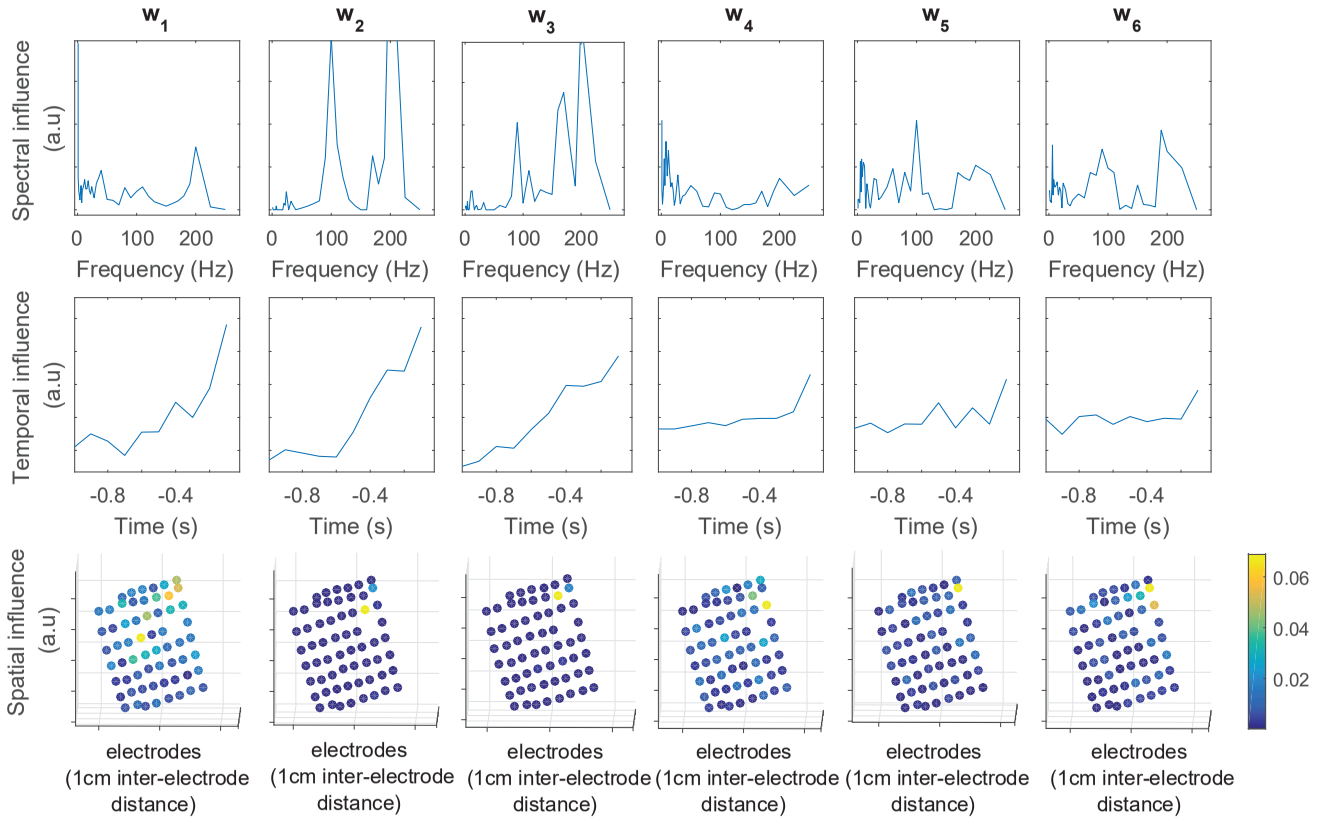

Figure 7.9: Clinical data set, example of the average contribution of frequency, temporal and spatial modalities to the MSLM's gate (subject zt, one session). The vector set $\{\mathbf{w}_1, \ldots, \mathbf{w}_6\}$ parametrizes the 6-class multinomial logistic regression such that $P(z^t = i|\mathbf{x}^t) = \frac{e^{\mathbf{w}_i' \mathbf{x}^t}}{\sum_j e^{\mathbf{w}_j' \mathbf{x}^t}}$. Contributions were assessed as the normalized summation of absolute values of each vector $\mathbf{w}_i$ along each modality. "a.u." refers to the utilization of arbitrary units.

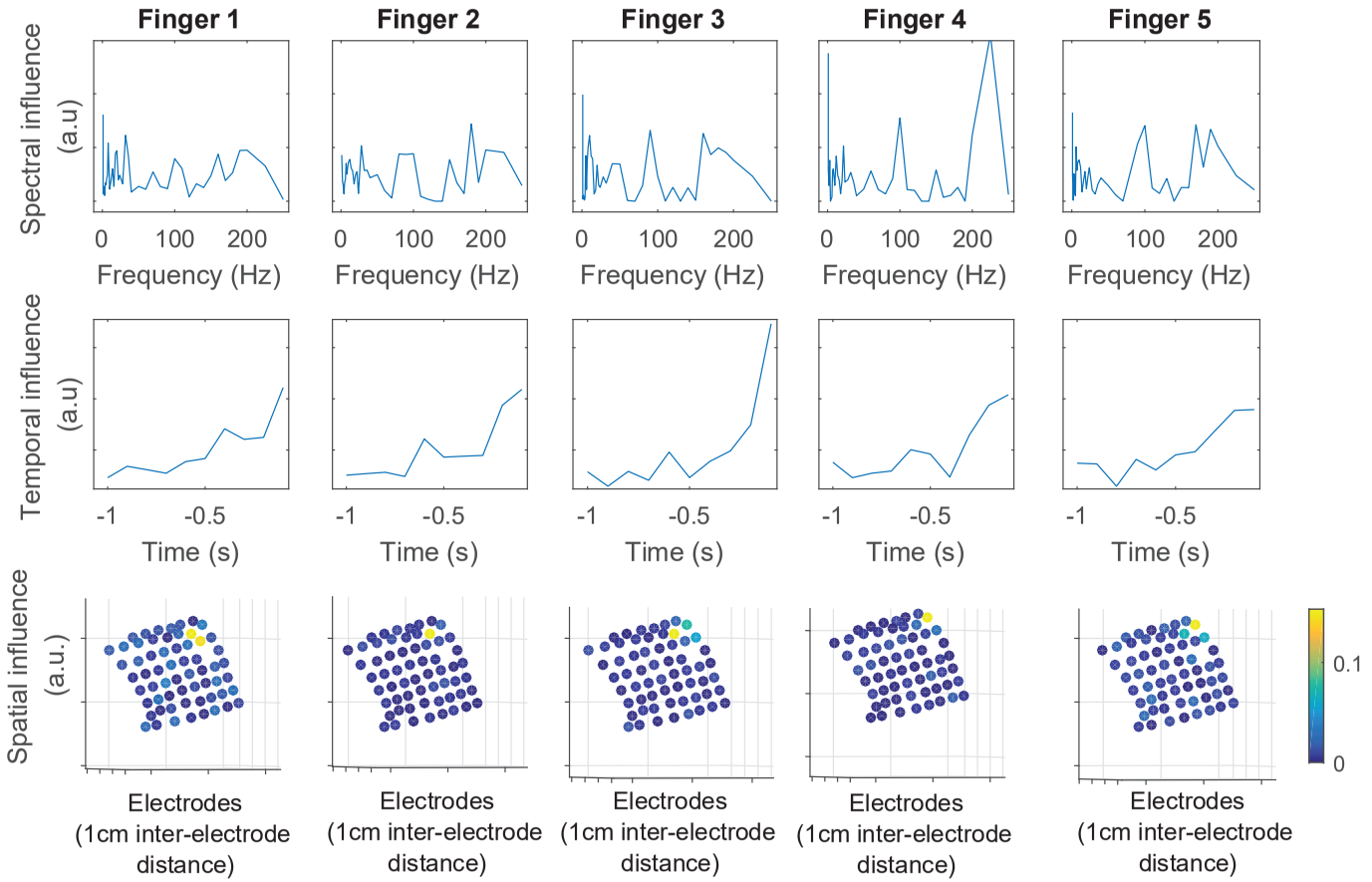

Figure 7.10: Clinical data set, example of the average contribution of frequency, temporal and spatial modalities to the MSLM's IC experts (subject zt, one session). Contributions were assessed as the normalized summation of absolute values of the considered expert's coefficients along each modality. "a.u." refers to the utilization of arbitrary units.

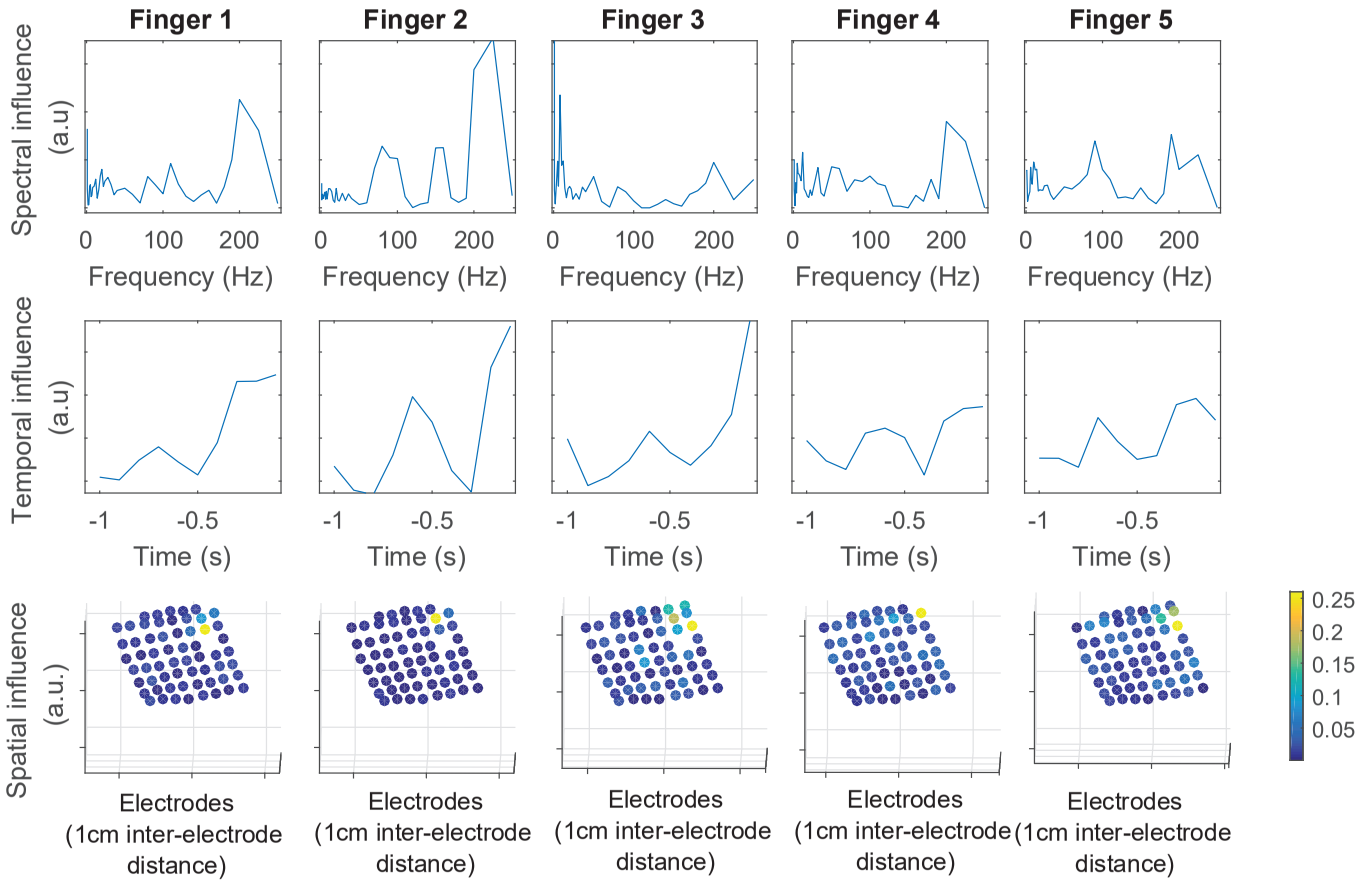

Figure 7.11: Clinical data set, example of the average contribution of frequency, temporal and spatial modalities to the MpWF linear model (subject zt, one session). Contributions were assessed as the normalized summation of absolute values of models' coefficients along each modality. "a.u." refers to the utilization of arbitrary units.

7.3 Discussion

The present results permit to assess the efficiency of the MSLM for accurate asynchronous mono-limb and sequential multi-limb decoding. The relevance of the three key features of the MSLM, namely switching modelling, discriminative modelling and dynamic gating, is of particular interest.

The pertinence of switching decoding, which is particularly related to the potential existence of context-specific modifications in a model of dependence between neural features and kinematic parameters, has been suggested in several MUA/SUA BCI studies [Kim et al., 2006c] [Ifft et al., 2013] [Velliste et al., 2014]. It was observed in [Velliste et al., 2014] that neural patterns generated during NC states did not fit a linear model trained on IC periods. The models permitting to decode bimanual movements from MUA/SUA in monkeys were found in [Ifft et al., 2013] to be different from the models independently trained for the decoding of unimanual left and right movements. These findings seemed to indicate the existence of an unimanual and a bimanual control modes associated with different models between neural features and kinematic parameters. In [Yu et al., 2007], switches between goal-specific models permitted to improve the decoding of reaching movements executed by monkeys. Different models were associated with specific phases of reaching movements in [Kim et al., 2003] and [Kang et al., 2012], and switches between emissions models were exploited in [Wu et al., 2004] to improve Kalman-based estimation of kinematic parameters from MUA/SUA signals. The interest of multi-model approaches has additionally been emphasized in recent reviews or studies on MUA/SUA signals [Rouse and Schieber, 2015] [Kao et al., 2017]. MUA/SUA and ECoG signals exhibit different spatial resolutions and spectral contents (see **Chapter 1**). While the pertinence of kinematic decoding is for example well established for MUA/SUA signals, it is still a matter of discussion for ECoG signals (see **Chapter 1**). The recent interest for MUA/SUA switching modelling may thus be irrelevant for ECoG decoding, for example if the model of dependence between ECoG features and kinematic parameters is well approximated by a context-independent linear model. While the use of switching models has been reported in a few ECoG studies for the integration of NC support into kinematic decoders [Williams et al., 2013] [Bundy et al., 2016] or for sequential multi-limb decoding [Flamary and Rakotomamonjy, 2012], the validity and interest of the switching hypothesis has only been partially investigated in [Williams et al., 2013], where it was observed that neural patterns elicited during NC and IC states exhibited a different baseline, and where a post-processing strategy was found suboptimal for the discrimination between NC and IC states. It has not, however, been clearly established whether fitting kinematic decoders on IC samples exclusively may permit to improve ECoG decoding accuracy during IC periods. In [Bundy et al., 2016] for example, the performance of linear kinematic models fitted on IC samples was not compared to the one of models fitted on both NC and IC samples, and the performance of these two training strategies was not specifically evaluated on IC periods in [Flamary and Rakotomamonjy, 2012]. The high decoding

accuracy achieved by means of a post-processed linear model in [Wang et al., 2011] further illustrates that the relevance of switching models is unclear for ECoG signals. The results here reported may contribute to the exploration of the relevance of switching modelling of ECoG signals, in particular for the task of accurate decoding during IC periods and for NC support integration into mono-limb and/or multi-limb kinematic decoders.

Several studies have been completed to assess the respective performance of discriminant (static) and generative (dynamic) models for kinematic decoding of MUA/SUA and ECoG signals (see **Chapter 2**). Although generative switching models have been regularly applied for the decoding of MUA/SUA signals, their use had not been reported for ECoG signals for which switching modelling is generally performed within the framework of discriminative models. The relevance of discriminative or generative frameworks for switching modelling thus remains to be clarified, here specifically for the task of accurate asynchronous mono-limb and sequential multi-limb control with linear models.

7.3.1 Mono-limb decoding

The MSLM and its key features were first evaluated for a task of accurate asynchronous mono-limb trajectory reconstruction from ECoG signals.

7.3.1.1 Strategies for NC support

The preclinical data set permitted to complete a first study on the efficiency of the MSLM for the introduction of NC support into mono-limb kinematic decoders. The training of the proposed MSLM decoder was considered in the case of complete training data sets, i.e. parameter identification was supervised with respect to both trajectory and state sequences. A comparison of the MSLM with a Wiener Filter with Markovian post-processing and a SKF was performed to assess the relevance of two of the MSLM key features, namely the use of switching models rather than of a single post-processed model and the framework of discriminative rather than generative modelling. An additional comparison with a SLM and pWF, i.e. variants of the MSLM and MpWF with static state detection, was carried out to evaluate the interest of the MSLM's third key feature, namely dynamical gating.

MSLM-based state detection was found to significantly outperform the state detection achieved by the MpWF as measured by the AUC. The TPR, FPR and ACC of the MpWF were additionally degraded when compared to the ones associated with the MSLM. The MSLM's FPR was in particular lower than the MpWF'one by a median of 26.6% and 14.4% for the subdural and epidural data sets, respectively. These median improvements rose to 34.7% and 17.7% when synchronized probabilities were considered on the subdural and epidural data sets, respectively, which seems to indicate that the MSLM may exhibit a temporal uncertainty slightly superior to the WF's one. MSLM- and the MpWF-based state estimation mainly differ

because the MSLM decoder is fed with neural signals while the WF is fed with the trajectory estimates issued by a single kinematic model, i.e. a continuous model trained and applied on both NC and IC states. The lower performance of the MpWF with respect to the MSLM for state detection may indicate that optimal states estimates cannot be inferred from the kinematic estimates, and thus that the kinematic estimates yielded by a single NC/IC linear kinematic decoder may be corrupted because the decoder is trained on both NC and IC samples. This finding seems to suggest the existence of NC/IC-specific modifications of the considered linear model of dependence between ECoG signals and kinematic parameters, and are consistent with the NC/IC-specific changes of ECoG baseline activity which were reported in [Williams et al., 2013].

In the case of the subdural data set, the MSLM's IC expert and the MpWF performed equivalently when they were applied on IC states (**Table 7.8**): excluding NC samples when training the MSLM IC expert did not permit to improve the model between subdural neural signals and kinematic parameters during IC periods. Thus, although using a single continuous decoder resulted in a degraded state detection, its ability to decode IC states was not inferior to the one of the IC-specialized expert. These results suggest that possible modifications of the ECoG encoding may not be sufficiently important to impact the quality of the linear kinematic model used to decode neural features during IC periods on the considered subdural ECoG data set.

Different results were obtained on the epidural data set, where the IC experts yielded IC kinematic estimates on known IC states with slightly but significantly lower errors than the single linear model embedded in the MpWF model. Similar results were observed when MSLMs embedding an IC expert trained on either IC samples exclusively or on both NC and IC samples were compared, as the NMRSE and NMAE were significantly better on the epidural data set when IC samples only were used to train the IC expert. These observations are consistent with an earlier ECoG study [Williams et al., 2013] where it was advocated to utilize NC- and IC-specific intercepts when applying a linear model to decode both NC and IC samples, so that as to take into account differences in the baseline activity during NC and IC states. Interestingly, this study had also been completed with epidural ECoG signals [Williams et al., 2013]. A median improvement of 6% was additionally found for the PCC (see **Table 7.11**). While this improvement was not significant, the PCC between synchronized trajectories was found to be significant.

The main performance difference between SKF-based and MSLM-based switching modelling arose for state detection. The AUC and ACC associated with the MSLM were found to be significantly higher than the ones obtained with the SKF on both the epidural and subdural data sets. No significant difference was observed between SKF- and MSLM-based kinematic decoding accuracy during IC state. The MSLM's IC expert was fitted using PLS regression. The fact that the SKF was fed with features obtained after a PLS-based dimensionality reduction may have contributed to this similarity between the MSLM's and SKF's IC decoding performances. If both the SKF and MSLM are switching decoders which model the switching state

succession as a first-order Markovian process, their respective statistical frameworks differ. The SKF being a generative model, its state estimation procedure relies on the consistency of the NC- and IC-specific emission and transition models with the observed neural signals. Dimensionality reduction is additionally required for online application of the SKF, and a suboptimal feature selection may result in degraded kinematic estimates. Because it was designed as a discriminative model, the MSLM can directly exploit differences between the distributions of NC and IC high dimensional neural features. The generative framework here appeared less relevant than the discriminative one for the considered task of state detection from high-dimensional neural signals. The SKF exhibited a higher FPR than the MSLM. This high FPR may impair the SKF's applicability for the integration of NC support into kinematic decoders (see **Figure 7.1**), for example when high-dimensional neural features are considered. Stable NC positions were here considered, because it is desirable for the limbs of CLINATEC's exoskeleton to be located at an unique position during all blocks of NC states. As the SKF is theoretically able to handle variable NC positions, its relevance may be increased for applications with non-unique NC positions.

Finally, the interest of the MSLM's Markovian hypothesis was illustrated when it was compared it with a SLM, i.e. its analogue with static gating. While performance indicators based on the confusion matrix were similar for the MSLM and the SLM, the MSLM corresponded to significantly fewer false activations and deactivations, i.e. blocks of consecutive false positives or false negatives, on the epidural data set. Although it was not found significant, the same trend was observed on the subdural data set. False activations and deactivations were longer (median increase: 30.1% and 19.2% respectively, on the subdural data set, and 26.3% and 28.9% respectively, on the subdural data set) but fewer for the MSLM than for the SLM (median improvement of the false activations' and deactivations' frequency of 32.7% and 16.7%, respectively, on the subdural data set, and of 31.9% and 26.0%, respectively, on the epidural data set). As a block of adjacent misclassified samples is expected to be less disturbing to BCI users than a few isolated erroneous state estimates, these results seem to confirm the interest of the dynamic state detection performed by the MSLM.

7.3.1.2 Multiple IC experts for accurate decoding

Decoding accuracy during IC is required for users to exert efficient kinematic control over an effector [Marathe and Taylor, 2011]. Three subdural acquisition sessions were further analysed because they exhibited complex reaching movements for which one IC expert could be insufficient for accurate kinematic modelling. MSLM with $k = 2$ IC experts were identified using the unsupervised training approach proposed in the present doctoral work. The majority of decoders which use has been reported for ECoG kinematic decoding are static or dynamic linear models [Wang et al., 2013c] [Schalk et al., 2007] [Chao et al., 2010] [Shimoda et al., 2012] [Nakanishi et al., 2013] [Williams et al., 2013] [Hammer et al., 2013] [Wang et al., 2013c] [Hotson

et al., 2014] [Bundy et al., 2016]. They do not exploit potential modifications of the model of dependence between ECoG signals and kinematic parameters within movement phases. While the use of mixtures of linear IC models has been considered for MUA/SUA decoding [Kim et al., 2006c], the relevance of switching modelling of ECoG signals is not straightforward because of the comparatively lower spatial resolution and spectral content of ECoG signals which even challenge the efficiency of generic kinematic decoding. For two of the sessions, the cross-validated PCC, NMAE and, for one of the session the NRMSE, were significantly improved when computed over hidden IC samples. The NMAE was significantly improved on the third data set. The preliminary study led on data sets exhibiting complex, multi-phase movements suggests that MSLMs with $k > 1$ IC experts may be profitably used to decode complex movements, and therefore that phase-specific model modifications may be observed in ECoG signals.

7.3.2 Multi-limb decoding

The relevance of the MSLM for asynchronous multi-limb decoding was investigated on a publicly available clinical data set acquired while human subjects were executing finger movements. Similarly to the mono-limb asynchronous study, the MSLM was compared to both a SKF and a Wiener filter with Markovian post-processing. The three models intrinsically prevent large parallel limb activations, as one limb only is associated with each possible value of the latent hidden state. Parameter identification was supervised with respect to both trajectory and state sequences.

The MSLM significantly surpassed the MpWF for state detection in terms of ACC and FPR (median improvements of 4% and 33%, respectively). The observed state detection improvement suggests that a linear model of dependence between neural signals and kinematic parameters may depend on the activated limb. Consistently with a possible existence of state-specific modifications of a linear model of dependence between neural features and kinematic parameters, the IC experts resulted in kinematic estimates with significantly lower errors for both hidden and known state sequences, and, in the case of synchronized estimated trajectories, higher correlation coefficients than the MpWF. Interestingly, significantly higher PCC were obtained on IC samples when the MSLM embedded experts fitted on IC rather than on both NC and IC samples (unknown state sequence, see **Table 7.27**). This observation is coherent with the results obtained on the epidural preclinical data set. The MSLM superior state and kinematic decoding abilities were associated with one drawback, namely the fact that errors were larger when they did occur (False Positives or activation of a wrong finger) (see **Figure 7.8**). It is expected that this drawback is profitably compensated for by the state detection and kinematic ability of the MSLM, as a slight increase of false activation amplitude is likely to be of limited importance to subjects.

The SKF was outperformed by the MSLM with respect to both state detection

and trajectory reconstruction during IC periods. Although it had been proved efficient for IC kinematic decoding when it is correctly initialized at the beginning of each movement (see **section C.3.2** in **Appendix C**), the SKF here presented a poor decoding ability when the switching states were known but the SKF was not reinitialized at the beginning of each movement (see illustration in **Figure 7.6**). The SKF performance was further degraded in the case of unknown state sequences. It exhibited both a poor state detection along with a limited decoding accuracy during IC periods. These results seem to indicate that the generative framework is less relevant than the discriminative one for switching modelling with more than 2 experts, in particular that the MSLM's direct exploitation of the distribution of the neural features is more efficient for state detection than the SKF's indirect strategy. The utilization of a SKF-like model had not yet been proposed for asynchronous multi-limb decoding, i.e. for $K > 2$. The results here obtained for the task of multi-limb decoding suggest that the use of the SKF may not be easily extended for multi-limb NC support, because it yields poor kinematic estimates which low accuracy may hinder its applicability for closed-loop ECoG-driven motor BCI systems. Further analyses are required to clarify the limits of the SKF. Mixtures of KF were for example considered in e.g. [Yu et al., 2007], where they were profitably used to extract goal-directed trajectories from MUA/SUA signals in monkeys. The SKF efficiency may have been here limited by ECoG spatial resolution or by the loss of informative neural features during dimensionality reduction.

Finally, similarly to the mono-limb asynchronous study, the performance of the MSLM for state detection was assessed by comparing it with the SLM, its analogue with static state detection. The MSLM corresponded to higher median ACC and CTPR than the SLM. It was additionally associated with less frequent false deactivations than the SLM.

7.3.3 Absolute decoding performance

The comparative relevance of the three key features of the MSLM has thus been assessed on both preclinical and clinical ECoG data sets, and the reported results suggest the interest of using the proposed decoder for ECoG asynchronous kinematic decoding. The absolute decoding performance of the MSLM, however, is also crucial. In particular, the ability of users to control reaching movements depends on the accuracy of the kinematic parameters extracted from their neural signals [Marathe and Taylor, 2011]. The IC decoding performances obtained on the preclinical data set (median PCC around 0.7 and 0.3 for the reconstruction of 3D wrist movements during known IC states for subdural and epidural ECoG signals, respectively) suggest that kinematic decoding is less challenging for subdural ECoG signals than for epidural ones. Incidentally, these results are consistent with the PCC reported in [Bundy et al., 2016], where 3D trajectory reconstructions from subdural ECoG signals reached average PCCs of 0.49, 0.28 and 0.38 for the x-, y- and z-axis, respectively. Similarly, state detection quality was higher for subdural acquisition sessions than for

epidural ones (average ACC of 0.89 and 0.96 for the subdurally implanted monkeys A and K, respectively, and of 0.83 and 0.77 for the epidurally implanted monkeys B and C). The FPRs, which are particularly important for clinical applications, range from 0.3 and 0.11 (subdural data set, Monkey K and A, respectively) to 0.14 and 0.17 (epidural data set, Monkey C and B, respectively). While these state detection performances may appear deterrent for clinical applications, they are consistent with results reported in [Bundy et al., 2016] (ACC ranging from 0.68 to 0.90 depending on subject, with an average of 0.80 for 5 subdurally-implanted subjects) and [Wang et al., 2013b] (average ACC of 0.91 for 3 subdurally-implanted subjects). Studies have suggested that state detection accuracy may be facilitated during closed-loop control [Williams et al., 2013] (ACC of 0.98-0.99). The observed difference between subdural and epidural signals is coherent with the findings reported in [Bundy et al., 2014], where the quality of signals acquired by subdural micro-ECoG array (inter-electrode distance of 1mm, against 3.5mm for the preclinical data set analysed in the present doctoral thesis) is superior to the epidural equivalents. The accuracy of the NC/IC state detection completed on the clinical data set was stable from one subject to another, and comparable to the results obtained on subural preclinical data (average ACC of 0.85 for 9 subjects, standard deviation of 0.02). The discrimination between fingers was found more challenging (average ratio of correctly classified fingers of 0.65), but results were consistent with results reported in [Saa et al., 2016], where a CRF permitted to achieve an ACC of 0.65 for 6-class classification. Closed-loop finger detection was performed with an accuracy of 0.92 for NC/IC detection and of 0.75 for inter-finger discrimination [Hotson et al., 2016], again suggesting that performance improvements are observed after training. An average PCC of 0.47 was obtained for MSLM-based multi-finger trajectory reconstruction on known IC samples. Thus, the IC decoding accuracy was lower for the clinical subdural data set when compared to the preclinical subdural one. The larger inter-electrode distance of the clinical ECoG arrays may have contributed to this difference [Bundy et al., 2014]. The PCC is slightly lower than the PCC reported in the acausal decoding study completed in [Flamary and Rakotomamonjy, 2012]. Further investigations are required to assess whether the exploitation of acausal information would permit to the MSLM to achieve the same decoding performance.

7.4 Conclusion

The decoding performance of the Markov Switching Linear Model has been presented and compared to the performance of alternative decoders for two decoding tasks, namely accurate asynchronous mono-limb and sequential multi-limb decoding. The reported results have permitted to validate the relevance of the three key features of the MSLM, namely switching modelling, discriminative modelling and dynamic state detection.

A comparison of the MSLM with a Wiener Filter with Markovian post-processing (MpWF) suggested the interest of switching modelling of ECoG signals, in particular

to improve the accuracy of kinematic estimates during IC states. This interest did not clearly arise from earlier studies led on ECoG signals. In particular, it is not well established whether linear models of dependence between ECoG neural features and kinematic parameters are state- or context-specific, and whether taking modifications of this model into account may permit to improve decoding accuracy. Improvements of the decoding accuracy were here observed for mono-limb kinematic decoding from epidural preclinical ECoG signals and from clinical signals measured during multi-limb movements. A reduction of spurious activations during NC states and of erroneous limb activations during IC periods was also observed. As the unsupervised training of MSLM with $k = 2$ IC experts permitted to improve IC decoding accuracy in three acquisition sessions exhibiting complex, multi-phase movements, switching modelling may also be of interest for multi-phase ECoG kinematic decoding.

An additional comparison was completed with a Switching Kalman Filter, i.e. a switching generative model, to assess the relevance of discriminative switching modelling with respect to generative switching modelling. To the best of our knowledge, such comparison had not been drawn in earlier MUA/SUA or ECoG studies. While the SKF permitted to achieve satisfying decoding accuracy over IC periods on the preclinical data set, it was outperformed by the MSLM in terms of state detection accuracy. The SKF additionally exhibited poor decoding performances when applied for multi-limb sequential ECoG decoding. While generative kinematic decoders are widely popular for kinematic decoding, these results seem to indicate that the generative framework may be suboptimal for switching ECoG modelling.

The Markovian hypothesis used to perform dynamic state detection was here found to be advantageous to limit short spurious system activations or deactivations.

The relevance of its three key features and of the associated training strategies permitted the MSLM to address the specific challenges of this thesis, i.e. namely accurate asynchronous mono-limb and sequential multi-limb decoding. The MSLM corresponded to improved state detection and/or kinematic parameter estimation on the considered ECoG data sets. Finally, the resulting accuracy is expected to be compatible with closed-loop control.

Chapter 8

Discussion and perspectives

Contents

8.1 Discussion

In spite of proofs of concept in laboratory environments [Collinger et al., 2013] [Wodlinger et al., 2015], clinical applications of motor BCIs remain rare [Mak and Wolpaw, 2009]. Technical obstacles to the clinical deployment of motor BCI systems are addressed in the present doctoral thesis.

8.1.1 Challenges for motor BCI clinical applications

Three technical challenges have been particularly considered, namely asynchronous mono-limb decoding, asynchronous sequential multi-limb decoding and decoding accuracy during active states.

A major issue for BCI clinical applications is the ability to provide users with accurate asynchronous control over the effector [Graimann et al., 2009]. Unlike synchronous BCIs which are periodically controllable by users, asynchronous BCI decoders are continuously available. The limitation of spurious effector activations during No-Control (NC) periods is desirable for asynchronous BCI systems. In contrast with communication BCIs which generally permit users to control virtual effectors (e.g., a cursor displayed on a computer screen), motor BCI systems embed physical effectors such as robotic prostheses or orthoses. False activations are likely to be particularly disturbing and stressful to users when they physically interact with the BCI effector. NC support, i.e. the ability of the BCI decoder to output

neutral values during NC states, is thus especially important for orthosis-based motor BCI systems.

Although it has been scarcely explored by the BCI community, multi-limb decoding is expected to improve the independence of severely motor impaired patients. Upper- and lower-limb control would for example be particularly beneficial to patients with tetraplegia. The extension of mono-limb asynchronous decoding to multi-limb asynchronous control is one of the challenges arising for multi-limb BCI systems. A sequential upper-/lower-limb activation strategy was here considered. It was chosen for CLINATEC's BCI system with the aim of improving the system's robustness and of facilitating neural control. The corresponding decoding objective lies in avoiding both spurious system activations and parallel, residual movements of the momentarily non-controlled limbs.

Finally, the ability of users to execute brain-controlled movements is impaired when the estimates of the user's intentions are not sufficiently correlated with his true intentions [Marathe and Taylor, 2011]. Decoding accuracy during Intentional Control (IC) periods is thus crucial for motor BCIs.

A decoder, namely a Markov Switching Linear Model (MSLM), has been developed in the present doctoral work for the task of accurate, asynchronous sequential multi-limb kinematic ECoG decoding.

The interest of switching models for kinematic reconstruction from ECoG signals has been investigated. Their relevance depends on the presence of state-related modifications in linear models of dependence between neural features and kinematic parameters. While the presence of such changes has regularly been illustrated in MUA/SUA studies, it has not yet been clearly established for ECoG signals.

The respective relevance of generative and discriminative approaches for switching modelling of ECoG signals has additionally been explored. Although generative switching models have been regularly applied for the decoding of MUA/SUA signals, their use had not been reported for ECoG signals for which switching modelling is generally performed within the framework of discriminative models. Their respective relevance of both switching frameworks remained to be investigated for ECoG signals.

Both supervised and unsupervised training procedures have been presented for the proposed discriminative Markov Switching Linear Model. An evaluation procedure, which includes block-wise criteria and permits to separate spatial from temporal errors, has been proposed to assess the comparative decoding performances of the MSLM and of alternative decoders which use has been reported within the BCI community.

Finally, the ability of the proposed model to address 3 challenges associated with clinical motor BCIs, namely efficient asynchronous mono-limb control, sequential multi-limb control and accurate decoding of kinematic parameters during active states, has been evaluated.

8.1.2 Proposed methodology

The Markov Switching Linear Model has been designed as a switching discriminative kinematic decoder with dynamic state detection.

8.1.2.1 Markov Switching Linear Model

The MSLM strategy to support NC periods in the case of mono-limb and multi-limb sequential asynchronous control consists of switching between NC and IC models to take into account state-specific modifications of a linear model of dependence between neural features and kinematic parameters (see Chapter 3). Several IC experts are used when complex movements are considered. Switches between IC models are therefore used to address the issue of multi-limb effector control (one IC model per limb) as well as to improve decoding accuracy during IC periods (several IC experts to model complex movements executed with one limb). The relevance of each model, i.e. of each state, is estimated from the neural data and is used to combine the available continuous models.

Three key features have been used to design the MSLM (see **Figure 5.3**). First, the MSLM has been developed as a switching model (or mixture of linear models), in contrast with post-processed decoders which rely on a single model to describe the dependence between neural signals and kinematic parameters. Switching models are liable to improve both the quality of NC support for asynchronous decoding and the accuracy of kinematic estimates during IC states if the model between neural signals and kinematic parameters is context-dependent.

Second, it has been conceived as a discriminative switching model, as opposed to generative switching models like the Switching Kalman Filter which utilization has been proposed by the BCI community. More precisely, the MSLM has been designed as an extension of Mixtures of Experts (ME) models. A static linear model between neural features and kinematic parameters is conditioned on the current hidden switching neural state. A probabilistic rule is used to weight the available regression models. This soft weighting approach results in smoother transitions than the widely used Winner-takes-all hard combination strategy, and is here expected to facilitate asynchronous effector control.

Finally, in contrast with the previously reported BCI switching regression models, the MSLM performs dynamic state detection to limit spurious effector activations in asynchronous settings. More specifically, the sequence of NC and (possibly multiple) IC states is assumed to be generated by a first-order Markov chain. A Hidden Markov Model-based (HMM) discrete decoder is used for state estimation. The corresponding MSLM thus extends Mixtures of Experts models by embedding sequential rather than static state detection.

Both supervised and unsupervised training procedures have been presented in the present dissertation. Supervised training is possible when the values taken by the switching state in the training data set are known. As NC/IC labels can

be inferred from the tracked limb trajectory, supervised training was utilized to train 2-state and 6-state MSLMs to introduce NC support into mono-limb and sequential multi-limb kinematic decoders. Unsupervised training permits to exploit internal states without precise a priori knowledge on their value in the training data set. It was here used to fit several IC experts to model complex unimanual reaching movements. The proposed unsupervised training procedure is based on the Expectation-Maximization (EM) algorithm, and extends both HMM- and ME- well known EM training algorithms.

8.1.2.2 Implementation and evaluation approach

The proposed MSLM is flexible inasmuch as it can embed different HMM-based gate and experts. The relevance of experts and gates may in particular be impacted by the measurement noise associated with the experts, by the dimensionality of the neural features and by their underlying distribution. The MSLM gate and expert structures and/or the corresponding training approaches were therefore optimized for the considered data sets in a preliminary study (see **Chapter 6** and **Appendix B**).

Benchmark decoders were chosen so as to reflect the strategies which use had been reported in earlier studies for asynchronous (possibly multi-limb) kinematic decoding (see **Chapter 5**). A Markovian post-processed Wiener Filter (MpWF) was chosen to represent a first alternative to the MSLM, namely a single model with post-processing-based (possibly multiple) NC support. A Switching Kalman Filter (SKF) was additionally implemented to assess the comparative performance of discriminative and generative switching models.

Performance indicators were selected to assess the ability of the MSLM to support asynchronous mono-limb and multi-limb and to accurately reconstruct kinematic parameters during IC periods (see **Chapter 5**). A first indicator subset focused on the capacity of each decoder to handle asynchronous mono-limb or multi-limb decoding, i.e. it gathered criteria measuring the quality of each decoder's NC support and/or of the discrimination between multiple active states. The performance of brain switches is generally evaluated using indicators extracted from the confusion matrix, for example the number of observations erroneously decoded as IC states. It was here proposed to additionally consider block-wise criteria, namely to measure the frequency and duration of false activation blocks. Because they take into account the dynamic of the errors made by the state decoder, block-wise criteria are expected to reflect the user's perception of NC support quality during asynchronous closed-loop control sessions. A second indicator subset was dedicated to the evaluation of the accuracy of the kinematic parameters reconstructed by each decoder during IC states. It was composed of the criteria typically associated with continuously-valued variables. These generic criteria are sensitive to slight desynchronizations between observed and estimated trajectories, and they therefore reflect both spatial and temporal decoding accuracies. For this reason, it was proposed to complement them by computing their synchronized analogues, i.e. to compute generic performance

indicators after having synchronized the estimated trajectory with the observed (tracked) trajectory. These synchronized criteria thus isolate spatial accuracy from temporal accuracy.

8.1.3 Results

The ability of the MSLM decoder to address the issue of accurate asynchronous (possibly multi-limb) decoding was assessed for two decoding tasks, namely asynchronous uni-manual trajectory reconstruction and multi-limb (multi-finger) trajectory reconstruction (see **Chapters 4** and **6**). Results suggest that the proposed MSLM decoder may be profitably used to introduce NC support into ECoG asynchronous kinematic decoders, possibly with sequential multi-limb control. It can additionally be used to improve the accuracy of complex reaching movement estimates by combining several IC experts. In particular, the results permit the validate the relevance of the MSLM three key features (see **Chapter 7**).

First, the switching framework was found to be more efficient than the post-processing framework for the two considered state detection tasks, namely NC/IC and NC/multiple IC state detection. Post-processing-based NC support has the advantage of a reduced training duration. Its efficiency is nevertheless degraded if linear models between neural signals and movement kinematics are not consistent across states, e.g. across NC and IC states. While the presence of such modifications has been regularly reported in MUA/SUA studies, it has not been clearly established for ECoG signals. Post-processing-based mono- and multi-limb asynchronous decoding was here less accurate than MSLM-based decoding. In the case of mono-limb control, a similar superiority of the switching hypothesis over the post-processing one had been reported in [Williams et al., 2013], where the estimated velocity of a cursor was found to provide a poor approximation of NC and IC states. Although a post-processed decoder designed for parallel multi-finger movements has been proposed in [Wang et al., 2011], to the best of our knowledge, post-processing approaches had not yet been considered for sequential multi-limb decoding. Our results suggest they may be less efficient that switching models for asynchronous multi-limb control. It should finally be noted that although the switching hypothesis appeared relevant for NC support, it did not systematically result in an improvement of kinematic decoding accuracy during IC periods. Improvements of the IC decoding accuracy were, however, observed for the epidural preclinical and clinical data sets. The utilization of several IC experts did permit, however, to improve the decoding of the complex movements exhibited in some sessions of the subdural preclinical data set.

Discriminative switching models were additionally found to be more efficient than generative switching models for asynchronous mono-limb and sequential multi-limb accurate decoding. Generative dynamic models, e.g. Kalman Filters, are frequently used for kinematic decoding. They have the benefit of embedding a model of the movement dynamic, which can be advantageously used to constrain

the estimated trajectory smoothness [Koyama et al., 2010b]. The application of generative dynamic models is nevertheless more complex than the application of generic discriminative models. In particular, Kalman filtering involves recurrent matrix inversions which are computationally expensive when neural features are high-dimensional. A supplementary dimensionality reduction step may thus be required for real-time application of the KF. Numerical issues due to roundoff errors may additionally arise when a generic implementation of the KF is used [Tusell, 2011]. The ability of generative models to handle (possibly) multi-limb asynchronous decoding was investigated in the present doctoral thesis. Switching generative models had been profitably used for NC support in a simulated study [Srinivasan et al., 2007]. Additionally, their application has been proposed to improve MUA/SUA decoding accuracy during reaching movements by combining trajectory models [Yu et al., 2007], and to enhance kinematic reconstruction from SUA/MUA signals during IC states by using several emission models [Wu et al., 2003b] [Wu et al., 2004]. In spite of these studies, the relevance of generative switching models on ECoG data had not yet been established, and their performance had not yet been compared with the one of their discriminative analogues. Despite the advantage conferred by explicitly modelling the kinematic dynamic during NC states, generative switching models here appeared as less efficient that discriminative switching models for asynchronous ECoG decoding. This suggests that Kalman filtering, which is increasingly popular for kinematic decoding, may suffer from difficulties to handle asynchronous decoding. Such difficulties were probably encountered in the MUA/SUA-based clinical trial completed by Hochberg and colleagues [Hochberg et al., 2012], where it was reported that the KF used for kinematic decoding was re-initialized at the beginning of each trial. Kalman-based IC experts additionally yielded IC kinematic estimates which were either equivalent or less accurate than the ones output by the MSLM linear experts.

Finally, dynamic gating permitted to limit false detections and/or modify their dynamic, namely to replace short and numerous false activations by fewer (but longer) spurious activations which are expected to be less disturbing to users. The relevance of dynamic detection is further suggested by the fact that recent clinical motor BCI studies relied on a first-order Markovian hypothesis for efficient (possibly NC) state detection in closed-loop control sessions [Fifer et al., 2014] [Hotson et al., 2016] [Kao et al., 2017].

The absolute IC decoding performance which was reached by the MSLM on the considered data sets is consistent with the decoding accuracy reported by several teams on private data sets (e.g., [Bundy et al., 2016]). It has been shown in [Marathe and Taylor, 2011] that position decoding with a PCC of 0.5 can be satisfyingly used to control the velocity of an upper-limb effector during reaching movements. For this reason, it is expected that the IC decoding ability of the MSLM makes it compatible with closed-loop kinematic control of upper-limb effectors.

The reported results thus suggest that the proposed decoder may be profitably used for ECoG-driven accurate closed-loop asynchronous mono-limb or multi-limb

decoding.

8.1.4 Limits of the present work

Because the present study was performed to prepare for CLINATEC's BCI clinical deployment, it was not completed with optimal chronic clinical ECoG data and therefore presents several limitations.

8.1.4.1 ECoG array and subjects

The relevance of the MSLM for asynchronous mono-limb decoding was assessed using preclinical ECoG data. Application of the MSLM on human ECoG data is needed to confirm the trends observed on an inherently limited monkey model [Ball et al., 2009b]. Although the MSLM ability to reconstruct asynchronous multi-limb movements has been measured on clinical ECoG data, additional tests are also required to address the shortcomings of this data set. While epileptic subjects are considered as safe models of non-epileptic patients if some guidelines are respected [Lachaux et al., 2012], data acquisition was performed using ECoG arrays which characteristics and implantation area had been chosen with the goal of localizing the patients' epileptic focus. In particular, the exposed electrode diameter was 2mm, and the 1cm-long inter-electrode distance resulted in a sparse sampling of the subjects' motor cortex. Subdural acquisition by macro-ECoG arrays has been shown to result in signals equivalent to those measured by epidural macro-ECoG arrays [Bundy et al., 2014]. Similarly, the results presented in [Wang et al., 2016] suggest that dense ECoG arrays are desirable for BCI applications.

The fact that the considered data sets were acquired from able-bodied subjects is another limit of the present study. Severely motor-impaired patients cannot execute overt (real) movements. Decoders are usually initialized by presenting subjects with instructions (e.g., movements executed by an experimenter), and by asking them to attempt to perform the proposed movements (e.g., [Hochberg et al., 2012] [Wodlinger et al., 2015]). A first potential difficulty lies in a possible difference in the strength of the patterns elicited by attempted movements and the ones generated by the execution of real movements. It has been suggested in several studies that attempted movements elicit ECoG patterns which are exploitable for decoder calibration [Collinger et al., 2014]. An accuracy drop may nevertheless be observed for motor-impaired patients. Additionally, the intentions of the subject are not accessible and are approximated by the instructions given to them. An unknown lag exists between these instructions and the corresponding movement attempts. If reaction times generally exhibit a small variability, episodic losses of concentration of the user may occur and be added to the reaction time.

8.1.4.2 Self-paced multi-limb complex 3D movements

Although a task of multi-finger control permitted to measure the MSLM ability to manage sequential limb activation, the ability of the MSLM to handle upper- and

lower-limb sequential activation remains to be explored. In particular, lower-limb movements may be harder to detect because they are mainly controlled in a deeper cortex area than fingers or arm (see the presentation of the somatotopic organization of the motor cortex in **section 1.2.2.1**). On the other hand, the approximate lateralization of motor control may make it easier to exploit spatial patterns to discriminate between left and right arm movements than to classify the activation of fingers of the same hand.

1D limb movements were considered in the case of the clinical data set. As for the preclinical data set, the wrist trajectories used to assess decoders' performance were associated with a limited space exploration. It is thus desirable to explore the ability of the MSLM to decode complex 3D (possibly multi-limb) trajectories. As a very limited number of complex reaching movements were present in the considered preclinical data set, additional studies are particularly required to further investigate the interest of using multiple IC experts to model complex reaching movements.

The NC states present in the preclinical and clinical acquisition sessions exhibited similar temporal characteristics, i.e. they were a few seconds long and the standard deviation of their duration was rather low (up to 3 seconds, see **Chapter 4**). The ability of the MSLM to detect NC states with more variable durations needs to be assessed.

8.2 Perspectives

Further investigations are necessary to strengthen and extend the hereby reported results, and to address the limits of the present study.

8.2.1 Technical optimizations

The relevance of several potential technical improvements remains to be explored.

8.2.1.1 Expert and gate structure

The study which was here completed to choose the gate and expert structure and training approach (see **Appendix B**) is not exhaustive. Several options may be considered to further optimize MSLM gate and experts.

Gate The MSLM gate relies on a HMM for dynamic state estimation. While the MSLM gate was able to detect state of moderately variable durations (see data set description in **Chapter 4**), the fact that the duration of each state is implicitly modelled by the HMM transition matrix (see **section 3.2.1**) may hinder the MSLM efficiency for some applications, for example in the case of highly variable state durations. Explicit modelling of the movement duration, for example using explicit-duration HMMs (also referred to as Hidden semi-Markov Models) [Yu, 2010], may be explored in the future.

Another structural modification could benefit state detection in the case of multiple IC states, namely the utilization of a hierarchical state detection strategy. This approach consists of training a first classifier to detect NC and IC states and a second one to discriminate between multiple IC states. Its use has been reported in several EEG [Murguialday et al., 2007] and ECoG studies [Hotson et al., 2016] [Wang et al., 2016].

State-specific experts (MSLM's continuous models) Although linear experts have been used in the present work, the utilization of nonlinear experts may be considered. The use of nonlinear decoders, for example Generalized Linear Models [Eliseyev and Aksenova, 2014], Multilayer Perceptrons [Kim et al., 2006b] or Support Vector Machine Regression [Kim et al., 2006b] [Mehring et al., 2003], has been proposed for the estimation of kinematic parameters from ECoG or MUA/SUA signals. The performance of the MSLM may additionally be compared to the previously mentioned nonlinear models for the task of asynchronous ECoG decoding.

8.2.1.2 Training procedures

The interest of several modifications of the MSLM training procedures may be investigated.

Supervised training Improvements may be brought to the gate training procedure. PLS-based dimensionality was here cascaded with a logistic regression model for state detection on the preclinical data. Generic and penalized PLS-based training approaches have been proposed for logistic regression [Marx, 1996] [Fort and Lambert-Lacroix, 2005], and may permit to perform optimal dimensionality reduction during the training of the logistic regression model. While their computational cost was here deemed deterrent, the interest of these identification algorithms needs to be thoroughly investigated.

The quality of the MSLM linear models may be enhanced by using the efficient N-way training approaches developed in CLINATEC. In particular, (penalized) N-way PLS algorithms have been proposed for sparse [Eliseyev et al., 2012] and/or smooth [Eliseyev and Aksenova, 2016] expert training, and may be profitably used to train the MSLM experts.

Unsupervised and semi-supervised training The approach here used for the selection of the number of IC experts may be improved by recoursing to variational Bayesian model identification methods [Bishop, 2006]. Variational Bayes training generalizes Expectation-Maximization training by including hyperparameters in the set of hidden variables. Though more complex, they here have the interest that the number of active models can thus be directly estimated during training.

State labelling can be a tedious and error-prone process. Difficulties to assign NC/IC labels to neural data were for example reported in [Suway et al., 2013]. Such

difficulties may additionally be encountered in the case of open-loop, observation-based model calibration for motor-impaired patients. Although presented with instructions, subjects are liable to initiate attempted movements with a certain time lag, and/or to loose concentration during a few instructed movements. Labelling these movements as IC may degrade the MSLM decoder. Semi-supervised training strategies may be used to address this issue, for example by using probabilistic NC/IC labels and by adjusting the probability during training. Semi-supervised training has for example been considered for HMMs [Bordes and Vandekerkhove, 2005] [Scheffer et al., 2001] [Ozkan et al., 2012]. Its extension to MSLM training may be explored.

Recursive supervised or unsupervised training Neural patterns have been shown to change during closed-loop training [Carmena et al., 2003] [Rouse and Moran, 2009]. Turn-taking training strategies, i.e. decoder identification steps inserted between user training sessions, have therefore been utilized in several motor BCI clinical studies [Hochberg et al., 2012] [Wang et al., 2013c]. A promising extension of turn-taking strategies consists of using adaptive training algorithms for online adjustments of the decoder parameters, i.e. parallel rather than sequential user and decoder training [Dangi et al., 2014]. Recursive training may thus be advantageously considered for online adaptation of the MSLM to the new neural patterns. Several algorithms have been proposed for HMM recursive training [Khreich et al., 2012]. Efficient recursive training algorithms have additionally been proposed in CLINATEC for N-way PLS model identification [Eliseyev and Aksenova, 2013]. In the case of supervised training, they may permit to perform recursive training of the MSLM.

8.2.2 Closed-loop evaluation during CLINATEC's clinical trial

CLINATEC "BCI and Tetraplegia" clinical research protocol, which Principal Investigator is Prof. A.-L. Benabid [ClinicalTrials.gov, 2016], will permit to further explore the relevance of the MSLM and of the above-mentioned potential improvements. Its objective is the chronic deployment of an accurate, multi-limb, multi-DoF (Degrees of Freedom) motor BCI in humans. The wireless 64-channel ECoG implant WIMAGINE® has been specifically designed for stable and long-term signal acquisition [Mestais et al., 2015]. CLINATEC's BCI platform [Eliseyev et al., 2014] also includes the 4-limb exoskeleton EMY [Morinière et al., 2015] and the software environments required for real time processing of the neural signals. WIMAGINE's inter-electrode distance of 4mm [Mestais et al., 2015] will permit a dense sampling of human users' ECoG motor activity. CLINATEC's clinical trial will permit to investigate the ability of the MSLM to decode complex, self-paced multi-limb movements during closed-loop experiments.

As the forward algorithm used by the MSLM for model gating is computationally efficient [Rabiner, 1989], the MSLM is expected to be compatible with real-time BCI decoding.

A soft combination strategy has been chosen for the MSLM experts because it was expected that hard combination results in disturbingly abrupt transitions and increases the delay the user perceives between his intention and the system's reaction. Closed-loop experiments only, however, will permit to validate or invalidate this choice. In particular, the possible impact of the combination strategy on subject training is yet to be investigated.

As mentioned ealier, changes of context between open-loop and closed-loop BCI settings result in differences between open-loop and closed-loop neural patterns [Leuthardt et al., 2006a] [Jackson and Fetz, 2011] [Jarosiewicz et al., 2013]. Decoder adaptation during closed-loop experiments has been regularly used in preclinical and clinical motor BCIs to handle these pattern modifications (e.g., [Wang et al., 2013c] [Wodlinger et al., 2015]). CLINATEC's clinical trial will permit to investigate the need for regular adaptation of the MSLM, and to select the best algorithmic strategy to achieve the closed-loop training of the MSLM, namely punctual recalibration or online adaptation via recursive training algorithms.

CLINATEC's clinical trial will additionally permit to further assess the interest of the MSLM for the decoding of complex movements, which was here suggested by the results obtained on a limited data set. Complex movements are crucial for the execution of everyday tasks, for example object manipulation. Object manipulation may in particular involve (coordinated) bimanual movements [Swinnen and Wenderoth, 2004]. It may thus be desirable for the MSLM to have the ability to handle parallel multi-limb movements, e.g. coordinated bimanual movements. The study completed by Ifft and colleagues in [Ifft et al., 2013] suggests that the MSLM switching hypothesis may be relevant to decode parallel multi-limbs movements, as they reported that modifications of the tuning properties of individual neurons were induced when NHPs performed unimanual or bimanual reaching movements (see **Chapter 3**). The relevance of the MSLM for object manipulation also depends on its ability to handle multi-phase movements, in particular phases of stabilization over objects. Stabilization difficulties have been reported in several MUA/SUA motor BCIs, and have led several teams to address this issue [Kang et al., 2012] [Golub et al., 2014] [Gowda et al., 2014] [Kao et al., 2017]. Stabilization was for example improved by reducing the effector speed when error neural signals and/or hectic trajectories were detected [Gürel and Mehring, 2012] [Golub et al., 2014]. The ability of the MSLM to handle hold periods will be explored during CLINATEC's BCI clinical application. The detection of a "hold-on-target" switching state [Kang et al., 2012] may permit to help users to stabilize on targets.

Appendix A

Supplementary materials for MSLM training

Contents

Supplementary materials on the supervised and unsupervised training of the MSLM are gathered in the present appendix.

A.1 Supervised training: decoupling of parameter estimation

Let us consider a MSLM composed of K linear experts gated by a HMM-based sequential decoder. Each linear expert i is parametrized by $\mathbf{B}_i$, where, depending on context, $\mathbf{B}_i$ refers either to the full set of parameters associated with expert k (i.e., parameters of the linear model and of the noise distribution) or only to the parameters of the linear model. Let $\mathbf{A} = (a_{ij})$ be the transition matrix $a_{ij} = P(z^{t+1} = j|z^t = i), i, j = 1, ..., K$ and $\boldsymbol{\pi}$ be the initial state distribution $\pi_i = P(z^1 = z_i),\ i = 1, ..., K$ associated with the first-order Markovian hidden state variable z^t. Finally, let $\{b_1, \ldots, b_K\}$ gather the parameters necessary to characterize the distributions $P(\mathbf{x}^t|z^t = j), j = 1, \ldots, K$, e.g. mean and variance in the case of Gaussian emissions.

Knowledge about the values taken by the switching variable in the training data set permits to decouple the identification of the gate and experts parameters.
Let us consider the MSLM data complete log-likelihood

$$L_c(\boldsymbol{\Theta}, \mathbf{X}, \mathbf{Y}, \mathbf{z}) = \ln P(\mathbf{X}, \mathbf{Y}, \mathbf{z}|\boldsymbol{\Theta}), \tag{A.1}$$

where

$$P(\mathbf{X}, \mathbf{Y}, \mathbf{z}|\boldsymbol{\Theta}) = P(\mathbf{X}, \mathbf{Y}|\mathbf{z}, \boldsymbol{\Theta})\ P(\mathbf{z}|\boldsymbol{\Theta}) = P(\mathbf{x}^{1:T}, \mathbf{y}^{1:T}|z^{1:T}, \boldsymbol{\Theta})\ P(z^{1:T}|\boldsymbol{\Theta}). \tag{A.2}$$

Let us consider the first term of (A.2), i.e. $P(\mathbf{x}^{1:T}, \mathbf{y}^{1:T}|z^{1:T}, \boldsymbol{\Theta})$. Following the MSLM hypotheses presented in **section 3.3**,

$$P(\mathbf{x}^{1:T}, \mathbf{y}^{1:T} | z^{1:T}, \mathbf{\Theta}) = \prod_{t=1}^{T} P\left(\mathbf{x}^t, \mathbf{y}^t | z^{1:T}, \mathbf{\Theta}\right).$$

As the emission distribution of $\mathbf{x}^t$ is only conditioned on z^t and as $\mathbf{y}^t$ only depends on $\mathbf{x}^t$ and z^t,

$$P(\mathbf{x}^{1:T}, \mathbf{y}^{1:T} | z^{1:T}, \mathbf{\Theta}) = \prod_{t=1}^{T} P(\mathbf{x}^t, \mathbf{y}^t | z^t, \mathbf{\Theta}),$$

with $P(\mathbf{x}^t, \mathbf{y}^t | z^t, \mathbf{\Theta}) = P(\mathbf{y}^t | \mathbf{x}^t, z^t, \mathbf{\Theta}) P(\mathbf{x}^t | z^t, \mathbf{\Theta})$.
$P(\mathbf{y}^t | \mathbf{x}^t, z^t, \mathbf{\Theta}) = P(\mathbf{y}^t | \mathbf{x}^t, \mathbf{B}_{z^t})$ is the likelihood of $\mathbf{y}^t$ for the expert relevant at time t. $P(\mathbf{x}^t | z^t, \mathbf{\Theta})$ is the HMM emission model corresponding to the state z^t, and parametrized by b_{z^t}. Thus, $P(\mathbf{x}^t | z^t, \mathbf{\Theta}) = P(\mathbf{x}^t | b_{z^t})$ and

$$P(\mathbf{x}^{1:T}, \mathbf{y}^{1:T} | z^{1:T}, \mathbf{\Theta}) = \prod_{t=1}^{T} P(\mathbf{y}^t | \mathbf{x}^t, \mathbf{B}_{z^t}) P(\mathbf{x}^t | b_{z^t}). \tag{A.3}$$

The second term of A.2 can be rewritten as

$$P(z^{1:T} | \mathbf{\Theta}) = P(z^T | z^{1:T-1}, \mathbf{\Theta})\ P(z^{1:T-1} | \mathbf{\Theta})$$

Taking into account the first-order Markovian hypothesis associated with the latent state variable z^t and the chosen parametrization,

$$P(z^{1:T} | \mathbf{\Theta}) = P(z^T | z^{T-1}, \mathbf{A})\ P(z^{1:T-1} | \mathbf{A}, \pi).$$

By recurrence,

$$P(z^{1:T} | \mathbf{\Theta}) = P(z^1 | \pi) \prod_{t=1}^{T-1} P(z^{t+1} | z^t, \mathbf{A}) = \pi_{z^1} \prod_{t=1}^{T-1} a_{z^t,\ z^{t+1}}. \tag{A.4}$$

Finally, from A.2, A.3 and A.4,

$$L_c(\Theta, \mathbf{X}, \mathbf{Y}, \mathbf{z}) = \ln \pi_{z^1} + \sum_{t=1}^{T-1} \ln a_{z^t, z^{t+1}} + \sum_{t=1}^{T} \ln P(\mathbf{y}^t | \mathbf{x}^t, \mathbf{B}_{z^t}) + \sum_{t=1}^{T} \ln P(\mathbf{x}^t | b_{z^t}). \tag{A.5}$$

Let us introduce the following notation: $\delta_{z^t,k} = 1$ if $z^t = k$ and $\delta_{z^t,k} = 0$ otherwise, and $\omega_{i,j}^t = 1$ if $z^t = i$ and $z^{t+1} = j$, and $\omega_{i,j}^t = 0$ otherwise.
Then A.5 can be rewritten as

$$L_c(\Theta, \mathbf{X}, \mathbf{Y}, \mathbf{z}) = \sum_{i=1}^{K} \delta_{z^1,i} \ln(\pi_i) + \sum_{t=1}^{T-1} \sum_{i=1}^{K} \sum_{j=1}^{K} \omega_{i,j}^t \ln(a_{ij}) + \sum_{t=1}^{T} \sum_{i=1}^{K} \delta_{z^t,i} \ln(P(\mathbf{y}^t|\mathbf{x}^t, \mathbf{B}_i)) + \sum_{t=1}^{T} \sum_{i=1}^{K} \delta_{z^t,i} \ln P(\mathbf{x}^t|b_i). \quad \text{(A.6)}$$

Switching the sum operators,

$$L_c(\Theta, \mathbf{X}, \mathbf{Y}, \mathbf{z}) = \sum_{i=1}^{K} \delta_{z^1,i} \ln(\pi_i) + \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{t=1}^{T-1} \omega_{i,j}^t \ln(a_{ij}) + \sum_{i=1}^{K} \sum_{t=1}^{T} \delta_{z^t,i} \ln P(\mathbf{y}^t|\mathbf{x}^t, \mathbf{B}_i) + \sum_{i=1}^{K} \sum_{t=1}^{T} \delta_{z^t,i} \ln P(\mathbf{x}^t|b_i). \quad \text{(A.7)}$$

As the contribution of the parameters of interest to the data log-likelihood is decoupled in A.7, the maximization of each term of A.7 can be performed separately. Maximizing the data log-likelihood amounts to solving the following maximization problems, where the constraints which pertain to the transition and initial state probabilities have been indicated:

$$\underset{\mathbf{B}_i}{\text{maximize}} \quad \sum_{t=1}^{T} \delta_{z^t,i} \ln P(\mathbf{y}^t|\mathbf{x}^t, \mathbf{B}_i) \qquad i = 1, \ldots, K \quad \text{(A.8)}$$

$$\begin{aligned} &\underset{\pi}{\text{maximize}} && \sum_{j=1}^{K} \delta_{z^1,j} \ln \pi_j \\ &\text{subject to} && \sum_{j=1}^{K} \pi_j = 1 \end{aligned} \quad \text{(A.9)}$$

$$\begin{aligned} &\underset{\mathbf{A}}{\text{maximize}} && \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{t=1}^{T-1} \omega_{ij}^t \ln a_{ij} \\ &\text{subject to} && \sum_{j=1}^{K} a_{ij} = 1 \quad \forall i \in \{1, \ldots, K\}. \end{aligned} \quad \text{(A.10)}$$

$$\underset{b_i}{\text{maximize}} \quad \sum_{t=1}^{T} \delta_{z^t,i} \ln P(\mathbf{x}^t|b_i) \qquad i = 1, \ldots, K. \quad \text{(A.11)}$$

The maximization problems (A.8), (A.9), (A.10), and (A.11) correspond to ML estimation of the expert, initial state, transition and emission parameters, respectively. The impact of the expert and emission parameters on the likelihood is additionally separable with respect to each possible state (problems (A.8) and (A.11)).

A.2 Unsupervised training: E-step

The MSLM unsupervised training is completed via the Expectation-Maximization algorithm. During the E-step, the expectation of the data log-likelihood (A.7) is computed with respect to the latent vector $\mathbf{z}$, given the observations $\mathbf{X}$ and $\mathbf{Y}$ and using the current parameter estimate $\hat{\Theta}$:

$$E\left[L_c(\Theta^+, \mathbf{X}, \mathbf{Y}, \mathbf{z})|\Theta = \hat{\Theta}, \mathbf{X}, \mathbf{Y}\right] = \sum_{i=1}^{K} \gamma_i^1 \ln(\pi_i^+) + \sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{t=1}^{T-1} \xi_{i,j}^t \ln(a_{ij}^+) + \sum_{i=1}^{K}\sum_{t=1}^{T} \gamma_i^t \ln P(\mathbf{y}^t|\mathbf{x}^t, \mathbf{B}_i^+) + \sum_{i=1}^{K}\sum_{t=1}^{T} \gamma_i^t \ln P(\mathbf{x}^t|b_i^+), \quad \text{(A.12)}$$

where

$$\gamma_i^1 = E\left[\delta_{z^1,i}|\Theta = \hat{\Theta}, \mathbf{X}, \mathbf{Y}\right] = P(z^1 = i|\Theta = \hat{\Theta}, \mathbf{x}^{1:T}, \mathbf{y}^{1:T}), \quad \text{(A.13)}$$

$$\gamma_i^t = E\left[\delta_{z^t,i}|\Theta = \hat{\Theta}, \mathbf{X}, \mathbf{Y}\right] = P(z^t = i|\Theta = \hat{\Theta}, \mathbf{x}^{1:T}, \mathbf{y}^{1:T}), \quad \text{(A.14)}$$

$$\xi_{i,j}^t = E\left[\omega_{i,j}^t|\Theta = \hat{\Theta}, \mathbf{X}, \mathbf{Y}\right] = P(z^t = i, z^{t+1} = j|\Theta = \hat{\Theta}, \mathbf{x}^{1:T}, \mathbf{y}^{1:T}). \quad \text{(A.15)}$$

During the M-step, the parameter estimate $\hat{\Theta}$ is updated by maximizing the expectation of the data log-likelihood. It thus amounts to solving the following maximization problems:

$$\underset{\mathbf{B}_i}{\text{maximize}} \quad \sum_{t=1}^{T} \gamma_i^t \ln P(\mathbf{y}^t|\mathbf{x}^t, \mathbf{B}_i) \qquad i = 1, \ldots, K \quad \text{(A.16)}$$

$$\begin{aligned} &\underset{\pi}{\text{maximize}} && \sum_{j=1}^{K} \gamma_j^1 \ln \pi_j \\ &\text{subject to} && \sum_{j=1}^{K} \pi_j = 1 \end{aligned} \quad \text{(A.17)}$$

$$\begin{aligned} &\underset{\mathbf{A}}{\text{maximize}} && \sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{t=1}^{T-1} \xi_{i,j}^t \ln a_{ij} \\ &\text{subject to} && \sum_{j=1}^{K} a_{ij} = 1 \quad \forall i \in \{1, \ldots, K\}. \end{aligned} \quad \text{(A.18)}$$

$$\underset{\mathbf{b_i}}{\text{maximize}} \quad \sum_{t=1}^{T} \gamma_i^t \ln P(\mathbf{x}^t|b_i) \qquad i = 1, \ldots, K. \quad \text{(A.19)}$$

(A.16), (A.17), (A.18), and (A.19) correspond to ML estimation of the expert, initial state, transition and emission parameters, respectively. In contrast with the

maximization problems associated with MSLM supervised training, the impact of each state-specific parameter set is weighted by the state probabilities γ_1^t, γ_i^t or $\xi_{i,j}^t$ computed during the E-step. A decomposition of these probabilities is presented in the first two paragraphs below. The derivation of the MSLM-specific forward and backward algorithms necessary to compute intermediary probabilities is detailed in the next paragraphs.

State probability estimation The state probabilities $\gamma_i^t = P(z^t = i|\mathbf{x}^{1:T}, \mathbf{y}^{1:T}, \Theta)$ is decomposed as follows:

$$\begin{aligned}
\gamma_i^t &= P(z^t = i|\mathbf{x}^{1:T}, \mathbf{y}^{1:T}, \Theta) \\
&= \frac{P(z^t = i, \mathbf{x}^{1:t}, \mathbf{x}^{t+1:T}, \mathbf{y}^{1:t}, \mathbf{y}^{t+1:T}|\Theta)}{P(\mathbf{x}^{1:T}, \mathbf{y}^{1:T}|\Theta)} \\
&= \frac{P(\mathbf{y}^{t+1:T}, \mathbf{x}^{t+1:T}|z^t = i, \mathbf{y}^{1:t}, \mathbf{x}^{1:t}, \Theta)P(z^t = i, \mathbf{x}^{1:t}, \mathbf{y}^{1:t}|\Theta)}{P(\mathbf{x}^{1:T}, \mathbf{y}^{1:T}|\Theta)} \\
\gamma_i^t &= \frac{P(\mathbf{y}^{t+1:T}, \mathbf{x}^{t+1:T}|z^t = i, \Theta)P(z^t = i, \mathbf{x}^{1:t}, \mathbf{y}^{1:t}|\Theta)}{P(\mathbf{x}^{1:T}, \mathbf{y}^{1:T}|\Theta)}.
\end{aligned}$$

Transition probability estimation The transition probabilities $\xi_{i,j}^t = P(z^t = i, z^{t+1} = j|\mathbf{x}^{1:T}, \mathbf{y}^{1:T}, \Theta)$ are expressed as

$$
\begin{aligned}
\xi_{i,j}^t &= P(z^t = i, z^{t+1} = j | \mathbf{x}^{1:T}, \mathbf{y}^{1:T}, \Theta) \\
&= \frac{P(z^t = i, z^{t+1} = j, \mathbf{x}^{1:T}, \mathbf{y}^{1:T} | \Theta)}{P(\mathbf{x}^{1:T}, \mathbf{y}^{1:T} | \Theta)} \\
&= \frac{P(z^t = i, z^{t+1} = j, \mathbf{x}^{1:t}, \mathbf{x}^{t+1}, \mathbf{x}^{t+2:T}, \mathbf{y}^{1:t}, \mathbf{y}^{t+1}\mathbf{y}^{t+2:T} | \Theta)}{P(\mathbf{x}^{1:T}, \mathbf{y}^{1:T} | \Theta)} \\
&= \frac{P(\mathbf{x}^{t+2:T}, \mathbf{y}^{t+2:T} | z^t = i, z^{t+1} = j, \mathbf{x}^{1:t}, \mathbf{x}^{t+1}, \mathbf{y}^{1:t}, \mathbf{y}^{t+1}, \Theta)}{P(\mathbf{x}^{1:T}, \mathbf{y}^{1:T} | \Theta)} \\
&\qquad \times P(z^t = i, z^{t+1} = j, \mathbf{x}^{1:t}, \mathbf{x}^{t+1}, \mathbf{y}^{1:t}, \mathbf{y}^{t+1} | \Theta) \\
&= \frac{P(\mathbf{x}^{t+2:T}, \mathbf{y}^{t+2:T} | z^{t+1} = j, \Theta)}{P(\mathbf{x}^{1:T}, \mathbf{y}^{1:T} | \Theta)} \\
&\qquad \times P(\mathbf{x}^{t+1}, \mathbf{y}^{t+1} | z^t = i, z^{t+1} = j, \mathbf{x}^{1:t}, \mathbf{y}^{1:t}, \Theta) \\
&\qquad \times P(z^t = i, z^{t+1} = j, \mathbf{x}^{1:t}, \mathbf{y}^{1:t} | \Theta) \\
&= \frac{P(\mathbf{x}^{t+2:T}, \mathbf{y}^{t+2:T} | z^{t+1} = j, \Theta) P(\mathbf{x}^{t+1}, \mathbf{y}^{t+1} | z^{t+1} = j, \Theta)}{P(\mathbf{x}^{1:T}, \mathbf{y}^{1:T} | \Theta)} \\
&\qquad \times P(z^{t+1} = j | z^t = i, \mathbf{x}^{1:t}, \mathbf{y}^{1:t}, \Theta) P(z^t = i, \mathbf{x}^{1:t}, \mathbf{y}^{1:t} | \Theta) \\
&= \frac{P(\mathbf{x}^{t+2:T}, \mathbf{y}^{t+2:T} | z^{t+1} = j, \Theta) P(\mathbf{x}^{t+1}, \mathbf{y}^{t+1} | z^{t+1} = j, \Theta)}{P(\mathbf{x}^{1:T}, \mathbf{y}^{1:T} | \Theta)} \\
&\qquad \times P(z^{t+1} = j | z^t = i \Theta) P(z^t = i, \mathbf{x}^{1:t}, \mathbf{y}^{1:t} | \Theta).
\end{aligned}
$$

The computation of the intermediary probabilities $P(z^t = i, \mathbf{x}^{1:t}, \mathbf{y}^{1:t}|\Theta)$, $P(\mathbf{x}^{t+1:T}, \mathbf{y}^{t+1:T}|z^t = j, \Theta)$ and $P(\mathbf{x}^{1:T}, \mathbf{y}^{1:T}|\Theta)$ is performed using extensions of the forward and backward algorithms originally developed for HMMs [Rabiner, 1989]. They extend the HMM-specific ones by taking into account the likelihood of each continuous expert when estimating the probability associated with each possible hidden state.

MSLM forward algorithm The MSLM forward algorithm permits to compute the probabilities $P(z^t = i, \mathbf{x}^{1:t}, \mathbf{y}^{1:t}|\Theta), t = 1, \ldots, T$. The recurrence is initialized by $P(z^1 = j, \mathbf{x}^1, \mathbf{y}^1|\Theta) = P(\mathbf{y}^1|\mathbf{x}^1, z^1 = j, \Theta)P(\mathbf{x}^1|z^1 = j)P(z^1 = j)$. The following recurrence formula is then utilized:

$$
\begin{aligned}
P(z^{t+1}=j,\mathbf{x}^{1:t+1},\mathbf{y}^{1:t+1}|\Theta) &= P(z^{t+1}=j,\mathbf{x}^{1:t},\mathbf{y}^{1:t},\mathbf{x}^{t+1},\mathbf{y}^{t+1}|\Theta)\\
&= P(\mathbf{x}^{t+1},\mathbf{y}^{t+1}|\mathbf{x}^{1:t},\mathbf{y}^{1:t},z^{t+1}=j,\Theta)P(\mathbf{x}^{1:t},\mathbf{y}^{1:t},z^{t+1}=j|\Theta)\\
&= P(\mathbf{x}^{t+1},\mathbf{y}^{t+1}|z^{t+1}=j,\Theta)\sum_i P(\mathbf{x}^{1:t},\mathbf{y}^{1:t},z^{t+1}=j,z^t=i|\Theta)\\
&= P(\mathbf{x}^{t+1},\mathbf{y}^{t+1}|z^{t+1}=j,\Theta)\\
&\quad\times\sum_i P(z^{t+1}=j|z^t=i,\mathbf{x}^{1:t},\mathbf{y}^{1:t},\Theta)P(z^t=i,\mathbf{x}^{1:t},\mathbf{y}^{1:t}|\Theta)\\
P(z^{t+1}=j,\mathbf{x}^{1:t+1},\mathbf{y}^{1:t+1}|\Theta) &= P(\mathbf{x}^{t+1},\mathbf{y}^{t+1}|z^{t+1}=j,\Theta)\\
&\quad\times\sum_i P(z^{t+1}=j|z^t=i,\Theta)P(z^t=i,\mathbf{x}^{1:t},\mathbf{y}^{1:t}|\Theta),
\end{aligned}
$$

where $P(z^{t+1}=j|z^t=i,\Theta)=a_{ij}$ and $P(\mathbf{x}^t,\mathbf{y}^t|z^t=j,\Theta)=P(\mathbf{y}^t|\mathbf{x}^t,\mathbf{B}_j)P(\mathbf{x}^t|b_j)$.

MSLM backward algorithm Similarly, the MSLM backward algorithm permits to compute $P(\mathbf{x}^{t+1:T},\mathbf{y}^{t+1:T}|z^t=i,\Theta)$. Following the generic HMM-specific backward algorithm [Fink, 2014], the recurrence is initialized with $P(\mathbf{x}^T,\mathbf{y}^T|z^T=j,\Theta)=(1,\ldots,1)'$. The following iterations are then completed:

$$
\begin{aligned}
P(\mathbf{x}^{t+1:T},\mathbf{y}^{t+1:T}|z^t=i,\Theta) &= \sum_j P(\mathbf{x}^{t+1:T},\mathbf{y}^{t+1:T},z^{t+1}=j|z^t=i,\Theta)\\
&= \sum_j P(\mathbf{x}^{t+1},\mathbf{x}^{t+2:T},\mathbf{y}^{t+1},\mathbf{y}^{t+2:T},z^{t+1}=j|z^t=i,\Theta)\\
&= \sum_j P(\mathbf{x}^{t+1},\mathbf{y}^{t+1}|\mathbf{x}^{t+2:T},\mathbf{y}^{t+2:T},z^{t+1}=j,z^t=i,\Theta)\\
&\quad\times P(\mathbf{x}^{t+2:T},\mathbf{y}^{t+2:T},z^{t+1}=j|z^t=i,\Theta)\\
&= \sum_j P(\mathbf{x}^{t+1},\mathbf{y}^{t+1}|z^{t+1}=j,\Theta)\\
&\quad\times P(\mathbf{x}^{t+2:T},\mathbf{y}^{t+2:T}|z^{t+1}=j,z^t=i,\Theta)P(z^{t+1}=j|z^t=i,\Theta)\\
&= \sum_j P(\mathbf{x}^{t+1},\mathbf{y}^{t+1}|z^{t+1}=j,\Theta)\\
&\quad\times P(\mathbf{x}^{t+2:T},\mathbf{y}^{t+2:T}|z^{t+1}=j,\Theta)P(z^{t+1}=j|z^t=i,\Theta),
\end{aligned}
$$

where $P(z^{t+1}=j|z^t=i,\Theta)=a_{ij}$ and $P(\mathbf{x}^t,\mathbf{y}^t|z^t=j,\Theta)=P(\mathbf{y}^t|\mathbf{x}^t,\mathbf{B}_j)P(\mathbf{x}^t|b_j)$.

Appendix B

MSLM: expert training procedure and gate selection

Contents

The relevance of training procedures for linear experts and of approaches for the modelling of the HMM-based gate's emission probabilities depends on the problem at hand. In particular, it can be impacted by the measurement noise associated with the experts, by the dimensionality of the neural features and by their underlying distribution. A comparative study was therefore completed to choose the training approaches and/or structures which corresponded to optimal MSLM experts and gate for the two considered data sets, namely the preclinical and clinical data sets presented in **Chapter 4**. Supervised training procedures only were explored in this study, i.e. NC/IC and NC/IC$_i, i = 1, \ldots, 5$ labels were available in the preclinical and clinical training data subsets, respectively.

B.1 Expert training procedure selection

The MSLM exploits linear regression models (experts) to reconstruct kinematic parameters during IC periods. It was assumed that these linear models were associated with Gaussian noises. The corresponding generic ML training approach, i.e. Optimal Least Squares (OLS), is generally unsuitable for high-dimensional independent variables. A study was therefore completed to select the best training procedure to approximate the ML estimator of the IC and IC$_i, i = 1, \ldots, 5$ regression models for the preclinical and clinical data set, respectively.

B.1.1 Methods and implementation

Penalized and/or projection-based training approaches were chosen on the basis of their utilization in previous BCI studies, and used to identify the MSLM's (possibly multiple) IC experts.

Linear models have been used for kinematic reconstruction from ECoG signals in several BCI studies (see **section 2.4.3.1**). Several training strategies have been resorted to when high-dimensional neural features were considered. The relevance of popular training approaches, namely ℓ^1-penalized regression (i.e., LASSO), Principal Component Regression (PCR) and Partial Least Squares Regression (PLSR), was assessed for trajectory reconstruction during IC periods on both the preclinical and clinical data sets. All models were trained and tested on IC or $\text{IC}_i, i = 1, \ldots, 5$ samples exclusively.

Principal Component Regression (PCR) The utilization of PCR, i.e. the regression of kinematic parameters against neural signals' principal components, has been considered in several BCI studies, e.g. [Spüler et al., 2016]. The number of extracted principal components was here chosen so that the resulting reduced variable explained a specific percentage of $\mathbf{x}^t$ variance, here 20%, 40%, 60% and 80% of $\mathbf{x}^t$, respectively. The corresponding linear models are referred to as "PCR_{20}", "PCR_{40}", "PCR_{60}" and "PCR_{80}", respectively.

Partial Least Squares Regression (PLSR) PLS regression has been found efficient for kinematic trajectory reconstruction from ECoG features in several studies [Chao et al., 2010] [Shimoda et al., 2012] [Bundy et al., 2016]. This training method is able to handle high dimensional and/or correlated explanatory variables, especially when the explanatory variable's dimension is higher than the number of training samples. The PLS estimator is a shrinkage estimator [Lingjaerde and Christophersen, 2000], i.e. it shrinks parameters towards zero. It thus exploits the prior knowledge that most features are irrelevant. As a result, the PLS estimator is biased but its variance is lower than the OLS variance.

PLS-based estimation of the parameters $\mathbf{B}$ of a linear model between $\mathbf{x}^t$ and $\mathbf{y}^t$ relies on the assumption that the dependence between $\mathbf{x}^t$ and $\mathbf{y}^t$ can be satisfyingly modelled by a linear model between low-dimensional projections of $\mathbf{x}^t$ and $\mathbf{y}^t$. Let the latent variables $\mathbf{t}^t \in \mathbb{R}^F$ and $\mathbf{u}^t \in \mathbb{R}^F$ be the projections ("scores") of $\mathbf{x}^t$ and $\mathbf{y}^t$ onto the low-dimensional subspaces $\tilde{\text{X}} \subset \mathbb{R}^m$ and $\tilde{\text{Y}} \subset \mathbb{R}^n$, with $dim(\tilde{\text{X}}) = dim(\tilde{\text{Y}}) = F$. $\tilde{\text{X}}$ and $\tilde{\text{Y}}$ are found jointly, on the criterion that they maximize the covariance between $\mathbf{t}^t$ and $\mathbf{u}^t$. The Nonlinear Iterative Partial Least Squares (NIPALS) [Höskuldsson, 1988] or SIMPLS [de Jong, 1993] algorithms are generally used to perform PLS regression.

The number of PLS factors used to reflect the dependence between the independent and dependent variables is a hyperparameter which value has to be fixed before performing PLS regression. Although the utilization of information criteria (e.g., the Bayesian Information Criterion) has recently begun to be explored [Krämer, 2011],

the optimal dimension F, i.e. the optimal number of PLS factors, is typically found using k-fold Cross-Validation-based criteria [Geladi and Kowalski, 1986] [Wold et al., 2001] [Li et al., 2002]. For each considered number of factors f, a PLS model is trained using (k-1) folds of the data. The corresponding PRESS (PRedicted Error Sum of Squares) statistic is computed on the remaining fold [Li et al., 2002]. The total PRESS is obtained by repeating these steps for all (k-1) fold combinations, and by summing the corresponding fold-specific PRESS statistics [Li et al., 2002]. The utilization of several criteria has been proposed to infer the optimal number of factors f from the total PRESS statistic. Wold's R criterion was here applied on the 6-fold cross-validated PRESS statistic. It is defined as $R_{Wold} = \frac{\text{PRESS}(f+1)}{\text{PRESS}(f)}$ [Li et al., 2002]. The chosen F corresponds to $R_{Wold} > 1$, i.e. the procedure is stopped when the addition of a new factor increases the error. The corresponding model is denoted by "PLS".

Penalized regression Penalized training has been used for kinematic decoding in several BCI studies, in particular [Li et al., 2009] [Suminski et al., 2010] [Flamary and Rakotomamonjy, 2012] [Shanechi et al., 2013] [Williams et al., 2013] [Willett et al., 2013] [Spüler et al., 2016]. LASSO regression, which application has been reported for ECoG kinematic decoding in [Williams et al., 2013] and [Spüler et al., 2016], was here considered. It relies on ℓ^1-penalization to identify sparse linear models. The amount λ of ℓ^1-penalization was here chosen by 6-fold cross-validation. LASSO was chosen over Elastic Net (i.e., linear combination of ℓ^1 and ℓ^2 regularization) training because the latter requires the estimation of two hyperparameters, namely the amount of regularization and the balance between ℓ^1 and ℓ^2 regularization. Grid search over two parameters being time-consuming, Elastic Net was deemed less suited to BCI applications than LASSO training. The open-source glmnet toolbox [Qian et al., 2013], which relies on an efficient coordinate descend optimization procedure [Friedman et al., 2010], was used to perform LASSO training.

B.1.2 Results

Comparisons between the 6 training procedures (PCR_{20}, PCR_{40}, PCR_{60}, PCR_{80}, PLS and LASSO) were performed on the preclinical and clinical data sets. Performances were assessed by computing the PCC (see **section 5.2.3**) between the true and estimated limb positions during IC periods. The Friedman test was used to assess the significance of the observed performance differences between classifiers (see **section 5.3**). Post-hoc testing was carried out by applying the sign test to the best decoder and each one of the other decoders. The significance level α was corrected to account for the multiple pairwise comparisons which were completed, here 5 pairwise comparisons. The corrected significance level $\alpha_{post-hoc} = \frac{0.05}{5} = 0.01$ was thus utilized to measure the significance of the performance differences observed between the best decoder and the other ones.

Table B.1: Preclinical data set, median PCC between reconstructed and true trajectories, linear IC experts (median over 8 and 16 sessions for the subdural and epidural preclinical data sets, respectively).

		PCR_{20}	PCR_{40}	PCR_{60}	PCR_{80}	LASSO	PLS
subdural	y_1	0.26	0.28	0.36	0.24	0.42	0.43
	y_2	0.51	0.64	0.70	0.63	0.73	0.74
	y_3	0.42	0.55	0.63	0.59	0.69	0.67
epidural	y_1	0.08	0.07	0.09	0.10	0.11	0.09
	y_2	0.20	0.36	0.38	0.36	0.46	0.45
	y_3	0.21	0.28	0.32	0.27	0.41	0.35

B.1.2.1 Preclinical data set

Reconstruction performances obtained on the preclinical and clinical data sets are shown in **Table B.1**. Friedman-based decoder ranking is presented in **Table B.2**. The Friedman test indicated that the 6 decoders did not behave equivalently. Post-hoc pair-wise tests were therefore completed (see corresponding p-values in **Table B.2**).

B.1.2.2 Clinical data set

Reconstruction performances obtained on the clinical data set are shown in **Table B.3**. Friedman-based decoder ranking is presented in **Table B.4**. The Friedman test indicated that the 6 considered linear experts did not behave equivalently. Post-hoc pairwise tests were therefore completed (see corresponding p-values in **Table B.4**).

Table B.2: Preclinical data set, ranking of the linear IC experts' training approaches. The Friedman test was used to rank the continuous decoders on the basis of the associated PCC (averaged over axes). Pairwise post-hoc tests with the corrected significance level $\alpha_{post-hoc} = 0.01$ were performed to assess the significance of the performance differences between the best decoder and the other ones. Decoders which exhibited a non-significant performance difference with the best decoder are indicated with a bold font.

	Subdural		Epidural	
Rank	Decoder	p-value	Decoder	p-value
1	**LASSO**		**LASSO**	
2	**PLS**	0.727	**PLS**	0.455
3	PCR_{60}	0.008	$\mathbf{PCR_{80}}$	0.02
4	PCR_{40}	0.008	PCR_{60}	0.004
5	PCR_{80}	0.008	PCR_{40}	0.004
6	PCR_{20}	0.008	PCR_{20}	< 0.001

Table B.3: Clinical data set, median PCC between reconstructed and true trajectories, linear IC experts (median over 9 sessions).

		PCR_{20}	PCR_{40}	PCR_{60}	PCR_{80}	LASSO	PLS
fingers	1	0.05	0.15	0.12	0.20	0.35	0.25
	2	0.11	0.22	0.07	0.25	0.49	0.34
	3	0.17	0.24	0.20	0.27	0.51	0.34
	4	0.09	0.16	0.12	0.23	0.44	0.26
	5	0.12	0.19	0.17	0.28	0.45	0.34

Table B.4: Clinical data set, ranking of the linear IC experts' training approaches. The Friedman test was used to rank the continuous decoders on the basis of the associated PCC (averaged over fingers). Pairwise post-hoc tests with the corrected significance level $\alpha_{post-hoc} = 0.01$ were performed to assess the significance of the performance differences between the best decoder and the other ones. Decoders which exhibited a non-significant performance difference with the best decoder are indicated with a bold font.

Rank	Decoder	p-value
1	**LASSO**	
2	**PLS**	0.04
3	PCR_{80}	0.004
4	PCR_{40}	0.004
5	PCR_{60}	0.004
6	PCR_{20}	0.004

B.1.3 Discussion

PCR was found suboptimal when compared to LASSO and PLSR on both the preclinical and clinical data sets (see **Tables B.2** and **B.4**). The respective performances of PCR and PLSR are consistent with the theoritical properties of the PLSR, which has been shown to provide a closer fit than the Principal Component Regression [Phatak and De Hoog, 2002]. PCR and PLSR performances are additionally coherent with results reported in an earlier kinematic decoding study completed with ECoG clinical data [Spüler et al., 2016]. In this study, the extraction of principal components followed by ridge regression was surpassed by an analogue of PLSR, namely Canonical-Correlation Analysis [Spüler et al., 2016]. In contrast with the results obtained with LASSO linear models in this same study, LASSO models here slightly surpassed PLSR on both data sets. The amount λ of penalization was chosen with cross-validation. This fine tuning of λ may explain why LASSO models were outperformed by Canonical-Correlation Analysis in [Spüler et al., 2016], where λ was fixed at 0.1 rather than optimized for each acquisition session.

LASSO training is computationally more expensive than PLSR. The performance difference between LASSO and PLSR experts was found to be insignificant on both the preclinical and clinical data sets. This result may, however, reflect a lack of power of the statistical test. High p-values were associated with the difference between LASSO and PLSR on the preclinical data set (0.73 and 0.56 for the subdural and epidural data sets, respectively), and LASSO experts only outperformed PLSR experts by a median of 3% and 4% for the PCC on the subdural and epidural data sets, respectively (PCC averaged over axes). Because of these similar decoding performances and of its lower computational cost, PLSR was chosen over LASSO training for fast identification of linear experts on the preclinical data set. The

p-value corresponding to the performance difference between LASSO and PLSR was lower in the case of the clinical data set ($p = 0.04$), and was associated with a median improvement of 35% of the PCC (PCC averaged over fingers). The fact that the clinical ECoG electrodes were separated by an inter-electrode distance of 1cm and were not exclusively measuring motor activity (see data description in **Chapter 4**) may explain this difference between LASSO models and PLSR. Because of this sparse spatial sampling, it is probable that most electrodes do not contribute to the encoding of each finger movement. While PLS models are supposed to assign low weights to the features associated with such electrodes, the fact that these weights are low rather than null may introduce undesirable noise in the kinematic estimates. Because of the clear advantage of LASSO training over PLSR, it was decided to use LASSO training on the clinical data set despite its higher computational cost.

B.2 Gate selection

The state-specific emission distributions associated with the HMM-based gate of the MSLM can be modelled by means of generic distributions or by using an alternative approach based on discriminative modelling, i.e. both generative and discriminative state decoders can be embedded into the dynamic HMM-based gate of the MSLM (see **section 3.2.1**). Their relevance may in particular depend on the distribution of neural features within each state. If for example this distribution cannot be satisfyingly modelled with a generic distribution, e.g. a Gaussian distribution, using a discriminative approach to model the emission may permit to improve the gating procedure. A study was therefore completed to measure the relevance of different modelling approaches. For the sake of simplicity, this study was carried out on static decoders, i.e. the ability of several generative and discriminative static state decoders to distinguish between NC and (possible multiple) IC states was compared on the preclinical and clinical data sets considered in the present doctoral work. The best static decoders were chosen to be subsequently integrated into the MSLM's dynamic gate.

As BCI classifier performances have been found to depend on the dimension of the input variable [Bhattacharyya et al., 2011], the considered classifiers were tested after different projection-based dimensionality reduction procedures had been performed on the input variable.

B.2.1 Methods and implementation

5 classifiers, namely LDA, QDA, LR, linear and non-linear SVM, were selected on the basis of their frequent utilization in BCI studies and/or on the high decoding performance they reached in previous studies. LDA and QDA are generative classifiers based on Gaussian distributions (see **Chapter 2**). They are thus static analogues of generic HMMs with Gaussian emission distributions. By contrast, LR and SVMs are discriminative classifiers (see **Chapter 2**). As a result, they correspond to static analogues of the HMMs which exploit a discriminative approach

to model the emission probabilities. They were combined with unsupervised and supervised dimensionality reduction steps, and tested for the task of binary and multi-class state detection on the preclinical and clinical data sets, respectively.

B.2.1.1 Binary and multi-class state decoders

LDA and QDA ML training of the parameters $\{\mu_i, \mathbf{\Sigma}_i\}$ associated with the multivariate Gaussian distributions $P(\mathbf{x}^t|z^t = i) = \mathcal{N}(\mu_i, \mathbf{\Sigma}_i)$, where $i = 1, 2$ for the preclinical data set and $i = 1, \ldots, 6$ for the clinical data set, was performed on the training samples $\mathbf{x}^t$ observed at times t such that $z^t = i$. The LDA covariance matrix $\mathbf{\Sigma} = \mathbf{\Sigma}_i = \mathbf{\Sigma}_j, i \neq j$ being shared by all classes [Friedman et al., 2001], it was fitted using the totality of the training samples. Equiprobable class prior probabilities $P(z^t = i) = P(z^t = j)$ were utilized for both the LDA and QDA classifiers.

Linear and nonlinear SVMs Following [Bhattacharyya et al., 2011] and [Bashashati et al., 2015], the nonlinear SVM was built using a Radial Basis Function (RBF) kernel. The one-against-one strategy was utilized to perform multi-class SVM classification because it has been reported as generally more efficient that the one-against-all one [Hsu and Lin, 2002] and has been used in several BCI studies [Bashashati et al., 2015]. One SVM was thus trained for each possible state pair i, j, $i \neq j$, and an Error-Correcting Output Code was used to combine the state estimates yielded by each binary classifier [Dietterich and Bakiri, 1995]. Each class is associated with a specific combination of the SVM outputs. During application, the output of each SVM is computed for the considered observation sample. The class which corresponds to the combination of SVM outputs which is the closest to the one obtained on the observation features (i.e., the one which minimize the number of different outputs) is attributed to the input variable [Dietterich and Bakiri, 1995].

Logistic Regression A LR model and its multi-class extension, namely a multinomial LR, were fitted on each session of the preclinical and clinical data sets. The ML estimator of LR parameters is known to be unstable when it is applied to a training data set with linearly separable classes [Bishop, 2006]. Penalized LR training has been found to be efficient in several BCI studies [Gouy-Pailler et al., 2009] [Ryali et al., 2010] [Bashashati et al., 2015] [Bundy et al., 2016]. Two training approaches were thus compared for the identification of LR and multinomial LR models: ML [Czepiel, 2012] and LASSO training [Friedman et al., 2010]. LASSO training consists in introducing ℓ^1-penalization to stabilize parameter estimation [Friedman et al., 2001]. It was again chosen over Elastic Net penalization to limit the number of hyperparameters (see **section B.1.1**). The amount λ of ℓ^1regularization was here found by 6-fold cross-validation.

B.2.1.2 Dimensionality reduction

BCI classifier performances depend on the characteristics of the input variable [Bashashati et al., 2015], in particular on its dimensionality [Bhattacharyya et al., 2011]. Supervised and unsupervised projection methods were compared for the reduction of the dimensionality of the input variable $\mathbf{x}^t$. Thus, the above-mentioned classifiers were trained and compared after 6 dimensionality reduction procedures had been applied on the input variable.

PCA-based dimensionality reduction PCA-based dimensionality reduction is frequently performed in BCI studies [Wang et al., 2009b][Ke and Li, 2009] [Suk and Lee, 2010] [Argunşah and Çetin, 2010] [Bhattacharyya et al., 2011]. PCA-reduced input variables of different sizes were here considered. Following [Suk and Lee, 2010], the number of extracted principal components was chosen so that the resulting reduced variable explained a specific percentage of $\mathbf{x}^t$ variance, here 20%, 40%, 60% and 80% of $\mathbf{x}^t$, respectively. The corresponding reduced variables are henceforth referred to as "PCA_{20}", "PCA_{40}", "PCA_{60}" and "PCA_{80}", respectively.

PLS-based dimensionality reduction The PLS algorithm has been designed for the identification of regression models. As such, it has been profitably applied for kinematic decoding in several BCI studies [Shimoda et al., 2012] [Eliseyev and Aksenova, 2014] [Eliseyev and Aksenova, 2016] [Bundy et al., 2016] [van Gerven et al., 2012] (see **Chapter 2**). PLS-based dimensionality reduction combined with classification has nevertheless been shown to be efficient for the decoding of discrete variables [Kemsley, 1996] [Nguyen and Rocke, 2002] [Barker and Rayens, 2003] [Turkmen and Billor, 2012]. Satisfying classification accuracies have been reported when such combinations have been utilized in BCI studies [Eliseyev et al., 2011] [Eliseyev et al., 2012].

PLS-based projectors were therefore fitted to perform supervised dimensionality reduction before applying the considered generic classifiers. The first F PLS-scores of $\mathbf{x}^t$, where the state dummy variable $\mathbf{z}^t$ was regressed against the neural features $\mathbf{x}^t$ [Rosipal and Krämer, 2006], were fed to the classifiers. The optimal number of PLS factors was found using Wold's R criterion with 6-fold cross-validation on the training data set. The resulting reduced input variables are denoted by PLS_1 in the following sections.

The PLS_1 reduced input variables and the corresponding classifiers were fitted on the same data sets. Another training strategy consists in using independent data sets to train the PLS projectors and the classifiers, thereby reducing the risk of overfitting. A reduced variable denoted by "PLS_2" was therefore built by training a PLS model on the first 40% of the training data set. The second part of the training data set (60%) was used to fit the classifiers.

Finally, the full input variable $\mathbf{x}^t$ is referred to as "full" in the remaining of the present appendix.

Table B.5: Preclinical data set, average dimension of the feature subsets. The average was computed over 5, 3, 6 and 10 acquisition sessions for Monkeys A, K, B and C, respectively.

	PCA_{20}	PCA_{40}	PCA_{60}	PCA_{80}	PLS_1	PLS_2	full
NHP A	7 ± 3	49 ± 12	231 ± 53	751 ± 105	6 ± 1	5 ± 1	12160 ± 0
NHP K	8 ± 3	56 ± 22	268 ± 133	988 ± 395	8 ± 1	8 ± 1	24747 ± 0
NHP B	2 ± 1	12 ± 4	63 ± 22	332 ± 84	10 ± 1	9 ± 6	24320 ± 0
NHP C	3 ± 2	16 ± 9	61 ± 35	273 ± 196	8 ± 1	7 ± 3	24320 ± 0

B.2.2 Results

The classification accuracy $ACC = \frac{TP+TN}{TP+TN+FP+FN}$ was used to assess the performance of each combination reduced feature / classifier for NC/IC discrimination. In the case of the clinical data set, it was computed after having pooled the active states $IC_i, i = 1, \ldots, 5$ into a global active state IC. The ratio of correctly classified samples was additionally computed on TP observations to measure the ability of each decoder to distinguish between fingers. The Friedman test and post-hoc tests were used to assess the significance of the observed performance differences between classifiers (see **Chapter 5** for details on the validation methodology).

B.2.2.1 Preclinical data set

The number of extracted PCA components and PLS factors varied from one data set to another. **Table B.5** summarizes the average dimensionality of each feature subset. Some dimensionality reduction/classifier pairs were deemed irrelevant because of the inadequacy of the classifier for high-dimensional input variables, or for deterrent training durations. Generic LDA and QDA are for example known to be ill-suited for high-dimensional feature classification, in particular because of difficulties to estimate high-dimensional covariance matrices [Friedman, 1989]. Similarly, instabilities of the ML estimates of LR parameters arise in high-dimensional settings [Zhang et al., 2007]. LDA, QDA and ML LR were thus trained exclusively when the reduced variables exhibited limited sizes.

Classification performances obtained on the subdurally and epidurally implanted monkeys are shown in **Table B.6**. Discarded dimensionality reduction/classifier pairs are indicated by grey cells. The Friedman test indicated that all pairs did not perform similarly. Friedman-based classifier ranking is presented in **Table B.7**.

B.2.2.2 Clinical data set

The number of extracted PCA components and PLS factors is displayed in **Table B.8**. Average performance indicators are shown in **Table B.9**. Similarly to the

Table B.6: Preclinical data set, binary classification accuracy (ACC). The classification accuracy achieved by each couple reduced feature / classifier was averaged over the subdural and epidural acquisition sessions (8 and 16 sessions, respectively). Grey cells indicate dimensionality reduction/classifier pairs which were discarded because the classifier was ill-suited for the corresponding feature dimensionality.

		ACC						
		PCA_{20}	PCA_{40}	PCA_{60}	PCA_{80}	PLS_1	PLS_2	full
sub.	LDA	0.71				0.87	0.85	
	QDA	0.79				0.91	0.89	
	ML-LR	0.86				0.92	0.92	
	LASSO-LR	0.85	0.90	0.92	0.92	0.92	0.92	0.92
	linear-SVM	0.86	0.90	0.91	0.90	0.92	0.92	0.90
	rbf-SVM	0.84	0.68	0.68	0.68	0.91	0.89	0.64
epi.	LDA	0.56				0.79	0.76	
	QDA	0.58				0.80	0.76	
	ML-LR	0.64				0.79	0.77	
	LASSO-LR	0.64	0.71	0.74	0.78	0.80	0.77	0.81
	linear-SVM	0.64	0.71	0.74	0.77	0.79	0.77	0.77
	rbf-SVM	0.64	0.60	0.55	0.55	0.76	0.72	0.55

Table B.7: Preclinical data set, ranking of the 6 most efficient couples reduced feature / classifier. The Friedman test was used to rank couples reduced feature / classifier. Pairwise post-hoc tests with the corrected significance level $\alpha_{post-hoc} = \frac{0.05}{5} = 0.01$ were performed to assess the significance of the performance differences between the best decoder and the 5 other ones. Decoders which exhibited a non-significant performance difference with the best decoder are indicated with a bold font.

	Subdural		Epidural	
Rank	decoder	p-value	decoder	p-value
1	**PLS_1 / LASSO LR**		**full / LASSO LR**	
2	**PLS_1 / ML LR**	1	**PLS_1 / ML LR**	0.02
3	**PLS_1 / linear SVM**	1	**PLS_1 / QDA**	0.02
4	**PCA_{80} / LASSO LR**	0.73	**PLS_1 / LASSO LR**	0.02
5	**PLS_2 / linear SVM**	0.29	**PLS_1 / linear SVM**	0.02
6	full / LASSO LR	0.008	**PLS_1 / LDA**	0.02

Table B.8: Clinical data set, average dimension of the feature subsets (9 acquisition sessions).

PCA_{20}	PCA_{40}	PCA_{60}	PCA_{80}	PLS_1	PLS_2	full
24 ± 17	209 ± 83	631 ± 135	1393 ± 182	18 ± 6	7 ± 1	21913 ± 3070

Table B.9: Clinical data set, ACC for NC/IC detection and ratio of correctly classified TP (Correct TP Ratio, CTPR). Both indicators were computed for each couple reduced feature / classifier, and averaged over the acquisition sessions (9 sessions). Grey cell are associated with the CTPR of some couples reduced feature / rbf SVM, because the SVM failed to output TP on some data sets and that the CTPR was consequently not defined.

		PCA_{20}	PCA_{40}	PCA_{60}	PCA_{80}	PLS_1	PLS_2	full
ACC	LDA	0.71				0.82	0.75	
	QDA	0.74				0.85	0.80	
	ML-LR	0.65				0.82	0.80	
	LASSO-LR	0.64	0.82	0.83	0.83	0.83	0.83	0.86
	linear-SVM	0.65	0.81	0.83	0.83	0.73	0.71	0.83
	rbf-SVM	0.36	0.36	0.36	0.36	0.80	0.78	0.36
CTPR	LDA	0.29				0.56	0.29	
	QDA	0.26				0.56	0.28	
	ML-LR	0.27				0.51	0.47	
	LASSO-LR	0.29	0.49	0.51	0.52	0.55	0.50	0.60
	linear-SVM	0.20	0.47	0.48	0.50	0.46	0.30	0.51
	rbf-SVM					0.52	0.36	

preclinical data set, irrelevant feature subsets / classifier pairs are indicated by grey cells. Ranking of the feature subset / classifier pairs is presented in **Table B.10**.

Table B.10: Clinical data set, ranking of the 6 most efficient couples reduced feature / classifier. The Friedman test was used to rank couples reduced feature / classifier. Pairwise post-hoc tests with the corrected significance level $\alpha_{post-hoc} = \frac{0.05}{5} = 0.01$ were performed to assess the significance of the performance differences between the best decoder and the 5 other ones. Decoders which exhibited a non-significant performance difference with the best decoder are indicated with a bold font.

	ACC		CTPR	
Rank	decoder	p-value	decoder	p-value
1	**full / LASSO LR**		**full / LASSO LR**	
2	**PLS_1 / QDA**	0.51	**PLS_1 / LDA**	0.04
3	**PCA_{60} / LASSO LR**	0.04	**PLS_1 / QDA**	0.51
4	full / linear SVM	0.004	PLS_1 / LASSO LR	0.004
5	**PCA_{60} / LASSO LR**	0.04	**PLS_1 / rbf SVM**	0.04
6	**PLS_2 / LASSO LR**	0.04	PLS_1 / ML LR	0.004

B.2.3 Discussion

The application of generic classifiers on PLS-reduced variables permitted to achieve high decoding performances for both the subdural and epidural acquisition sessions of the preclinical data set. In particular, it was observed that classifiers fed with PCA components either fell behind classifiers fed with PLS factors, or equated these classifiers. In the latter case, a higher number of PCA components was required to achieve the same performance as PLS-fed classifiers. These results are consistent with both theoretical [Barker and Rayens, 2003] and practical [Kemsley, 1996] [Nguyen and Rocke, 2002] studies previously completed outside the BCI community. Fitting both PLS regression and classifiers on the same data set was found to yield better results than separate training. The number of training samples for both the PLS models and classifiers is reduced when training is performed independently. This diminution may explain the poor decoding performance observed in the case of independent training. Satisfying results were obtained with high-dimensional variables when they were utilized in combination with a penalization approach (i.e., LR training with a LASSO penalization).

While linear SVMs were here found to generally outperform nonlinear ones on the preclinical data set, LDA was surpassed by QDA. These results are consistent with earlier reports on the variable respective relevance of nonlinear and linear classifiers for BCI decoding (see **Chapter 2**).

Generative models (LDA and QDA) were mainly outperformed by discriminant ones on the preclinica data set, e.g. LR or linear SVM. This observation suggests that the distribution of neural features within NC and IC states cannot be satisfyingly modelled for the considered preclinical data set with generic Gaussian distributions. The modelling of the MSLM's HMM-based gate emission probabilities is thus likely

to benefit from the utilization of alternative approaches based on discriminative modelling (see **section 3.2.1**). ML LR fed with PLS-based features was here chosen for discriminative modelling of the MSLM emission on the preclinical data set because it exhibited a high decoding accuracy on both the subdural and epidural acquisition sessions. Its training has the additional advantage of being computationally less expensive that LASSO training of a LR on the full feature set.

The interest of PLS-based dimensionality reduction was further observed on the clinical data set. As, however, LASSO-trained LR corresponded to CTPRs in average 9% higher than the ones associated with LDA or QDA fed with PLS-based features, it appears that PLS-based dimensionality reduction may be suboptimal for the multi-class problem at hand, or that discriminative modelling of the MSLM gate emission probabilities may outperform traditional generative modelling with generic Gaussian distributions for the considered clinical data set. LASSO LR was thus chosen to model the MSLM gate emission probabilities on the clinical data set despite its higher computational cost.

Appendix C

Switching Kalman Filtering for asynchronous ECoG decoding

Contents

C.1 Introduction: Kalman filter

The Kalman Filter (KF) is an algorithm which permits to infer a hidden trajectory $\mathbf{y}^t$ from noisy measurements $\mathbf{x}^t \in \mathbb{R}^m$ [Bishop, 2006] (see **Figure C.1**). The continuous response (state) variable $\mathbf{y}^t \in \mathbb{R}^n$ is composed by the trajectory coordinates and derivatives (velocity, acceleration, etc.). The KF applies to linear Gaussian state-space models [Bishop, 2006]. The following state-space model is generally considered in BCI studies [Wu et al., 2002] [Pistohl et al., 2008] [Wang et al., 2013b]:

$$\mathbf{y}^{t+1} = \mathbf{A}\mathbf{y}^t + \mathbf{w}^t \ , \tag{C.1}$$

$$\mathbf{x}^t = \mathbf{C}\mathbf{y}^t + \mathbf{v}^t \tag{C.2}$$

Here, the emission (C.2) and transition (C.1) models are linear with constant emission and transition matrices ($\mathbf{C} \in \mathbb{R}^{m\times n}$ and $\mathbf{A} \in \mathbb{R}^{n\times n}$, respectively). The noise processes $\mathbf{w}^t$ and $\mathbf{v}^t$ are independent and identically distributed sequences of random variables [Krishnamurthy, 2016] with $P(\mathbf{w}^t) \sim \mathcal{N}(0, \mathbf{\Gamma}), \mathbf{\Gamma} \in \mathbb{R}^{n\times n}$ and $P(\mathbf{v}^t) \sim \mathcal{N}(0, \mathbf{\Sigma}), \mathbf{\Sigma} \in \mathbb{R}^{m\times m}$. The initial state is characterized by $\mathbf{y}^1 = \hat{\mathbf{y}}^0 + \mathbf{u}^0$, with $P(\mathbf{u}^0) \sim \mathcal{N}(0, \mathbf{P}^0)$.

After training, typically performed using Ordinary Least Squares [Wu et al., 2002], the KF issues the estimate $\hat{\mathbf{y}}^t = E(\mathbf{y}^t|\mathbf{x}^{1:t})$ (see **algorithm 1**).

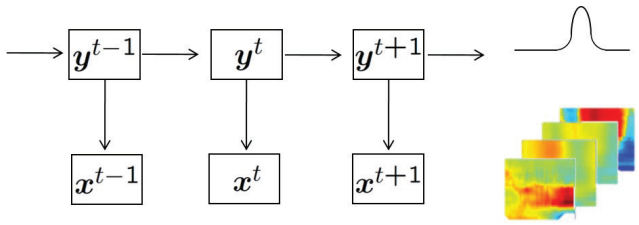

Figure C.1: KF structure. When the KF is used for ECoG filtering, the hidden trajectory $\mathbf{y}^t$ corresponds to the effector movement intended by the user, while the measurement $\mathbf{x}^t$ refers to the ECoG features.

Algorithm 1 Kalman Filter [Welch and Bishop, 2006]

Notations:

$$\hat{\mathbf{y}}^t = E(\mathbf{y}^t|\mathbf{x}^{1:t}) \tag{C.3}$$
$$\mathbf{P}^t = \mathrm{Var}(\hat{\mathbf{y}}^t - \mathbf{y}^t) \tag{C.4}$$
$$\hat{\mathbf{y}}^t_{prior} = E(\mathbf{y}^t|\mathbf{x}^{1:t-1}) \tag{C.5}$$
$$\mathbf{P}^t_{prior} = \mathrm{Var}(\hat{\mathbf{y}}^t_{prior} - \mathbf{y}^t). \tag{C.6}$$

Initialization: $\hat{\mathbf{y}}^0$, $\mathbf{P}^0$.

Time update:

$$\hat{\mathbf{y}}^t_{prior} = \mathbf{A}\hat{\mathbf{y}}^{t-1} \tag{C.7}$$
$$\mathbf{P}^t_{prior} = \mathbf{A}\mathbf{P}^{t-1}\mathbf{A}^T + \mathbf{\Gamma}. \tag{C.8}$$

Measurement update:

$$\mathbf{K}^t = \mathbf{P}^t_{prior}\mathbf{C}^T(\mathbf{C}\mathbf{P}^t_{prior}\mathbf{C}^T + \mathbf{\Sigma})^{-1} \tag{C.9}$$
$$\hat{\mathbf{y}}^t = \hat{\mathbf{y}}^t_{prior} + \mathbf{K}^t(\mathbf{x}^t - \mathbf{C}^T\hat{\mathbf{y}}^t_{prior}) \tag{C.10}$$
$$\mathbf{P}^t = (\mathbf{I} - \mathbf{K}^t\mathbf{C})\mathbf{P}^t_{prior}. \tag{C.11}$$

C.2 Switching Kalman Filter

C.2.1 Switching State-Space Model

The Switching Kalman Filter is a hybrid decoder which probabilistically combines K Kalman Filters [Murphy, 1998]. It extends the KF for hybrid discrete/continuous decoding.

Emission and transition distributions are conditioned on a switching variable $z^t \in Z$, with $Z = \{z_1, z_2 \ldots, z_K\} \subset \mathbb{N}$. z^t is not observed, and is assumed to be generated by a first-order Markov chain:

$$z^{t+1} = \mathbf{A}_{switch} z^t \tag{C.12}$$
$$\mathbf{y}^{t+1} = \mathbf{A}_{z^t} \mathbf{y}^t + \mathbf{w}_{z^t}^t \tag{C.13}$$
$$\mathbf{x}^t = \mathbf{C}_{z^t} \mathbf{y}^t + \mathbf{v}_{z^t}^t, \tag{C.14}$$

where $\mathbf{A}_{switch}$ is the switching transition matrix. Similarly to the matrices $\mathbf{A}_k$, $\mathbf{C}_k$, $\mathbf{\Gamma}_k$ and $\mathbf{\Sigma}_k$, the probability of the initial state $\mathbf{y}^1$ is conditioned on z^1: $\mathbf{y}^1 = \hat{\mathbf{y}}_{z^1}^0 + \mathbf{u}_{z^1}$, with $P(\mathbf{u}_{z^1}) \sim \mathcal{N}(0, \mathbf{P}_{z^1}^0)$. The initial probabilities $P(z^1 = z_k), k = 1 \dots K$ of the switching variable are gathered in $\pi \in \mathbb{R}^K$. The dependencies between variables are illustrated in **Figure C.2**.

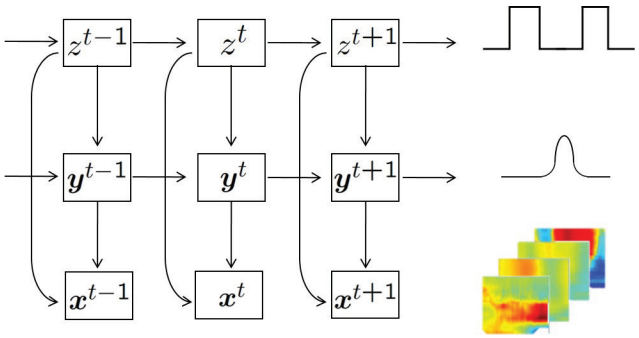

Figure C.2: SKF structure. When the SKF is used for ECoG filtering, the hidden trajectory $\mathbf{y}^t$ corresponds to the effector movement intended by the user, the hidden switching variable z^t to the neural state (for example, $z^t = 0$ and $z^t = 1$ for NC and IC states, respectively) and the measurement $\mathbf{x}^t$ to the ECoG features.

C.2.2 Training

Expectation-Maximization formula for unsupervised Maximum-Likelihood training of Switching Kalman Filters are presented in [Murphy, 1998]. Supervised Maximum-Likelihood training is possible when a complete training data set $\{\mathbf{X}, \mathbf{Y}, \mathbf{z}\}$ is available. Maximization formula [Murphy, 1998] for EM-based SKF training are applied, taking into account that the hidden switching state sequence is known. ML estimates of $\mathbf{A}_k$, $\mathbf{C}_k$, $\mathbf{\Gamma}_k$ and $\mathbf{\Sigma}_k$ are computed on the training data sets $\{\mathbf{X}_k, \mathbf{Y}_k\}$. The switching transition matrix $\mathbf{A}_{switch}$ is fitted on the basis of transition frequencies in the sequence $(z^t)_{t=1}^T$.

C.2.3 Application

The **algorithms 2** and **3** gather inference formula which permit to perform Switching Kalman Filtering. In contrast with Kalman filtering, exact estimation of $\hat{\mathbf{y}}^t = E(\mathbf{y}^t | \mathbf{x}^{1:t})$ is intractable because of the exponential number of possible state sequences. The SKF therefore relies on approximate solutions [Murphy, 1998]. The

collapsing routine exposed in [Murphy, 1998] and implemented in [Quinn, 2005] is here presented.

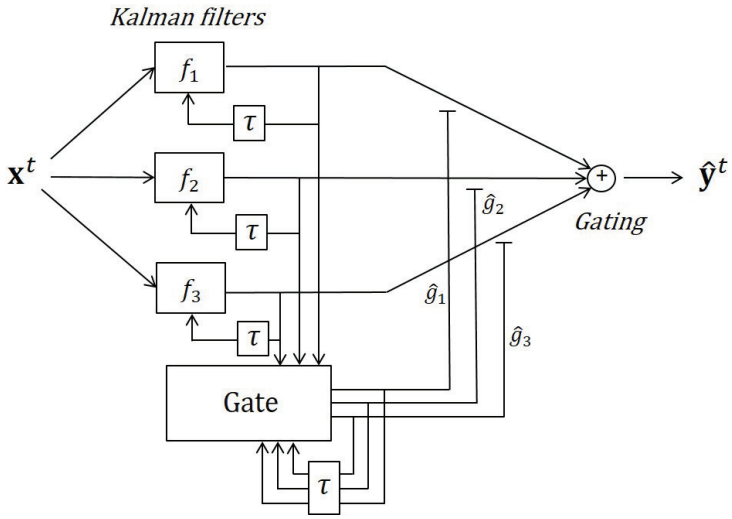

Figure C.3: SKF application.

Algorithm 2 Switching Kalman Filter [Murphy, 1998]

Notations:

$$\begin{aligned}
&\mathbf{P}_j^{t|\tau} = \mathrm{Cov}(\mathbf{y}^t|\mathbf{x}^{1:\tau}, z^t = j)\\
&\mathbf{P}_j^{t,t-1|\tau} = \mathrm{Cov}(\mathbf{y}^t, \mathbf{y}^{t-1}|\mathbf{x}^{1:\tau}, z^t = j)\\
&\mathbf{P}_{i(j)}^{t,t-1|\tau} = \mathrm{Cov}(\mathbf{y}^t, \mathbf{y}^{t-1}|\mathbf{x}^{1:\tau}, z^{t-1} = i, z^t = j)\\
&M^{t-1,t|\tau}(i,j) = P(z^{t-1} = i, z^t = j|\mathbf{x}^{1:\tau})\\
&M^{t|\tau}(j) = P(z^t = j|\mathbf{x}^{1:\tau})\\
&L_j^t = P(\mathbf{x}^t|\mathbf{x}^{1:t-1}, z^t = j)
\end{aligned}$$

Initialization: π, $\hat{\mathbf{y}}_i^0$, $\mathbf{P}_i^0$, $i = 1, \ldots, K.$,

Iterations:

$$(\mathbf{y}_{i(j)}^{t|t}, \mathbf{P}_{i(j)}^{t|t}, \mathbf{P}_{i(j)}^{t,t-1|t}, L_{i(j)}^t) = \mathrm{filter}(\mathbf{y}_i^{t-1|t-1}, \mathbf{P}_i^{t-1|t-1}, \mathbf{x}^t; \mathbf{A}_j, \mathbf{C}_j, \mathbf{\Gamma}_j, \mathbf{\Sigma}_j) \tag{C.15}$$

$$M^{t-1,t|t}(i,j) = P(z^{t-1} = i, z^t = j|\mathbf{x}^{1:t}) \tag{C.16}$$

$$= \frac{L^t(i,j)Z(i,j)M^{t-1|t-1}(i))}{\sum_i \sum_j L^t(i,j)Z(i,j)M^{t-1|t-1}(i)} \tag{C.17}$$

$$M^{t|t}(j) = \sum_i M^{t-1,t|t}(i,j) \tag{C.18}$$

$$W_{i|j} = P(z^{t-1} = i|z^t = j, \mathbf{x}^{1:t}) = \frac{M^{t-1,t|t(i,j)}}{M^{t|t(j)}} \tag{C.19}$$

$$(\mathbf{x}_j^{t|t}, \mathbf{P}_j^{t|t}) = \mathrm{Collapse}(\mathbf{y}_{i(j)}^{t|t}, \mathbf{P}_{i|j}^{t|t}, W_{i|j}^t) \tag{C.20}$$

Algorithm 3 Filter operator [Murphy, 1998]:

$(\mathbf{y}^{t|t}, \mathbf{P}^{t|t}, \mathbf{P}^{t,t-1|t}, L^t) = \text{filter}(\mathbf{y}^{t-1|t-1}, \mathbf{P}^{t-1|t-1}, \mathbf{x}^t; \mathbf{A}, \mathbf{C}, \mathbf{\Gamma}, \mathbf{\Sigma})$

Prediction:

$$\begin{aligned}\mathbf{y}^{t|t-1} &= \mathbf{A}\mathbf{y}^{t-1|t-1}\\ \mathbf{P}^{t|t-1} &= \mathbf{A}\mathbf{P}^{t-1|t-1}\mathbf{A}' + \mathbf{\Gamma}\end{aligned}$$

Innovation:

$$\begin{aligned}\mathbf{e}^t &= \mathbf{x}^t - \mathbf{C}\mathbf{y}^{t|t-1}\\ \mathbf{S}^t &= \mathbf{C}\mathbf{P}^{t|t-1}\mathbf{C}' + \mathbf{\Sigma}\\ \mathbf{K}^t &= \mathbf{P}^{t|t-1}\mathbf{C}'(\mathbf{S}^t)^{-1}\\ L^t &= \mathcal{N}(\mathbf{e}; 0, \mathbf{S}^t)\end{aligned}$$

Updated estimates:

$$\begin{aligned}\mathbf{y}^{t|t} &= \mathbf{y}^{t|t-1} + \mathbf{K}^t\mathbf{e}^t\\ \mathbf{P}^{t|t} &= (\mathbf{I} - \mathbf{K}^t\mathbf{C})\mathbf{P}^{t|t-1} = \mathbf{P}^{t|t-1} - \mathbf{K}^t\mathbf{S}^t(\mathbf{K}^t)'\\ \mathbf{P}^{t,t-1|t} &= (\mathbf{I} - \mathbf{K}^t\mathbf{C})\mathbf{A}\mathbf{P}^{t-1|t-1}\end{aligned}$$

Details on the collapsing operator used to approximate mixtures of Gaussians and mentioned in equation (C.20) can be found in [Murphy, 1998].

C.3 Optimization of Kalman Filtering for high-dimensional ECoG signals

The SKF exploits Kalman filters to reconstruct kinematic parameters during IC periods. The accuracy of the Kalman filters can be improved by choosing an optimal composition for the continuous state variable [Pistohl et al., 2008] and by introducing a relevant lag between observed neural features and resulting movements in the emission model [Wu et al., 2002]. The computation of each KF's update involves the inversion of a $m \times m$ matrix (see **section C.1**), where m is the dimension of the neural features. Matrix inversion is liable to be time-consuming when high-dimensional features are fed to the KF. The reduction of the neural features' dimension may thus be additionally required to make KF suitable for online application. A preliminary study was completed to select relevant state variable compositions, lags between neural features and kinematic parameters and dimensionality reduction procedures for efficient SKF-based ECoG decoding on the considered preclinical and clinical data sets.

C.3.1 Methods and implementation

C.3.1.1 Optimal lag

Because causal neural patterns are likely to precede the resulting limb movements with a specific time lag, the accuracy of KF estimates is liable to be improved by introducing a relevant lag between the observations and the corresponding hidden trajectory in the emission model. Uniform lags, i.e. lags which are similar for all neural features [Wu et al., 2002], were here investigated. 2 different lags were tested. Kalman filters were fed with features either associated with the last 100ms before the instant t considered for trajectory estimation (10^{th} time bin of the generic feature $\mathbf{x}^t$ presented in **Chapter 6**) or with features observed between t-200 to t-100ms (9^{th} time bin of the generic feature $\mathbf{x}^t$). The corresponding features are denoted by $\mathbf{x}^t_{KF}$.

C.3.1.2 Composition of the state variable

Following earlier ECoG decoding studies [Pistohl et al., 2008] [Wang et al., 2013b] [Marathe and Taylor, 2013], a first continuous state variable $\mathbf{y}^t_{KF}$ was composed of the wrist's or fingers' position and velocity. An alternative state variable composed of position only was additionally considered for comparison purposes.

C.3.1.3 Dimensionality reduction

The recourse to projection approaches such as PCA [Wu et al., 2003b] [Kao et al., 2013] or Factor Analysis [Sadtler, 2014] and to feature selection strategies [Malik et al., 2015] has been reported for Bayesian filtering in BCI studies. Projection-based methods were here considered because of their frequent application [Wu et al., 2003b] [Kao et al., 2013] [Sadtler, 2014] and of their simple implementation.

PCA Following [Wu et al., 2003b] [Kao et al., 2013], PCA-based dimensionality was here considered. The number of extracted principal components was chosen so that the resulting reduced variable explained a specific percentage of $\mathbf{x}^t_{KF}$ variance, here 20%, 40% and 60% of $\mathbf{x}^t_{KF}$, respectively. Higher percentages were not considered because they often corresponded to a high number of principal components, and were therefore ill-suited for online application of the KF. The reduced independent variables are referred to as "PCA_{20}", "PCA_{40}" and "PCA_{60}", respectively.

PLS An alternative PLS-based supervised dimensionality reduction was additionally considered. PLS regression between $\mathbf{x}^t_{KF}$ and $\mathbf{y}^t_{KF}$ [Rosipal and Krämer, 2006] was used to identify an informative low-dimensional subspace, which dimension was chosen by applying Wold's R criterion on the 6-fold cross-validated PRESS statistic [Li et al., 2002].

Table C.1: Preclinical data set, median PCC between reconstructed and true trajectories. Kalman filters were fed with neural features either associated with the last 100ms before the considered instant t (10^{th} time bin of the generic feature $\mathbf{x}^t$) or with features observed between t-200 to t-100ms (9^{th} time bin of the generic feature $\mathbf{x}^t$) (median over 8 and 16 sessions for the subdural and epidural data sets, respectively). The PCCs were averaged over the 3 considered axes within each acquisition session.

		Position				Position and velocity			
	neural features	PCA_{20}	PCA_{40}	PCA_{60}	PLS	PCA_{20}	PCA_{40}	PCA_{60}	PLS
sub.	t-100:t (ms)	0.22	0.40	0.59	0.64	0.38	0.52	0.64	0.66
	t-200:t-100 (ms)	0.23	0.39	0.56	0.64	0.37	0.49	0.59	0.64
epi.	t-100:t (ms)	0.16	0.18	0.24	0.36	0.35	0.35	0.39	0.47
	t-200:t-100 (ms)	0.16	0.18	0.22	0.35	0.34	0.36	0.39	0.44

OLS estimates of the transition and emission matrices were then computed using the reduced neural features (observations) and the position/velocity of the limbs of interest. Similarly, ML estimates of the emission and transition covariance matrices were computed following [Aggarwal et al., 2013].

C.3.2 Results

Comparisons between the different lags, state variable compositions and dimensionality reduction procedures were performed on the preclinical and clinical data sets presented in **Chapter 4**. Performance was assessed by computing the PCC (see **section 5.2.3**) between the true and estimated limb positions during IC periods. Importantly, the KFs were re-initialized using the tracked position and/or velocity at the beginning of each IC (or IC_i) period.

C.3.2.1 Preclinical data set

KF reconstruction performances over IC periods are displayed in **Table C.1**.

C.3.2.2 Clinical data set

KF reconstruction performances over IC periods are displayed in **Table C.2**.

C.3.3 Discussion

Continuous state variables composed of limb position and velocity were associated with higher PCC between true and estimated trajectories than continuous state variables only constituted by limb positions. These results are consistent with

Table C.2: Clinical data set, median PCC between reconstructed and true trajectories. Kalman filters were fed with features associated either with the neural signals elicited during the last 100ms before the instant t considered for the continuous state variable $\mathbf{y}_{KF}^{t}$ (10^{th} time bin of the generic feature $\mathbf{x}^{t}$) or with features observed between t-200 and t-100ms (9^{th} time bin of the generic feature $\mathbf{x}^{t}$) . The PCCs were averaged over the 5 considered fingers within each acquisition session.

	Position				Position and velocity			
neural features $\mathbf{x}^{t}$	PCA_{20}	PCA_{40}	PCA_{60}	PLS	PCA_{20}	PCA_{40}	PCA_{60}	PLS
t-100:t (ms)	0.28	0.32	0.38	0.41	0.35	0.42	0.45	0.52
t-200:t-100 (ms)	0.24	0.27	0.26	0.32	0.33	0.40	0.36	0.40

findings reported in earlier ECoG studies [Pistohl et al., 2008]. Additionally, the best decoding accuracy was obtained when the KF was fed with the neural features corresponding to the neural activity generated during the last 100ms before the considered instant t (10^{th} time bin of the generic feature $\mathbf{x}^{t}$). High numbers of PCA components (i.e., 60% of explained variance) led to improved KF decoding accuracies when optimal continuous state variable and lag were considered. A similar observation has been reported in [Kao et al., 2013], where feeding a KF with principal components accounting for 60% of MUA/SUA features permitted to achieve an online accuracy similar to the one obtained by using all the neural features. PLS-based dimensionality reduction nevertheless yielded the best KF IC decoding accuracy. Thus, somehow counter-intuitively, the application of a discriminant dimensionality reduction technique proved to be efficient for a subsequent application of a generative model. Although a better optimization of the number of principal components may permit to improve the accuracy of PCA-based KF, such procedure is likely to be time-consuming and to correspond to a less compact feature representation than the one obtained after PLS-based dimensionality reduction.

Importantly, the high decoding performance exhibited by Kalman filters is specific to the trial-per-trial analysis which was adopted in the present preliminary study, i.e. to that fact that the KF were correctly initialized at the beginning of each movement. In particular, Kalman Filtering has been found to suffer from drift effects in earlier BCI studies (e.g., [Hochberg et al., 2012]), and a degradation of the KF decoding accuracy is likely to be observed in real-time settings where positions and velocity are not correctly re-initialized at the beginning of each movement.

Bibliography

[Acharya et al., 2010] Acharya, S., Fifer, M. S., Benz, H. L., Crone, N. E., and Thakor, N. V. (2010). Electrocorticographic amplitude predicts finger positions during slow grasping motions of the hand. Journal of neural engineering, 7(4):046002. (Cited on pages 31 and 46.)

[Aflalo et al., 2015] Aflalo, T., Kellis, S., Klaes, C., Lee, B., Shi, Y., Pejsa, K., Shanfield, K., Hayes-Jackson, S., Aisen, M., Heck, C., Liu, C., and Andersen, R. (2015). Decoding motor imagery from the posterior parietal cortex of a tetraplegic human. Science, 348(6237):906–910. (Cited on pages 5 and 12.)

[Aggarwal et al., 2008] Aggarwal, V., Acharya, S., Tenore, F., Shin, H. C., Etienne-Cummings, R., Schieber, M. H., and Thakor, N. V. (2008). Asynchronous decoding of dexterous finger movements using M1 neurons. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 16(1):3–14. (Cited on page 47.)

[Aggarwal et al., 2013] Aggarwal, V., Mollazadeh, M., Davidson, A. G., Schieber, M. H., and Thakor, N. V. (2013). State-based decoding of hand and finger kinematics using neuronal ensemble and LFP activity during dexterous reach-to-grasp movements. Journal of neurophysiology, 109(12):3067–81. (Cited on pages 42, 52, 62, 65, 66, 120, 125 and 208.)

[Aggarwal et al., 2009] Aggarwal, V., Tenore, F., Acharya, Soumyadipta Schieber, M. H., and Thakor, V. N. (2009). Cortical decoding of individual finger and wrist kinematics for an upper-limb neuroprosthesis. In Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE, pages 4535–4538. (Cited on page 63.)

[Ajiboye et al., 2012] Ajiboye, a. B., Simeral, J. D., Donoghue, J. P., Hochberg, L. R., and Kirsch, R. F. (2012). Prediction of imagined single-joint movements in a person with high-level tetraplegia. IEEE Transactions on Biomedical Engineering, 59(10):2755–2765. (Cited on page 46.)

[Alessandri et al., 2008] Alessandri, A., Baglietto, M., and Battistelli, G. (2008). Maximum-likelihood Kalman Filtering for Switching Discrete-time Linear Systems. In Proceedings of the 47th IEEE Conference on Decision and Control, pages 3192–3198. (Cited on page 70.)

[Amiri et al., 2013] Amiri, S., Rabbi, A., Azinfar, L., and Fazel-Rezai, R. (2013). A Review of P300, SSVEP, and Hybrid P300/SSVEP Brain- Computer Interface Systems. Brain-Computer Interface Systems - Recent Progress and Future Prospects, 2013:1–8. (Cited on pages 2 and 3.)

[Anderson and Sijercic, 1996] Anderson, C. and Sijercic, Z. (1996). Classification of EEG signals from four subjects during five mental tasks. In Solving engineering problems with neural networks: proceedings of the conference on engineering applications in neural networks (EANN'96), pages 407–414. (Cited on pages 13 and 14.)

[Anderson et al., 2012] Anderson, N. R., Blakely, T., Schalk, G., Leuthardt, E. C., and Moran, D. W. (2012). Electrocorticographic (ECoG) correlates of human arm movements. Experimental brain research, 223(1):1–10. (Cited on pages 18, 23, 24, 27, 28, 39, 44 and 115.)

[Argunşah and Çetin, 2010] Argunşah, a. Ö. and Çetin, M. (2010). AR-PCA-HMM approach for sensorimotor task classification in EEG-based brain-computer interfaces. Proceedings - International Conference on Pattern Recognition, pages 113–116. (Cited on pages 44, 47, 56, 57 and 195.)

[Ashmore et al., 2012] Ashmore, R. C., Endler, B. M., Smalianchuk, I., Degenhart, A. D., Hatsopoulos, N. G., Tyler-Kabara, E. C., Batista, A. P., and Wang, W. (2012). Stable online control of an electrocorticographic brain-computer interface using a static decoder. Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference, 2012:1740–4. (Cited on pages 17, 18, 26 and 44.)

[Bai et al., 2011] Bai, O., Rathi, V., Lin, P., Huang, D., Battapady, H., Fei, D.-Y., Schneider, L., Houdayer, E., Chen, X., and Hallett, M. (2011). Prediction of human voluntary movement before it occurs. Clinical Neurophysiology, 122(2):364–372. (Cited on page 53.)

[Ball et al., 2009a] Ball, T., Kern, M., Mutschler, I., Aertsen, A., and Schulze-Bonhage, A. (2009a). Signal quality of simultaneously recorded invasive and non-invasive EEG. NeuroImage, 46(3):708–16. (Cited on pages 18 and 39.)

[Ball et al., 2009b] Ball, T., Schulze-Bonhage, A., Aertsen, A., and Mehring, C. (2009b). Differential representation of arm movement direction in relation to cortical anatomy and function. Journal of neural engineering, 6(1):016006. (Cited on pages 23, 24, 27, 42, 43 and 173.)

[Barachant et al., 2010] Barachant, A., Bonnet, S., Congedo, M., and Jutten, C. (2010). Riemannian geometry applied to BCI classification. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 6365 LNCS:629–636. (Cited on page 53.)

[Barachant et al., 2012] Barachant, A., Bonnet, S., Congedo, M., and Jutten, C. (2012). Multiclass Brain-Computer Interface Classification by Riemannian Geometry. IEEE Transactions on Biomedical Engineering, 59:920–928. (Cited on page 53.)

[Barker and Rayens, 2003] Barker, M. and Rayens, W. (2003). Partial least squares for discrimination. Journal of Chemometrics, 17(3):166–173. (Cited on pages 195 and 199.)

[Bashashati et al., 2007a] Bashashati, A., Fatourechi, M., Ward, R. K., and Birch, G. E. (2007a). A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals. Journal of neural engineering, 4(2):R32–57. (Cited on pages 4, 5, 13, 37, 41 and 75.)

[Bashashati et al., 2007b] Bashashati, A., Ward, R. K., and Birch, G. E. (2007b). Towards development of a 3-state self-paced brain-computer interface. Computational intelligence and neuroscience, 2007:84386. (Cited on pages 29 and 63.)

[Bashashati et al., 2015] Bashashati, H., Ward, R. K., Birch, G. E., and Bashashati, A. (2015). Comparing different classifiers in sensory motor brain computer interfaces. PLoS ONE, 10(6):1–17. (Cited on pages 43, 52, 53, 54, 55, 56, 119, 194 and 195.)

[Baxter et al., 2013] Baxter, B. S., Decker, A., and He, B. (2013). Noninvasive control of a robotic arm in multiple dimensions using scalp electroencephalogram. International IEEE/EMBS Conference on Neural Engineering, NER, pages 45–47. (Cited on pages 10, 11, 12, 13, 19, 20, 29 and 40.)

[Bell et al., 2008] Bell, C. J., Shenoy, P., Chalodhorn, R., and Rao, R. P. (2008). Control of a humanoid robot by a noninvasive brain-computer interface in humans. Journal of neural engineering, 5(2 PG - 214-220):214–220. (Cited on pages 12 and 14.)

[Benavoli et al., 2015] Benavoli, A., Corani, G., and Mangili, F. (2015). Should we really use post-hoc tests based on mean-ranks? Journal of Machine Learning Research, pages 1–10. (Cited on pages 112 and 128.)

[Bengio and Frasconi, 1996] Bengio, Y. and Frasconi, P. (1996). Input-output HMM's for sequence processing. IEEE Transactions on Neural Networks, 7(5):1231–1249. (Cited on pages 57, 75, 76 and 84.)

[Bhattacharyya et al., 2015] Bhattacharyya, S., Basu, D., Konar, A., and Tibarewala, D. N. (2015). Interval type-2 fuzzy logic based multiclass ANFIS algorithm for real-time EEG based movement control of a robot arm. Robotics and Autonomous Systems, 68:104–115. (Cited on pages 19, 29, 30, 46 and 51.)

[Bhattacharyya et al., 2011] Bhattacharyya, S., Khasnobish, A., Konar, A., Tibarewala, D. N., and Nagar, A. K. (2011). Performance analysis of left/right hand movement classification from EEG signal by intelligent algorithms. IEEE SSCI 2011 - Symposium Series on Computational Intelligence - CCMB 2011: 2011 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind,

and Brain, pages 1–8. (Cited on pages 43, 47, 50, 51, 52, 53, 54, 55, 56, 119, 193, 194 and 195.)

[Bilmes, 1998] Bilmes, J. a. (1998). A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. International Computer Science Institute, 4(510):126. (Cited on page 86.)

[Birbaumer et al., 1999] Birbaumer, N., Ghanayim, N., Hinterberger, T., Iversen, I., Kotchoubey, B., Kübler, a., Perelmouter, J., Taub, E., and Flor, H. (1999). A spelling device for the paralysed. Nature, 398(6725):297–298. (Cited on page 13.)

[Bishop, 2006] Bishop, C. M. (2006). Pattern Recognition and Machine Learning, volume 4. (Cited on pages 47, 51, 52, 53, 54, 61, 62, 72, 73, 119, 123, 175, 194 and 201.)

[Blabe et al., 2015] Blabe, C. H., Gilja, V., Chestek, C. a., Shenoy, K. V., Anderson, K. D., and Henderson, J. M. (2015). Assessment of brain-machine interfaces from the perspective of people with paralysis. Journal of neural engineering, 12(4):043002. (Cited on page 8.)

[Blackrock, 2016] Blackrock (2016). Blackrock's Utah Array. http://blackrockmicro.com/neuroscience-research-products/low-noise-ephys-electrodes/blackrock-utah-array/. (Cited on page 15.)

[Blackrock, 2017] Blackrock (2017). Blackrock Neuromed ECoG grid. http://blackrockneuromed.com/products/elite/. (Cited on page 18.)

[Blakely et al., 2009] Blakely, T., Miller, K. J., Zanos, S. P., Rao, R. P. N., and Ojemann, J. G. (2009). Robust, long-term control of an electrocorticographic brain-computer interface with fixed parameters. Neurosurgical focus, 27(July):E13. (Cited on pages 17, 18 and 44.)

[Blankertz et al., 2008] Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., and Müller, K. R. (2008). Optimizing spatial filters for robust EEG single-trial analysis. IEEE Signal Processing Magazine, 25(1):41–56. (Cited on page 41.)

[Boostani et al., 2007] Boostani, R., Graimann, B., Moradi, M. H., and Pfurtscheller, G. (2007). A comparison approach toward finding the best feature and classifier in cue-based BCI. Medical and Biological Engineering and Computing, 45(4):403–412. (Cited on pages 45 and 48.)

[Boostani and Moradi, 2004] Boostani, R. and Moradi, M. H. (2004). A new approach in the BCI research based on fractal dimension as feature and Adaboost as classifier. J. Neural Eng, 1:212–217. (Cited on page 45.)

[Bordes and Vandekerkhove, 2005] Bordes, L. and Vandekerkhove, P. (2005). Statistical Inference for Partially Hidden Markov Models. Communications in Statistics - Theory and Methods, 34(5):1081–1104. (Cited on page 176.)

[Bougrain et al., 2012] Bougrain, L., Rochel, O., Boussaton, O., and Havet, L. (2012). From the decoding of cortical activities to the control of a JACO robotic arm: a whole processing chain. Control Architecture of Robots (CAR), pages 1–7. (Cited on page 29.)

[Bourlard and Morgan, 1998] Bourlard, H. and Morgan (1998). Hybrid HMM/ANN systems for speech recognition: Overview and new research directions. Adaptive Processing of Sequences and Data Structures, 1387(1):389–417. (Cited on pages 77 and 118.)

[Bouton et al., 2016] Bouton, C. E., Shaikhouni, A., Annetta, N. V., Bockbrader, M. A., Friedenberg, D. A., Nielson, D. M., Sharma, G., Sederberg, P. B., Glenn, B. C., Mysiw, W. J., Morgan, A. G., Deogaonkar, M., and Rezai, A. R. (2016). Restoring cortical control of functional movement in a human with quadriplegia. Nature, 000(7602):1–13. (Cited on pages 12, 15, 29, 31, 35 and 43.)

[Bouveyron and Brunet, 2013] Bouveyron, C. and Brunet, C. (2013). Model-Based clustering of high-dimensional data : A review. Computational statistics and data analysis, Elservier. (Cited on page 121.)

[Bradberry et al., 2010] Bradberry, T. J., Gentili, R. J., and Contreras-Vidal, J. L. (2010). Reconstructing three-dimensional hand movements from noninvasive electroencephalographic signals. The Journal of neuroscience : the official journal of the Society for Neuroscience, 30(9):3432–7. (Cited on pages 23, 24, 27 and 59.)

[Bradberry et al., 2009] Bradberry, T. J., Rong, F., and Contreras-Vidal, J. L. (2009). Decoding center-out hand velocity from MEG signals during visuomotor adaptation. NeuroImage, 47(4):1691–700. (Cited on page 23.)

[BrainMaps, 2017] BrainMaps (2017). BrainMaps. BrainMaps.org. (Cited on page 10.)

[Brockwell et al., 2004] Brockwell, a. E., Rojas, a. L., and Kass, R. E. (2004). Recursive Bayesian Decoding of Motor Cortical Signals by Particle Filtering. Journal of Neurophysiology, 91(4):1899–1907. (Cited on pages 61 and 62.)

[Brodu et al., 2011] Brodu, N., Lotte, F., and Lécuyer, A. (2011). Comparative study of band-power extraction techniques for Motor Imagery classification. IEEE SSCI 2011 - Symposium Series on Computational Intelligence - CCMB 2011: 2011 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain, pages 95–100. (Cited on pages 43 and 44.)

[Brunner et al., 2007] Brunner, C., Naeem, M., Leeb, R., Graimann, B., and Pfurtscheller, G. (2007). Spatial filtering and selection of optimized components in four class motor imagery EEG data using independent components analysis. Pattern Recognition Letters, 28(8):957–964. (Cited on page 40.)

[Brunner et al., 2006] Brunner, C., Scherer, R., Graimann, B., Supp, G., and Pfurtscheller, G. (2006). Online control of a brain-computer interface using phase synchronization. IEEE Transactions on Biomedical Engineering, 53(12):2501–2506. (Cited on page 45.)

[Bruns, 2004] Bruns, A. (2004). Fourier-, Hilbert- and wavelet-based signal analysis: are they really different approaches? Journal of neuroscience methods, 137(2):321–32. (Cited on pages 44 and 45.)

[Bundy et al., 2016] Bundy, D. T., Pahwa, M., Szrama, N., and Leuthardt, E. C. (2016). Decoding three-dimensional reaching movements using electrocorticographic signals in humans. Journal of Neural Engineering, 13(2):026021. (Cited on pages 24, 27, 28, 39, 41, 44, 47, 52, 53, 59, 61, 63, 65, 70, 71, 72, 73, 74, 77, 78, 115, 119, 159, 163, 164, 165, 172, 188, 194 and 195.)

[Bundy et al., 2014] Bundy, D. T., Zellmer, E., Gaona, C. M., Sharma, M., Szrama, N., Hacker, C., Freudenburg, Z. V., Daitch, A., Moran, D. W., and Leuthardt, E. C. (2014). Characterization of the effects of the human dura on macro- and microelectrocorticographic recordings. Journal of neural engineering, 11(1):016006. (Cited on pages 18, 165 and 173.)

[Buzsáki et al., 2012] Buzsáki, G., Anastassiou, C. a., and Koch, C. (2012). The origin of extracellular fields and currents–EEG, ECoG, LFP and spikes. Nature reviews. Neuroscience, 13(6):407–20. (Cited on pages 4, 9, 14, 15 and 19.)

[Cacciatore and Nowlan, 1994] Cacciatore, T. W. and Nowlan, S. J. (1994). Mixtures of Controllers for Jump Linear and Non-Linear Plants. Advances in Neural Information Processing Systems 6, (i):719–726. (Cited on page 75.)

[Cappé et al., 2006] Cappé, O., Moulines, E., and Ryden, T. (2006). Inference in Hidden Markov models. (Cited on pages 81 and 82.)

[Cardillo, 2009] Cardillo, G. (2009). myfriedman: Friedman test for non parametric two way ANalysis Of VAriance. http://www.mathworks.com/matlabcentral/fileexchange/25882. (Cited on page 112.)

[Carmena, 2013] Carmena, J. M. (2013). Advances in neuroprosthetic learning and control. PLoS biology, 11(5):e1001561. (Cited on pages 6 and 25.)

[Carmena et al., 2003] Carmena, J. M., Lebedev, M. a., Crist, R. E., O'Doherty, J. E., Santucci, D. M., Dimitrov, D. F., Patil, P. G., Henriquez, C. S., and Nicolelis, M. a. L. (2003). Learning to control a brain-machine interface for reaching and grasping by primates. PLoS biology, 1(2):E42. (Cited on pages 15, 16, 29, 46, 59 and 176.)

[Carvalhaes and De Barros, 2015] Carvalhaes, C. and De Barros, J. A. (2015). The surface Laplacian technique in EEG: Theory and methods. International Journal of Psychophysiology, 97(3):174–188. (Cited on page 41.)

[Cassisi et al., 2012] Cassisi, C., Montalto, P., Aliotta, M., Cannata, A., and Pulvirenti, A. (2012). Similarity Measures and Dimensionality Reduction Techniques for Time Series Data Mining. Advances in Data Mining Knowledge Discovery and Applications, (January 2017):72–95. (Cited on page 109.)

[Chae et al., 2012] Chae, Y., Jeong, J., and Jo, S. (2012). Toward brain-actuated humanoid robots: Asynchronous direct control using an EEG-Based BCI. IEEE Transactions on Robotics, 28(5):1131–1144. (Cited on pages 3, 19, 40, 41 and 50.)

[Chao et al., 2010] Chao, Z. C., Nagasaka, Y., and Fujii, N. (2010). Long-term asynchronous decoding of arm motion using electrocorticographic signals in monkeys. Frontiers in neuroengineering, 3(March):3. (Cited on pages 17, 19, 24, 29, 41, 43, 44, 63, 64, 89, 90, 91, 115, 162 and 188.)

[Chatterjee et al., 2007] Chatterjee, A., Aggarwal, V., Ramos, A., Acharya, S., and Thakor, N. V. (2007). A brain-computer interface with vibrotactile biofeedback for haptic information. Journal of neuroengineering and rehabilitation, 4(1):40. (Cited on page 30.)

[Chen et al., 2014a] Chen, C., Shin, D., Watanabe, H., Nakanishi, Y., Kambara, H., Yoshimura, N., Nambu, A., Isa, T., Nishimura, Y., and Koike, Y. (2014a). Decoding grasp force profile from electrocorticography signals in non-human primate sensorimotor cortex. Neuroscience research, 83:1–7. (Cited on page 46.)

[Chen et al., 2009] Chen, C. W., Lin, C. C. K., and Ju, M. S. (2009). Hand orthosis controlled using brain-computer interface. Journal of Medical and Biological Engineering, 29(5):234–241. (Cited on page 29.)

[Chen et al., 2014b] Chen, W., Liu, X., and Litt, B. (2014b). Logistic-weighted regression improves decoding of finger flexion from electrocorticographic signals. Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference, 2014:2629–2632. (Cited on pages 52, 53, 74, 77, 78, 115 and 116.)

[Chiappa and Bengio, 2003] Chiappa, S. and Bengio, S. (2003). Hmm and iohmm modeling of eeg rhythms for asynchronous bci systems. Technical report, IDIAP. (Cited on pages 13, 14, 50, 51, 56, 57 and 58.)

[Chin et al., 2007] Chin, C. M., Popovic, M. R., Thrasher, A., Cameron, T., Lozano, A., and Chen, R. (2007). Identification of arm movements using correlation of electrocorticographic spectral components and kinematic recordings. Journal of neural engineering, 4:146–158. (Cited on pages 17, 26, 27, 42, 43, 51, 52 and 54.)

[Choi et al., 2009] Choi, K., Hirose, H., Sakurai, Y., Iijima, T., and Koike, Y. (2009). Prediction of arm trajectory from the neural activities of the primary motor cortex with modular connectionist architecture. Neural Networks, 22(9):1214–1223. (Cited on page 46.)

[Cincotti et al., 2007] Cincotti, F., Kauhanen, L., Aloise, F., Palomäki, T., Caporusso, N., Jylänki, P., Mattia, D., Babiloni, F., Vanacker, G., Nuttin, M., Marciani, M. G., and Millán, J. D. R. (2007). Vibrotactile feedback for brain-computer interface operation. Computational Intelligence and Neuroscience, 2007. (Cited on pages 5 and 30.)

[Cincotti et al., 2008] Cincotti, F., Mattia, D., Aloise, F., Bufalari, S., Schalk, G., Oriolo, G., Cherubini, A., Marciani, M. G., and Babiloni, F. (2008). Non-invasive brain-computer interface system: towards its application as assistive technology. Brain research bulletin, 75(6):796–803. (Cited on page 5.)

[Cincotti et al., 2003] Cincotti, F., Scipione, A., Timperi, A., Mattia, D., Marciani, M. G., Millán, J., Salinari, S., Bianchi, L., and Babiloni, F. (2003). Comparison of different feature classifiers for brain computer interfaces. In International IEEE/EMBS Conference on Neural Engineering, NER, pages 645–647. (Cited on pages 53 and 58.)

[ClinicalTrials.gov, 2016] ClinicalTrials.gov (2016). Brain Computer Interface: Neuroprosthetic Control of a Motorized Exoskeleton (BCI). https://clinicaltrials.gov/ct2/show/record/NCT02550522?term=clinatec. (Cited on pages 31 and 176.)

[Collinger et al., 2014] Collinger, J. L., Vinjamuri, R., Degenhart, A. D., Weber, D. J., Sudre, G. P., Boninger, M. L., Tyler-Kabara, E. C., and Wang, W. (2014). Motor-related brain activity during action observation: a neural substrate for electrocorticographic brain-computer interfaces after spinal cord injury. Frontiers in integrative neuroscience, 8(February):17. (Cited on page 173.)

[Collinger et al., 2013] Collinger, J. L., Wodlinger, B., Downey, J. E., Wang, W., Tyler-Kabara, E. C., Weber, D. J., McMorland, A. J. C., Velliste, M., Boninger, M. L., and Schwartz, A. B. (2013). High-performance neuroprosthetic control by an individual with tetraplegia. Lancet, 381(9866):557–64. (Cited on pages 5, 10, 11, 12, 15, 16, 24, 29, 30, 31, 42, 49, 59, 60 and 167.)

[Costecalde, 2012] Costecalde, T. (2012). Brain computer interface with chronic cortical electrode arrays for motor deficit compensation in motor disabled patients. Experimental study in rodents. PhD thesis, Université de Grenoble. (Cited on page 34.)

[Coyle et al., 2005] Coyle, D., Prasad, G., and McGinnity, T. M. (2005). A time-series prediction approach for feature extraction in a brain-computer interface.

IEEE Transactions on Neural Systems and Rehabilitation Engineering, 13(4):461–467. (Cited on page 45.)

[Coyle et al., 2007] Coyle, S. M., Ward, T. E., Markham, C. M., Ward, E., and Markham, C. M. (2007). Brain-computer interface using a simplified functional near-infrared spectroscopy system. Journal of neural engineering, 4(3):219–26. (Cited on page 4.)

[Curran et al., 2004] Curran, E., Sykacek, P., Stokes, M., Roberts, S., Penny, W., Johnsrude, I., and Owen, A. (2004). Cognitive Tasks for Driving a Brain Computer Interfacing System: A Pilot Study. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 12(1):48–54. (Cited on page 25.)

[Czepiel, 2012] Czepiel, S. A. (2012). Maximum Likelihood Estimation of Logistic Regression Models: Theory and Implementation. https://pdfs.semanticscholar.org/fd70/35ab3fca912c1f77c34ff0c4435d41ebd0d3.pdf. (Cited on page 194.)

[Dangi et al., 2014] Dangi, S., Gowda, S., Moorman, H. G., Orsborn, A. L., So, K., and Shanechi, M. (2014). Continuous Closed-Loop Decoder Adaptation with a Recursive Maximum Likelihood Algorithm Allows for Rapid Performance Acquisition in Brain-Machine Interfaces. Neural computation, 1872:1840–1872. (Cited on page 176.)

[Darmanjian et al., 2003] Darmanjian, S., Kim, S. P. K. S. P., Nechyba, M., Morrison, S., Principe, J., Wessberg, J., and Nicolelis, M. (2003). Bimodal brain-machine interface for motor control of robotic prosthetic. Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No.03CH37453), 4(October):3–8. (Cited on pages 56 and 57.)

[Davis and Goadrich, 2006] Davis, J. and Goadrich, M. (2006). The Relationship Between Precision-Recall and ROC Curves. Proceedings of the 23rd international conference on Machine learning, pages 233–240. (Cited on page 107.)

[de Jong, 1993] de Jong, S. (1993). SIMPLS: An alternative approach to partial least squares regression. Chemometrics and Intelligent Laboratory Systems, 18(3):251–263. (Cited on page 188.)

[Degenhart et al., 2016] Degenhart, A. D., Eles, J., Dum, R., Mischel, J. L., Smalianchuk, I., Endler, B., Ashmore, R. C., Tyler-Kabara, E. C., Hatsopoulos, N. G., Wang, W., Batista, A. P., and Cui, X. T. (2016). Histological evaluation of a chronicallyimplanted electrocorticographic electrode grid in a non-human primate. Journal of Neural Engineering, 13(4):1–14. (Cited on page 19.)

[Delgado Saa and Cetin, 2012] Delgado Saa, J. F. and Cetin, M. (2012). A latent discriminative model-based approach for classification of imaginary motor tasks from EEG data. Journal of neural engineering, 9(2). (Cited on pages 57 and 58.)

[Demirer et al., 2009] Demirer, R. M., Ozerdem, M. S., and Bayrak, C. (2009). Classification of imaginary movements in ECoG with a hybrid approach based on multi-dimensional Hilbert-SVM solution. Journal of Neuroscience Methods, 178(1):214–218. (Cited on page 53.)

[Dempster et al., 1977] Dempster, A., Laird, N., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society Series B Methodological, 39(1):1–38. (Cited on pages 81 and 82.)

[Demšar, 2006] Demšar, J. (2006). Statistical Comparisons of Classifiers over Multiple Data Sets. The Journal of Machine Learning Research, 7:1–30. (Cited on pages 112 and 128.)

[Devulapalli, 1996] Devulapalli, S. (1996). Non-Linear Principal Component Analysis and Classification of EEG during mental tasks. PhD thesis, Colorado State University. (Cited on page 47.)

[Dietterich, 2009] Dietterich, T. (2009). Machine learning for sequential data: A review. Structural, Syntactic, and Statistical Pattern Recognition, 2396:227–246. (Cited on pages 56, 57, 59, 74 and 81.)

[Dietterich and Bakiri, 1995] Dietterich, T. G. and Bakiri, G. (1995). Solving Multiclass Learning Problems via Error-Correcting Output Codes. Journal of Artificial Intelligence Research, 2:263–286. (Cited on page 194.)

[Do et al., 2013] Do, A. H., Wang, P. T., King, C. E., Chun, S. N., and Nenadic, Z. (2013). Brain-Computer Interface Controlled Robotic Gait Orthosis. Journal of NeuroEngineering and Rehabilitation, 10(1):111. (Cited on page 29.)

[Doud et al., 2011] Doud, A. J., Lucas, J. P., Pisansky, M. T., and He, B. (2011). Continuous three-dimensional control of a virtual helicopter using a motor imagery based brain-computer interface. PloS one, 6(10):e26322. (Cited on pages 13, 19, 20, 29, 30 and 40.)

[Duprès et al., 2014] Duprès, A., Cabestaing, F., and Rouillard, J. (2014). SSVEP-based BCIs : study of classifier stability over time and effects of human learning on classification accuracy. AMSE, Journal of the Association for the Advancement of Modelling and Simulation Techniques in Enterprises (Special edition HANDICAP). (Cited on page 22.)

[Eberly, 2014] Eberly, D. (2014). Derivative Approximation by Finite Differences Derivatives of Univariate Functions. Magic Software, Inc, pages 1–7. (Cited on page 116.)

[Eliseyev and Aksenova, 2013] Eliseyev, A. and Aksenova, T. (2013). Recursive N-way partial least squares for brain-computer interface. PloS one, 8(7):e69962. (Cited on pages 34 and 176.)

[Eliseyev and Aksenova, 2014] Eliseyev, A. and Aksenova, T. (2014). Stable and artifact-resistant decoding of 3D hand trajectories from ECoG signals using the generalized additive model. Journal of neural engineering, 11(6):066005. (Cited on pages 34, 40, 43, 44, 47, 59, 60, 65, 78, 109, 116, 128, 175 and 195.)

[Eliseyev and Aksenova, 2016] Eliseyev, A. and Aksenova, T. (2016). Penalized Multi-Way Partial Least Squares for Smooth Trajectory Decoding from Electrocorticographic (ECoG) Recording. Plos One, 11(5):e0154878. (Cited on pages 47, 59, 63, 175 and 195.)

[Eliseyev et al., 2014] Eliseyev, A., Mestais, C., Charvet, G., Sauter, F., Abroug, N., Arizumi, N., Cokgungor, S., Costecalde, T., Foerster, M., Korczowski, L., Morinière, B., Porcherot, J., Pradal, J., Ratel, D., Tarrin, N., Torres, N., Verney, A., Aksenova, T., and Benabid, A.-l. (2014). CLINATEC ® BCI platform based on the ECoG-recording implant WIMAGINE ® and the innovative signal-processing : preclinical results. 36th Annual International Conference of the IEEE Engineering in Medicince and Biology Society, pages 1222–1225. (Cited on pages 31, 34 and 176.)

[Eliseyev et al., 2011] Eliseyev, A., Moro, C., Costecalde, T., Torres, N., Gharbi, S., Mestais, C., Benabid, A. L., and Aksenova, T. (2011). Iterative N-way partial least squares for a binary self-paced brain-computer interface in freely moving animals. Journal of neural engineering, 8(4):046012. (Cited on pages 34, 53 and 195.)

[Eliseyev et al., 2012] Eliseyev, A., Moro, C., Faber, J., Wyss, A., Torres, N., Mestais, C., Benabid, A. L., and Aksenova, T. (2012). L1-penalized N-way PLS for subset of electrodes selection in BCI experiments. Journal of neural engineering, 9(4):045010. (Cited on pages 34, 43, 44, 53, 56, 175 and 195.)

[Emotiv, 2016] Emotiv (2016). EMOTIV Epoc+. https://www.emotiv.com/epoc/. (Cited on page 19.)

[Fan et al., 2011] Fan, J., Fan, Y., and Wu, Y. (2011). High-Dimensional Classification. High-dimensional Data Analysis, pages 3–37. (Cited on page 51.)

[Fatourechi et al., 2007a] Fatourechi, M., Bashashati, A., Ward, R. K., and Birch, G. E. (2007a). EMG and EOG artifacts in brain computer interface systems: A survey. Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology, 118(3):480–94. (Cited on pages 20, 39 and 40.)

[Fatourechi et al., 2007b] Fatourechi, M., Birch, G. E., and Ward, R. K. (2007b). Application of a hybrid wavelet feature selection method in the design of a self-paced brain interface system. Journal of neuroengineering and rehabilitation, 4(Ic):11. (Cited on page 48.)

[Fazli et al., 2011] Fazli, S., Danóczy, M., Schelldorfer, J., and Müller, K.-R. (2011). L1-penalized linear mixed-effects models for high dimensional data with application to BCI. NeuroImage, 56(4):2100–2108. (Cited on page 48.)

[Fazli et al., 2009] Fazli, S., Grozea, C., Danoczy, M., Blankertz, B., Popescu, F., and Muller, K.-R. (2009). Subject independent EEG-based BCI decoding. Advances in Neural Information Processing Systems, 22:1–9. (Cited on page 21.)

[Felton et al., 2007] Felton, E. a., Wilson, J. A., Williams, J. C., and Garell, P. C. (2007). Electrocorticographically controlled brain-computer interfaces using motor and sensory imagery in patients with temporary subdural electrode implants. Report of four cases. Journal of neurosurgery, 106(3):495–500. (Cited on page 44.)

[Fernandes et al., 2003] Fernandes, F. C., van Spaendonck, R. L., and Burrus, C. S. (2003). A new framework for complex wavelet transforms. IEEE Transactions on signal processing, 51(7):1825–1837. (Cited on page 43.)

[Ferreira et al., 2008] Ferreira, A., Celeste, W. C., Cheein, F. a., Bastos-Filho, T. F., Sarcinelli-Filho, M., and Carelli, R. (2008). Human-machine interfaces based on EMG and EEG applied to robotic systems. Journal of neuroengineering and rehabilitation, 5:10. (Cited on page 11.)

[Fifer et al., 2014] Fifer, M. S., Hotson, G., Wester, B. A., McMullen, D. P., Wang, Y., Johannes, M. S., Katyal, K. D., Helder, J. B., Para, M. P., Vogelstein, R. J., Anderson, W. S., Thakor, N. V., and Crone, N. E. (2014). Simultaneous neural control of simple reaching and grasping with the modular prosthetic limb using intracranial EEG. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 22(3):695–705. (Cited on pages 17, 18, 31, 40, 41, 44, 50, 52, 55, 56, 57, 58, 74, 76, 118 and 172.)

[Fink, 2014] Fink, G. (2014). Markov models for pattern recognition. Springer Science & Business Media. (Cited on pages 84 and 185.)

[Fitzsimmons, 2009] Fitzsimmons, N. A. (2009). Extracting kinematic parameters for monkey bipedal walking from cortical neuronal ensemble activity. Frontiers in Integrative Neuroscience, 3:3. (Cited on page 24.)

[Flamary and Rakotomamonjy, 2012] Flamary, R. and Rakotomamonjy, A. (2012). Decoding Finger Movements from ECoG Signals Using Switching Linear Models. Frontiers in neuroscience, 6(March):29. (Cited on pages 31, 42, 56, 66, 70, 71, 73, 74, 77, 78, 115, 116, 159, 165 and 189.)

[Flesher et al., 2016] Flesher, S. N., Collinger, J. L., Foldes, S. T., Weiss, J. M., Downey, J. E., Tyler-Kabara, E. C., Bensmaia, S. J., Schwartz, A. B., Boninger, M. L., and Gaunt, R. A. (2016). Intracortical microstimulation of human somatosensory cortex. Science Translational Medicine, pages 1–11. (Cited on page 5.)

[Flint et al., 2012] Flint, R. D., Lindberg, E. W., Jordan, L. R., Miller, L. E., and Slutzky, M. W. (2012). Accurate decoding of reaching movements from field potentials in the absence of spikes. Journal of Neural Engineering, 9(4):046006. (Cited on page 60.)

[Flint et al., 2013] Flint, R. D., Wright, Z. a., Scheid, M. R., and Slutzky, M. W. (2013). Long term, stable brain machine interface performance using local field potentials and multiunit spikes. Journal of Neural Engineering, 10(5):056005. (Cited on page 42.)

[Flotzinger et al., 1994] Flotzinger, D., Pregenzer, M., and Pfurtscheller, G. (1994). Feature selection with distinction sensitive learning vector quantisation and genetic algorithms. Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94), 6:3448–3451. (Cited on pages 41 and 48.)

[Foodeh et al., 2016] Foodeh, R., Khorasani, A., Shalchyan, V., and Daliri, M. R. (2016). Minimum Noise Estimate filter: a Novel Automated Artifacts Removal method for Field Potentials. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 4320(c):1–1. (Cited on page 40.)

[Fort and Lambert-Lacroix, 2005] Fort, G. and Lambert-Lacroix, S. (2005). Classification using partial least squares with penalized logistic regression. Bioinformatics, 21(7):1104–1111. (Cited on page 175.)

[Friedman et al., 2001] Friedman, J., Hastie, T., and Tibshirani, R. (2001). The Elements of Statistical Learning. (Cited on pages 49, 50, 53, 59 and 194.)

[Friedman et al., 2010] Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of statistical software, 33(1). (Cited on pages 123, 189 and 194.)

[Friedman, 1989] Friedman, J. H. (1989). Regularized Discriminant Analysis. Journal of the American Statistical Association, 84(405):165–175. (Cited on page 196.)

[Fruitet et al., 2010] Fruitet, J., McFarland, D. J., and Wolpaw, J. R. (2010). A comparison of regression techniques for a two-dimensional sensorimotor rhythm-based brain-computer interface. Journal of neural engineering, 7:16003. (Cited on page 19.)

[Galán et al., 2008] Galán, F., Nuttin, M., Lew, E., Ferrez, P. W., Vanacker, G., Philips, J., and Millán, J. D. R. (2008). A brain-actuated wheelchair: Asynchronous and non-invasive Brain-computer interfaces for continuous control of robots. Clinical Neurophysiology, 119(9):2159–2169. (Cited on page 41.)

[Gancet et al., 2012] Gancet, J., Ilzkovitz, M., Motard, E., Nevatia, Y., Letier, P., Weerdt, D. D., Hoellinger, T., Seetharaman, K., Petieau, M., Ivanenko, Y., Molinari, M., Pisotta, I., Tamburella, F., Labini, F. S., Avella, A., Kooij, H. V. D., Wang, L., Helm, F. V. D., Wang, S., Zanow, F., Hauffe, R., and Thorsteinsson, F. (2012). MINDWALKER: Going One Step Further with Assistive Lower Limbs Exoskeleton for SCI Condition Subjects. IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics, 2010:1794–1800. (Cited on page 29.)

[Ganguly and Carmena, 2009] Ganguly, K. and Carmena, J. M. (2009). Emergence of a stable cortical map for neuroprosthetic control. PLoS biology, 7(7):e1000153. (Cited on pages 24 and 25.)

[Gao et al., 2003] Gao, Y., Black, M. J., Bienenstock, E., and Donoghue, J. P. (2003). A quantitative comparison of linear and non-linear models of motor cortical activity for the encoding and decoding of arm motions. First International IEEE EMBS Conference on Neural Engineering, 2003. Conference Proceedings., pages 189–192. (Cited on pages 61, 62 and 70.)

[Garrett et al., 2003] Garrett, D., Peterson, D. a., Anderson, C. W., and Thaut, M. H. (2003). Comparison of linear, nonlinear, and feature selection methods for EEG signal classification. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 11(2):141–144. (Cited on pages 48, 52, 54 and 55.)

[Geladi and Kowalski, 1986] Geladi, P. and Kowalski, B. R. (1986). Partial least-squares regression: a tutorial. Analytica Chimica Acta, 185(C):1–17. (Cited on page 189.)

[Georgopoulos et al., 1982] Georgopoulos, A. P., Kalaska, J. F., Caminiti, R., and Massey5, J. T. (1982). On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. The Journal of Neuroscience, 2(11):1527–1537. (Cited on pages 23 and 24.)

[Georgopoulos et al., 1986] Georgopoulos, a. P., Schwartz, a. B., Kettner, R. E., and Schwartz, a. B. (1986). Neuronal population coding of movement direction. Science, 233(March):1416–1419. (Cited on page 59.)

[Ghahramani, 2001] Ghahramani, Z. (2001). An Introduction to Hidden Markov Models and Bayesian Networks. International Journal of Pattern Recognition and Artificial Intelligence, 15(01):9–42. (Cited on pages 77 and 79.)

[Goldfeld and Quandt, 1973] Goldfeld, S. M. and Quandt, R. E. (1973). A Markov model for switching regressions. Journal of Econometrics, 1(1):3–15. (Cited on page 70.)

[Golub et al., 2014] Golub, M. D., Yu, B. M., Schwartz, A. B., and Chase, S. M. (2014). Motor cortical control of movement speed with implications for brain-machine interface control. Journal of neurophysiology, 112(2):411–29. (Cited on page 177.)

[Gouy-Pailler et al., 2009] Gouy-Pailler, C., Mattout, J., Congedo, M., and Jutten, C. (2009). Uncued Brain-Computer Interfaces: a Variational Hidden Markov Model of Mental State Dynamics. ESANN 09, Bruges : Belgium (2009). (Cited on pages 53, 56, 108 and 194.)

[Gowda et al., 2014] Gowda, S., Orsborn, A. L., Overduin, S. a., Moorman, H. G., and Carmena, J. M. (2014). Designing dynamical properties of brain-machine

interfaces to optimize task-specific performance. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 22(5):911–920. (Cited on pages 15 and 177.)

[Graimann et al., 2009] 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, The Frontiers Collection, pages 1–28. Springer Berlin Heidelberg, Berlin, Heidelberg. (Cited on pages 1, 8 and 167.)

[Graimann et al., 2004] Graimann, B., Huggins, J. E., Levine, S. P., and Pfurtscheller, G. (2004). Toward a direct brain interface based on human subdural recordings and wavelet-packet analysis. IEEE transactions on bio-medical engineering, 51(6):954–62. (Cited on page 48.)

[Gray, 1918] Gray, H. (1918). Anatomy of the Human Body. Lea \& Febiger. (Cited on page 10.)

[G.tex, 2016] G.tex (2016). intendiX - User-Ready Brain-Computer Interface Applications. http://www.gtec.at/Products/Complete-Solutions/intendiX-Specs-Features. (Cited on page 30.)

[Gunduz, 2008] Gunduz, A. (2008). Human motor control through electrocorticographic brain machine interfaces. Doctoral dissertation, University of Florida. (Cited on pages 23, 24 and 27.)

[Gürel and Mehring, 2012] Gürel, T. and Mehring, C. (2012). Unsupervised adaptation of brain-machine interface decoders. Frontiers in neuroscience, 6(November):164. (Cited on page 177.)

[Gwin and Ferris, 2011] Gwin, J. T. and Ferris, D. (2011). High-density EEG and independent component analysis mixture models distinguish knee contractions from ankle contractions. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, pages 4195–4198. (Cited on page 19.)

[Gysels and Celka, 2004] Gysels, E. and Celka, P. (2004). Phase synchronization for the recognition of mental tasks in a brain-computer interface. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 12(4):406–415. (Cited on page 45.)

[Hammer et al., 2013] Hammer, J., Fischer, J., Ruescher, J., Schulze-Bonhage, A., Aertsen, A., and Ball, T. (2013). The role of ECoG magnitude and phase in decoding position, velocity, and acceleration during continuous motor behavior. Frontiers in neuroscience, 7(November):200. (Cited on pages 27, 28, 41, 44, 45, 46, 63 and 162.)

[Hammer et al., 2016] Hammer, J., Pistohl, T., Fischer, J., Kršek, P., Tomášek, M., Marusič, P., Schulze-Bonhage, A., Aertsen, A., and Ball, T. (2016). Predominance of Movement Speed Over Direction in Neuronal Population Signals of Motor Cortex: Intracranial EEG Data and A Simple Explanatory Model. Cerebral Cortex, page bhw033. (Cited on pages 27, 28, 42, 46, 61, 78 and 115.)

[Hamner et al., 2011] Hamner, B., Leeb, R., Tavella, M., and Del R. Millán, J. (2011). Phase-based features for motor imagery brain-computer interfaces. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, pages 2578–2581. (Cited on pages 44 and 45.)

[Hasan and Gan, 2009] Hasan, B. a. S. and Gan, J. Q. (2009). Unsupervised adaptive GMM for BCI. 2009 4th International IEEE/EMBS Conference on Neural Engineering, NER '09, (June):295–298. (Cited on pages 50 and 51.)

[Hasan and Gan, 2010] Hasan, B. A. S. and Gan, J. Q. (2010). Unsupervised movement onset detection from EEG recorded during self-paced real hand movement. Medical and Biological Engineering and Computing, 48(3):245–253. (Cited on page 57.)

[Hasan and Gan, 2011] Hasan, B. A. S. and Gan, J. Q. (2011). Temporal modeling of EEG during self-paced hand movement and its application in onset detection. Journal of Neural Engineering, 8(5):056015. (Cited on pages 51, 57 and 75.)

[Haselsteiner and Pfurtscheller, 2000] Haselsteiner, E. and Pfurtscheller, G. (2000). Using time-dependent neural networks for EEG classification. IEEE transactions on rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society, 8(4):457–463. (Cited on pages 52, 54 and 58.)

[Hatsopoulos et al., 2004] Hatsopoulos, N., Joshi, J., and O'Leary, J. G. (2004). Decoding continuous and discrete motor behaviors using motor and premotor cortical ensembles. Journal of neurophysiology, 92(2):1165–1174. (Cited on pages 50, 51, 52, 54 and 60.)

[Héliot et al., 2010] Héliot, R., Ganguly, K., Jimenez, J., and Carmena, J. M. (2010). Learning in closed-loop brainmachine interfaces: Modeling and experimental validation. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 40(5):1387–1397. (Cited on page 21.)

[Herman et al., 2008] Herman, P., Prasad, G., McGinnity, T. M., and Coyle, D. (2008). Comparative analysis of spectral approaches to feature extraction for EEG-based motor imagery classification. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 16(4):317–326. (Cited on pages 39, 41 and 44.)

[Hill et al., 2006] Hill, N. J., Lal, T. N., Schröder, M., Hinterberger, T., Wilhelm, B., Nijboer, F., Mochty, U., Widman, G., Elger, C., Schölkopf, B., Kübler, A., and Birbaumer, N. (2006). Classifying EEG and ECoG signals without subject training for

fast BCI implementation: comparison of nonparalyzed and completely paralyzed subjects. IEEE transactions on neural systems and rehabilitation engineering, 14(2):183–186. (Cited on pages 27, 44, 51 and 53.)

[Hochberg et al., 2012] 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. (Cited on pages 5, 12, 15, 16, 23, 24, 25, 29, 30, 31, 35, 49, 61, 62, 64, 172, 173, 176 and 209.)

[Hochberg et al., 2006] Hochberg, L. R., Serruya, M. D., Friehs, G. M., Mukand, J. a., Saleh, M., Caplan, A. H., Branner, A., Chen, D., Penn, R. D., and Donoghue, J. P. (2006). Neuronal ensemble control of prosthetic devices by a human with tetraplegia. Nature, 442(7099):164–71. (Cited on pages 15, 16, 23, 26, 35, 59 and 64.)

[Homer and Nurmikko, 2013] Homer, M. and Nurmikko, A. (2013). Sensors and decoding for intracortical brain computer interfaces. Annual review of biomedical engineering, 15:383–405. (Cited on pages 3, 4, 9, 15, 16 and 22.)

[Hong et al., 2015] Hong, K. S., Naseer, N., and Kim, Y. H. (2015). Classification of prefrontal and motor cortex signals for three-class fNIRS-BCI. Neuroscience Letters, 587:87–92. (Cited on page 4.)

[Horki et al., 2011] Horki, P., Solis-Escalante, T., Neuper, C., and Müller-Putz, G. (2011). Combined motor imagery and SSVEP based BCI control of a 2 DoF artificial upper limb. Medical and Biological Engineering and Computing, 49(5):567–577. (Cited on page 14.)

[Hortal et al., 2015] Hortal, E., Planelles, D., Costa, A., Iáñez, E., Úbeda, A., Azorín, J. M., and Fernández, E. (2015). SVM-based Brain-Machine Interface for controlling a robot arm through four mental tasks. Neurocomputing, 151(P1):116–121. (Cited on pages 20, 30, 42, 46, 50, 51, 52, 55 and 118.)

[Höskuldsson, 1988] Höskuldsson, A. (1988). PLS Regression Methods. Journal of Chemometrics, 2(July):211–228. (Cited on pages 47, 118 and 188.)

[Hotson et al., 2014] Hotson, G., Fifer, M. S., Acharya, S., Benz, H. L., Anderson, W. S., Thakor, N. V., and Crone, N. E. (2014). Coarse electrocorticographic decoding of ipsilateral reach in patients with brain lesions. PLoS ONE, 9(12):1–20. (Cited on pages 28, 41, 42, 46, 48, 63 and 163.)

[Hotson et al., 2016] Hotson, G., McMullen, D. P., Fifer, M. S., Johannes, M. S., Katyal, K. D., Para, M. P., Armiger, R., Anderson, W. S., Thakor, N. V., Wester, B. A., and Crone, N. E. (2016). Individual finger control of a modular prosthetic limb using high-density electrocorticography in a human subject. Journal of Neural Engineering, 13(2):026017. (Cited on pages 12, 17, 18, 31, 46, 49, 50, 52, 55, 56, 57, 58, 74, 75, 76, 118, 165, 172 and 175.)

[Hsu and Lin, 2002] Hsu, C. W. and Lin, C. J. (2002). A comparison of methods for multiclass support vector machines. IEEE Transactions on Neural Networks, 13(2):415–425. (Cited on page 194.)

[Hudson and Burdick, 2007] Hudson, N. and Burdick, J. W. (2007). Learning hybrid system models for supervisory decoding of discrete state, with applications to the parietal reach region. Proceedings of the 3rd International IEEE EMBS Conference on Neural Engineering, pages 587–592. (Cited on page 21.)

[Hyndman and Koehler, 2006] Hyndman, R. J. and Koehler, A. B. (2006). Another look at measures of forecast accuracy. International Journal of Forecasting, 22(4):679–688. (Cited on page 108.)

[Ifft et al., 2013] Ifft, P. J., Shokur, S., Li, Z., Lebedev, M. A., and Nicolelis, M. A. L. (2013). A brain-machine interface enables bimanual arm movements in monkeys. Science translational medicine, 5(210):210ra154. (Cited on pages 12, 15, 16, 21, 24, 30, 31, 35, 61, 62, 67, 70, 159 and 177.)

[Imamizu et al., 2007] Imamizu, H., Sugimoto, N., Osu, R., Tsutsui, K., Sugiyama, K., Wada, Y., and Kawato, M. (2007). Explicit contextual information selectively contributes to predictive switching of internal models. Experimental Brain Research, 181(3):395–408. (Cited on page 69.)

[Iriarte et al., 2003] Iriarte, J., Urrestarazu, E., Valencia, M., Alegre, M., Malanda, A., Viteri, C., and Artieda, J. (2003). Independent component analysis as a tool to eliminate artifacts in EEG: a quantitative study. Journal of clinical neurophysiology : official publication of the American Electroencephalographic Society, 20(4):249–257. (Cited on page 40.)

[Jackson and Fetz, 2011] Jackson, A. and Fetz, E. E. (2011). Interfacing with the computational brain. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 19(5):534–541. (Cited on pages 22, 25 and 177.)

[Jacobs et al., 1991] Jacobs, R. a., Jordan, M. I., Nowlan, S. J., and Hinton, G. E. (1991). Adaptive Mixtures of Local Experts. Neural Computation, 3(1):79–87. (Cited on pages 72, 75 and 81.)

[Jarosiewicz et al., 2013] Jarosiewicz, B., Masse, N. Y., Bacher, D., Cash, S. S., Eskandar, E., Friehs, G., Donoghue, J. P., and Hochberg, L. R. (2013). Advantages of closed-loop calibration in intracortical brain–computer interfaces for people with tetraplegia. Journal of Neural Engineering, 10(4):046012. (Cited on pages 15, 16, 21, 22 and 177.)

[Jerbi et al., 2011] Jerbi, K., Vidal, J., Mattout, J., Maby, E., Lecaignard, F., Ossandon, T., Hamamé, C., Dalal, S., Bouet, R., Lachaux, J.-P., Leahy, R., Baillet, S., Garnero, L., Delpuech, C., and Bertrand, O. (2011). Inferring hand movement kinematics from MEG, EEG and intracranial EEG: From brain-machine interfaces to motor rehabilitation. IRBM, 32(1):8–18. (Cited on pages 24 and 27.)

[Kandel et al., 2000] Kandel, E., Schwartz, J., and Jessell, T. (2000). Principles Of Neural Science. (Cited on pages 9, 10, 11 and 15.)

[Kang et al., 2012] Kang, X., Schieber, M. H., and Thakor, N. V. (2012). Decoding of finger, hand and arm kinematics using switching linear dynamical systems with pre-motor cortical ensembles. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, pages 1732–1735. (Cited on pages 159 and 177.)

[Kao et al., 2017] Kao, J. C., Nuyujukian, P., Ryu, S. I., and Shenoy, K. V. (2017). A high-performance neural prosthesis incorporating discrete state selection with hidden Markov models. IEEE Transactions on Biomedical Engineering, 64(4):935–945. (Cited on pages 47, 57, 58, 159, 172 and 177.)

[Kao et al., 2013] Kao, J. C., Nuyujukian, P., Stavisky, S., Ryu, S. I., Ganguli, S., and Shenoy, K. V. (2013). Investigating the role of firing-rate normalization and dimensionality reduction in brain-machine interface robustness. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, volume 2010, pages 293–298. (Cited on pages 47, 72, 207 and 209.)

[Kapeller et al., 2015] Kapeller, C., Gergondet, P., Kamada, K., Ogawa, H., Takeuchi, F., Ortner, R., Prückl, R., Kheddar, A., Scharinger, J., and Guger, C. (2015). Online Control of a Humanoid Robot through Hand Movement Imagination using CSP and ECoG based Features. In Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE, pages 1765–1768. (Cited on pages 12, 17, 18, 21, 31, 41, 45, 50, 52, 55 and 118.)

[Karin and Andres, 2014] Karin, C. A. and Andres, V. A. (2014). Virtual hand prosthesis moved by encephalographic signals. 2014 III International Congress of Engineering Mechatronics and Automation (CIIMA), pages 1–5. (Cited on page 40.)

[Kayikcioglu and Aydemir, 2010] Kayikcioglu, T. and Aydemir, O. (2010). A polynomial fitting and k-NN based approach for improving classification of motor imagery BCI data. Pattern Recognition Letters, 31(11):1207–1215. (Cited on pages 52, 54 and 55.)

[Ke and Li, 2009] Ke, L. and Li, R. (2009). Classification of EEG signals by multiscale filtering and PCA. Proceedings - 2009 IEEE International Conference on Intelligent Computing and Intelligent Systems, ICIS 2009, 1:362–366. (Cited on pages 47 and 195.)

[Kellis et al., 2012] Kellis, S., Hanrahan, S., Davis, T., House, P. A., Brown, R., and Greger, B. (2012). Decoding hand trajectories from micro-electrocorticography in human patients. In Proceedings of the Annual International Conference of

the IEEE Engineering in Medicine and Biology Society, EMBS, pages 4091–4094. (Cited on pages 10, 11, 12, 13, 17, 18, 30, 42, 62 and 63.)

[Kellis et al., 2015] Kellis, S., Sorensen, L., Darvas, F., Sayres, C., O'Neill, K., Brown, R., House, P., Ojemann, J., and Greger, B. (2015). Multi-scale analysis of neural activity in humans: implications for micro-scale electrocorticography. Clinical Neurophysiology, pages 1–11. (Cited on page 18.)

[Kelly et al., 2012] Kelly, J. W., Degenhart, A. D., Siewiorek, D. P., Smailagic, A., and Wang, W. (2012). Sparse linear regression with elastic net regularization for brain-computer interfaces. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, pages 4275–4278. (Cited on page 48.)

[Kelly et al., 2011] Kelly, J. W., Siewiorek, D. P., Smailagic, A., Collinger, J. L., Weber, D. J., and Wang, W. (2011). Fully automated reduction of ocular artifacts in high-dimensional neural data. IEEE Transactions on Biomedical Engineering, 58(3 PART 1):598–606. (Cited on page 40.)

[Kemere et al., 2008] Kemere, C., Santhanam, G., Yu, B. M., Afshar, A., Ryu, S. I., Meng, T. H., and Shenoy, K. V. (2008). Detecting neural-state transitions using hidden Markov models for motor cortical prostheses. Journal of Neurophysiology, 100(4):2441–2452. (Cited on pages 57 and 75.)

[Kemere et al., 2004a] Kemere, C., Santhanam, G., Yu, B. M., Ryu, S., Meng, T., and Shenoy, K. V. (2004a). Model-based decoding of reaching movements for prosthetic systems. Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference, 6:4524–4528. (Cited on pages 67 and 71.)

[Kemere et al., 2004b] Kemere, C., Shenoy, K. V., and Meng, T. H. (2004b). Model-based neural decoding of reaching movements: A maximum likelihood approach. IEEE Transactions on Biomedical Engineering, 51(6):925–932. (Cited on pages 67 and 71.)

[Kemsley, 1996] Kemsley, E. (1996). Discriminant analysis of high-dimensional data: a comparison of principal components analysis and partial least squares data reduction methods. Chemometrics and Intelligent Laboratory Systems, 33(1):47–61. (Cited on pages 195 and 199.)

[Khreich et al., 2012] Khreich, W., Granger, E., Miri, A., and Sabourin, R. (2012). A survey of techniques for incremental learning of HMM parameters. Information Sciences, 197:105–130. (Cited on page 176.)

[Kim, 1994] Kim, C.-J. (1994). Dynamic linear models with Markov-switching. Journal of Econometrics, 60:1–22. (Cited on page 70.)

[Kim et al., 2006a] Kim, H. K., Biggs, S. J., Schloerb, D. W., Carmena, J. M., Lebedev, M. a., Nicolelis, M. a. L., and Srinivasan, M. a. (2006a). Continuous shared control for stabilizing reaching and grasping with brain-machine interfaces. IEEE transactions on bio-medical engineering, 53(6):1164–73. (Cited on page 8.)

[Kim et al., 2006b] Kim, K. H., Kim, S. S., and Kim, S. J. (2006b). Superiority of nonlinear mapping in decoding multiple single-unit neuronal spike trains: A simulation study. Journal of Neuroscience Methods, 150(2):202–211. (Cited on pages 60 and 175.)

[Kim et al., 2003] Kim, S.-P., Sanchez, J. C., Erdogmus, D., Rao, Y. N., Wessberg, J., Principe, J. C., and Nicolelis, M. (2003). Divide-and-conquer approach for brain machine interfaces: nonlinear mixture of competitive linear models. Neural networks : the official journal of the International Neural Network Society, 16(5-6):865–71. (Cited on pages 67, 70, 71, 72, 73 and 159.)

[Kim et al., 2006c] Kim, S.-P., Sanchez, J. C., Rao, Y. N., Erdogmus, D., Carmena, J. M., Lebedev, M. a., Nicolelis, M. a. L., and Principe, J. C. (2006c). A comparison of optimal MIMO linear and nonlinear models for brain-machine interfaces. Journal of neural engineering, 3:145–161. (Cited on pages 47, 60, 61, 63, 66, 70, 159 and 163.)

[Kim et al., 2008] Kim, S.-P., Simeral, J. D., Hochberg, L. R., Donoghue, J. P., and Black, M. J. (2008). Neural control of computer cursor velocity by decoding motor cortical spiking activity in humans with tetraplegia. Journal of Neural Engineering, 5(4):455–76. (Cited on pages 15, 16 and 30.)

[Kim et al., 2011] Kim, S.-P., Simeral, J. D., Hochberg, L. R., Donoghue, J. P., Friehs, G. M., and Black, M. J. (2011). Point-and-click cursor control with an intracortical neural interface system by humans with tetraplegia. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 19(2):193–203. (Cited on pages 15, 16, 30, 51 and 56.)

[Kindermans et al., 2012] Kindermans, P.-J., Verschore, H., Verstraeten, D., and Schrauwen, B. (2012). A P300 BCI for the Masses: Prior Information Enables Instant Unsupervised Spelling. Advances in Neural Information Processing Systems 25, pages 719–727. (Cited on page 2.)

[King et al., 2015] King, C. E., Wang, P. T., McCrimmon, C. M., Chou, C. C., Do, A. H., and Nenadic, Z. (2015). The feasibility of a brain-computer interface functional electrical stimulation system for the restoration of overground walking after paraplegia. Journal of NeuroEngineering and Rehabilitation, 12(1):80. (Cited on pages 19, 29, 46, 63, 74 and 75.)

[King-Stephens et al., 2015] King-Stephens, D., Mirro, E., Weber, P. B., Laxer, K. D., Van Ness, P. C., Salanova, V., Spencer, D. C., Heck, C. N., Goldman, A., Jobst, B., Shields, D. C., Bergey, G. K., Eisenschenk, S., Worrell, G. A., Rossi,

M. A., Gross, R. E., Cole, A. J., Sperling, M. R., Nair, D. R., Gwinn, R. P., Park, Y. D., Rutecki, P. A., Fountain, N. B., Wharen, R. E., Hirsch, L. J., Miller, I. O., Barkley, G. L., Edwards, J. C., Geller, E. B., Berg, M. J., Sadler, T. L., Sun, F. T., and Morrell, M. J. (2015). Lateralization of mesial temporal lobe epilepsy with chronic ambulatory electrocorticography. Epilepsia, 56(6):959–967. (Cited on page 19.)

[Koike et al., 2006] Koike, Y., Hirose, H., Sakurai, Y., and Iijima, T. (2006). Prediction of arm trajectory from a small number of neuron activities in the primary motor cortex. Neuroscience Research, 55(2):146–153. (Cited on page 46.)

[Kousarrizi et al., 2009] Kousarrizi, M., Ghanbari, A., Teshnehlab, M., Aliyari, M., and Gharaviri, A. (2009). Feature Extraction and Classification of EEG Signals Using Wavelet Transform, SVM and Artificial Neural Networks for Brain Computer Interfaces. 2009 International Joint Conference on Bioinformatics, Systems Biology and Intelligent Computing, pages 16–19. (Cited on page 43.)

[Koyama et al., 2010a] Koyama, S., Castellanos Pérez-Bolde, L., Shalizi, C. R., and Kass, R. E. (2010a). Approximate Methods for State-Space Models. Journal of the American Statistical Association, 105(489):170–180. (Cited on page 62.)

[Koyama et al., 2010b] Koyama, S., Chase, S. M., Whitford, A. S., Velliste, M., Schwartz, A. B., and Kass, R. E. (2010b). Comparison of brain-computer interface decoding algorithms in open-loop and closed-loop control. Journal of computational neuroscience, 29(1-2):73–87. (Cited on pages 59, 61, 62, 63 and 172.)

[Krämer, 2011] Krämer, N. (2011). The Degrees of Freedom of Partial Least Squares Regression. Journal of the American Statistical Association, 106(494):697–705. (Cited on page 188.)

[Krishnamurthy, 2016] Krishnamurthy, V. (2016). Partially Observed Markov Decision Processes. Cambridge University Press. (Cited on pages 61 and 201.)

[Krusienski et al., 2012] Krusienski, D. J., McFarland, D. J., and Wolpaw, J. R. (2012). Value of amplitude, phase, and coherence features for a sensorimotor rhythm-based brain-computer interface. Brain Research Bulletin, 87(1):130–134. (Cited on pages 44 and 45.)

[Krusienski et al., 2006] Krusienski, D. J., Sellers, E. W., Cabestaing, F., Bayoudh, S., McFarland, D. J., Vaughan, T. M., and Wolpaw, J. R. (2006). A comparison of classification techniques for the P300 Speller. Journal of neural engineering, 3(4):299–305. (Cited on page 2.)

[Kubánek et al., 2009] Kubánek, J., Miller, K. J., Ojemann, J. G., Wolpaw, J. R., and Schalk, G. (2009). Decoding flexion of individual fingers using electrocortico-

graphic signals in humans. Journal of Neural Engineering, 6(6):066001. (Cited on pages 31, 39, 59 and 78.)

[Kuratowski, 2014] Kuratowski, K. (2014). Topology. Elsevier. (Cited on page 110.)

[Kwak et al., 2015] Kwak, N.-S., Müller, K.-R., and Lee, S.-W. (2015). A lower limb exoskeleton control system based on steady state visual evoked potentials. Journal of Neural Engineering, 12(5):056009. (Cited on page 29.)

[Lachaux et al., 2012] Lachaux, J.-P., Axmacher, N., Mormann, F., Halgren, E., and Crone, N. E. (2012). High-frequency neural activity and human cognition: Past, present and possible future of intracranial EEG research. Progress in Neurobiology, 98(3):279–301. (Cited on page 173.)

[Lafferty et al., 2001] Lafferty, J., McCallum, A., and Pereira, F. C. N. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. ICML '01 Proceedings of the Eighteenth International Conference on Machine Learning, 8(June):282–289. (Cited on page 57.)

[LaFleur et al., 2013] LaFleur, K., Cassady, K., Doud, A., Shades, K., Rogin, E., and He, B. (2013). Quadcopter control in three-dimensional space using a noninvasive motor imagery-based brain-computer interface. Journal of neural engineering, 10(4):046003. (Cited on pages 4, 12, 19, 20, 25, 26, 29, 30, 40, 42, 48 and 50.)

[Lal et al., 2005] Lal, T. T. N., Hinterberger, T., Widman, G., Schröder, M., Schroeder, M., Hill, J., Rosenstiel, W., Elger, C., Schölkopf, B., and Birbaumer, N. (2005). Methods towards invasive human brain computer interfaces. Advances in Neural Information Processing System, 17:737–744. (Cited on pages 27, 44, 51 and 53.)

[Le Van Quyen et al., 2001] Le Van Quyen, M., Foucher, J., Lachaux, J., Rodriguez, E., Lutz, a., Martinerie, J., and Varela, F. J. (2001). Comparison of Hilbert transform and wavelet methods for the analysis of neuronal synchrony. Journal of neuroscience methods, 111(2):83–98. (Cited on page 44.)

[Lebedev and Nicolelis, 2006] Lebedev, M. a. and Nicolelis, M. a. L. (2006). Brain-machine interfaces: past, present and future. Trends in neurosciences, 29(9):536–46. (Cited on pages 2, 4, 5, 6, 11, 20 and 30.)

[Leeb et al., 2007] Leeb, R., Friedman, D., Müller-Putz, G. R., Scherer, R., Slater, M., and Pfurtscheller, G. (2007). Self-paced (asynchronous) BCI control of a wheelchair in virtual environments: a case study with a tetraplegic. Computational intelligence and neuroscience, 2007:79642. (Cited on pages 26, 35 and 46.)

[Lemm et al., 2004] Lemm, S., Schafer, C., and Curio, G. (2004). BCI competition–Data set III: Probabilistic modelling of sensorimotor m rhythms for classification of imaginary hand movements. IEEE Trans. Biomed. Eng, 51(6):1077–1080. (Cited on pages 43, 51 and 116.)

[Leuthardt et al., 2011] Leuthardt, E. C., Gaona, C., Sharma, M., Szrama, N., Roland, J., Freudenberg, Z., Solis, J., Breshears, J., and Schalk, G. (2011). Using the electrocorticographic speech network to control a brain-computer interface in humans. Journal of neural engineering, 8(3):036004. (Cited on pages 11, 31 and 44.)

[Leuthardt et al., 2006a] Leuthardt, E. C., Miller, K. J., Schalk, G., Rao, R. P. N., and Ojemann, J. G. (2006a). Electrocorticography-based brain computer interface–the Seattle experience. IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society, 14(2):194–8. (Cited on pages 14, 16, 17, 18, 20, 22, 27, 30 and 177.)

[Leuthardt et al., 2006b] Leuthardt, E. C., Schalk, G., Moran, D., and Ojemann, J. G. (2006b). The emerging world of motor neuroprosthetics: a neurosurgical perspective. Neurosurgery, 59(1):1–14; discussion 1–14. (Cited on pages 1, 6, 8, 12 and 13.)

[Leuthardt et al., 2004] Leuthardt, E. C., Schalk, G., Wolpaw, J. R., Ojemann, J. G., and Moran, D. W. (2004). A brain-computer interface using electrocorticographic signals in humans. Journal of neural engineering, 1(2):63–71. (Cited on pages 17, 18, 27, 39 and 44.)

[Lew, 2012] Lew, E. (2012). Detection of self-paced reaching movement intention from EEG signals. Frontiers in Neuroengineering, 5(July):1–17. (Cited on pages 2 and 41.)

[Li et al., 2002] Li, B., Morris, J., and Martin, E. B. (2002). Model selection for partial least squares regression. Chemometrics and Intelligent Laboratory Systems, 64:79–89. (Cited on pages 118, 119, 189 and 207.)

[Li, 2014] Li, Z. (2014). Decoding methods for neural prostheses: where have we reached? Frontiers in Systems Neuroscience, 8(July):1–6. (Cited on pages 46, 58, 59, 61 and 62.)

[Li et al., 2009] Li, Z., O'Doherty, J. E., Hanson, T. L., Lebedev, M. a., Henriquez, C. S., and Nicolelis, M. a. L. (2009). Unscented Kalman Filter for Brain-Machine Interfaces. PLoS ONE, 4(7):e6243. (Cited on pages 59, 62, 63 and 189.)

[Liang and Bougrain, 2009] Liang, N. and Bougrain, L. (2009). Decoding Finger Flexion using Amplitude Modulation form Band-specific ECoG. In European Symposium on Artificial Neural Networks - ESANN 2009, pages 467–472. (Cited on pages 78, 115 and 116.)

[Liang and Bougrain, 2012] Liang, N. and Bougrain, L. (2012). Decoding finger flexion from band-specific ecog signals in humans. Frontiers in Neuroscience, 6(JUN):1–6. (Cited on pages 31, 48, 59 and 63.)

[Liehr et al., 1999] Liehr, S., Pawelzik, K., Kohlmorgen, J., and Müller, K.-R. (1999). Hidden Markov Mixtures of Experts with an Application to EEG Recordings from Sleep. Theory in Biosciences, 118(3-4):246–260. (Cited on page 70.)

[Lingjaerde and Christophersen, 2000] Lingjaerde, O. C. and Christophersen, N. (2000). Shrinkage Structure of Partial Least Squares. Scandinavian Journal of Statistics, 27(3):459–473. (Cited on page 188.)

[Liu et al., 2004] Liu, H., Hild, K. E., Gao, J. B., Erdogmus, D., Príncipe, J. C., and Chris Sackellares, J. (2004). Evaluation of a BSS algorithm for artifacts rejection in epileptic seizure detection. Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference, 1:91–4. (Cited on page 40.)

[Loboda et al., 2014] Loboda, A., Margineanu, A., and Rotariu, G. (2014). Discrimination of EEG-Based Motor Imagery Tasks by Means of a Simple Phase Information Method. International Journal of Advanced Research in Artificial Intelligence (IJARAI), 3(10), 3(10):11–15. (Cited on page 45.)

[López-Larraz et al., 2016] López-Larraz, E., Trincado-Alonso, F., Rajasekaran, V., Pérez-Nombela, S., Del-Ama, A. J., Aranda, J., Minguez, J., Gil-Agudo, A., and Montesano, L. (2016). Control of an Ambulatory Exoskeleton with a Brain-Machine Interface for Spinal Cord Injury Gait Rehabilitation. Frontiers in neuroscience, 10(August):359. (Cited on pages 29 and 39.)

[Lotte et al., 2007] Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., and Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain-computer interfaces. Journal of Neural Engineering, 4(2):R1–R13. (Cited on pages 53, 56, 58 and 59.)

[Ma et al., 2017] Ma, X., Ma, C., Huang, J., Zhang, P., Xu, J., and He, J. (2017). Decoding Lower Limb Muscle Activity and Kinematics from Cortical Neural Spike Trains during Monkey Performing Stand and Squat Movements. Frontiers in Neuroscience, 11(February):1–16. (Cited on page 24.)

[Mahmoudi and Erfanian, 2002] Mahmoudi, B. and Erfanian, a. (2002). Single-channel EEG-based prosthetic hand grasp control for amputee subjects. Proceedings of the Second Joint 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society] [Engineering in Medicine and Biology, 3:2406–2407. (Cited on page 54.)

[Mak and Wolpaw, 2009] Mak, J. N. and Wolpaw, J. R. (2009). Clinical Applications of Brain – Computer Interfaces : Current State and Future Prospects. IEEE reviews in biomedical engineering, 2:187–199. (Cited on pages 1, 2, 3, 4, 5, 8, 9, 16, 30, 42 and 167.)

[Malik et al., 2015] Malik, W. Q., Hochberg, L. R., Donoghue, J. P., and Brown, E. N. (2015). Modulation depth estimation and variable selection in state-space models for neural interfaces. IEEE Transactions on Biomedical Engineering, 62(2):570–581. (Cited on page 207.)

[Marathe and Taylor, 2011] Marathe, a. R. and Taylor, D. M. (2011). Decoding position, velocity, or goal: does it matter for brain-machine interfaces? Journal of neural engineering, 8(2):025016. (Cited on pages 6, 7, 27, 46, 162, 164, 168 and 172.)

[Marathe and Taylor, 2013] Marathe, A. R. and Taylor, D. M. (2013). Decoding continuous limb movements from high-density epidural electrode arrays using custom spatial filters. Journal of Neural Engineering, 10(3):036015. (Cited on pages 17, 18, 21, 31, 41, 45, 46, 62, 63, 115 and 207.)

[Marathe and Taylor, 2015] Marathe, a. R. and Taylor, D. M. (2015). The impact of command signal power distribution, processing delays, and speed scaling on neurally-controlled devices. Journal of neural engineering, 12(4):046031. (Cited on pages 6, 8, 64, 78 and 109.)

[Marx, 1996] Marx, B. D. (1996). Iteratively reweighted partial least squares estimation for generalized linear regression. Technometics, 38:374–381. (Cited on page 175.)

[Mason et al., 2006] Mason, S., Kronegg, J., Huggins, J., Fatourechi, M., and Schlögl, A. (2006). Evaluating the performance of self-paced brain computer interface technology. neil squire soc., vancouver, bc. Technical report, Canada, Tech. Rep. 19. (Cited on pages 35 and 107.)

[Mason and Birch, 2000] Mason, S. G. and Birch, G. E. (2000). A brain-controlled switch for asynchronous control applications. IEEE transactions on bio-medical engineering, 47(10):1297–307. (Cited on pages 29, 46, 54, 63 and 75.)

[Maynard et al., 1997] Maynard, E. M., Nordhausen, C. T., and Normann, R. A. (1997). The Utah Intracortical Electrode Array: A recording structure for potential brain-computer interfaces. Electroencephalography and Clinical Neurophysiology, 102(3):228–239. (Cited on page 15.)

[McCreadie et al., 2014] McCreadie, K. A., Coyle, D. H., and Prasad, G. (2014). Is sensorimotor BCI performance influenced differently by mono, stereo, or 3-D auditory feedback? IEEE Transactions on Neural Systems and Rehabilitation Engineering, 22(3):431–440. (Cited on page 5.)

[McFarland et al., 2010] McFarland, D. J., Sarnacki, W. A., and Wolpaw, J. R. (2010). Electroencephalographic (EEG) control of three-dimensional movement. Journal of neural engineering, 7(3):036007. (Cited on pages 4, 12, 13, 19, 20, 25, 26, 30, 41 and 48.)

[McFarland and Wolpaw, 2011] McFarland, D. J. and Wolpaw, J. R. (2011). Brain-Computer Interfaces for Communication and Control. Communications of the ACM, 54(5):60–66. (Cited on pages 8, 21 and 22.)

[Mehring et al., 2004] Mehring, C., Nawrot, M. P., de Oliveira, S. C., Vaadia, E., Schulze-Bonhage, A., Aertsen, A., and Ball, T. (2004). Comparing information about arm movement direction in single channels of local and epicortical field potentials from monkey and human motor cortex. Journal of Physiology-Paris, 98(4-6):498–506. (Cited on page 23.)

[Mehring et al., 2003] Mehring, C., Rickert, J., Vaadia, E., de Oliveira, S. C., Aertsen, A., and Rotter, S. (2003). Inference of hand movements from local field potentials in monkey motor cortex. Nature Neuroscience, 6(12):1253–1254. (Cited on pages 23, 60 and 175.)

[Meila and Jordan, 1996] Meila, M. and Jordan, M. I. (1996). Learning Fine Motion by Markov Mixtures of Experts. Advances in Neural Information Processing Systems 8, (1567):1003–1009. (Cited on pages 75 and 76.)

[Mestais et al., 2015] Mestais, C. S., Charvet, G., Sauter-Starace, F., Foerster, M., Ratel, D., and Benabid, A. L. (2015). WIMAGINE: Wireless 64-channel ECoG recording implant for long term clinical applications. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 23(1):10–21. (Cited on pages 33 and 176.)

[Milan and Carmena, 2010] Milan, J. D. R. and Carmena, J. (2010). Invasive or Noninvasive: Understanding Brain-Machine Interface Technology [Conversations in BME]. IEEE Engineering in Medicine and Biology Magazine, 29(February):16–22. (Cited on pages 20 and 26.)

[Milekovic et al., 2012] Milekovic, T., Fischer, J., Pistohl, T., Ruescher, J., Schulze-Bonhage, A., Aertsen, A., Rickert, J., Ball, T., and Mehring, C. (2012). An online brain-machine interface using decoding of movement direction from the human electrocorticogram. Journal of neural engineering, 9(4):046003. (Cited on pages 12, 17, 18, 30, 41 and 42.)

[Millán and Mouriño, 2003] Millán, J. D. R. and Mouriño, J. (2003). Asynchronous BCI and local neural classifiers: An overview of the adaptive brain interface project. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 11(2):159–161. (Cited on pages 29, 63 and 75.)

[Miller et al., 2012] Miller, K. J., Hermes, D., Honey, C. J., Hebb, A. O., Ramsey, N. F., Knight, R. T., Ojemann, J. G., and Fetz, E. E. (2012). Human Motor Cortical Activity Is Selectively Phase-Entrained on Underlying Rhythms. PLoS Computational Biology, 8(9). (Cited on page 95.)

[Miller and Schalk, 2008] Miller, K. J. and Schalk, G. (2008). Prediction of Finger Flexion: 4th Brain-Computer Interface Data Competition. BCI Competition IV. (Cited on pages 95 and 96.)

[Moerland, 1997] Moerland, P. (1997). Some methods for training mixtures of experts. Technical report, IDIAP. (Cited on pages 81 and 123.)

[Moerland, 2000] Moerland, P. (2000). Mixture models for unsupervised and supervised learning. PhD thesis, Ecole Polytechnique Fédérale de Lausanne. (Cited on page 123.)

[Morinière et al., 2015] Morinière, B., Verney, a., Abroug, N., Garrec, P., and Perrot, Y. (2015). EMY: A dual arm exoskeleton dedicated to the evaluation of Brain Machine Interface in clinical trials. IEEE International Conference on Intelligent Robots and Systems, 2015-Decem:5333–5338. (Cited on pages 34 and 176.)

[Morshed and Khan, 2014] Morshed, B. I. and Khan, A. (2014). Biomedical Science A Brief Review of Brain Signal Monitoring Technologies for BCI Applications : Challenges and Prospects. Journal of Bioengineering & Biomedical Science, 4(1):1–10. (Cited on pages 10 and 17.)

[Müller et al., 2003] Müller, K.-R., Anderson, C. W., and Birch, G. E. (2003). Linear and nonlinear methods for brain-computer interfaces. IEEE transactions on neural systems and rehabilitation engineering, 11(2):165–169. (Cited on pages 52 and 55.)

[Muller, 2007] Muller, M. (2007). Dynamic Time Warping. Information retrieval for music and motion, pages 69–84. (Cited on page 109.)

[Müller et al., 2011] Müller, S. M. T., Bastos-Filho, T. F., and Sarcinelli-Filho, M. (2011). Using a SSVEP-BCI to command a robotic wheelchair. In Industrial Electronics (ISIE), 2011 IEEE International Symposium on, number 1-2, pages 957–962. (Cited on page 2.)

[Müller-Putz et al., 2010] Müller-Putz, G. R., Kaiser, V., Solis-Escalante, T., and Pfurtscheller, G. (2010). Fast set-up asynchronous brain-switch based on detection of foot motor imagery in 1-channel EEG. Medical & Biological Engineering & Computing, 48(3):229–233. (Cited on page 46.)

[Murguialday et al., 2007] Murguialday, A. R., Aggarwal, V., Chatterjee, A., Cho, Y., Rasmussen, R., O'Rourke, B., Acharya, S., and Thakor, N. V. (2007). Brain-Computer Interface for a Prosthetic Hand Using Local Machine Control and Haptic Feedback. 2007 IEEE 10th International Conference on Rehabilitation Robotics, 00(c):609–613. (Cited on pages 29, 30 and 175.)

[Murphy, 1998] Murphy, K. (1998). Switching Kalman Filters. Technical Report August. (Cited on pages 104, 202, 203, 204, 205 and 206.)

[Murphy, 2002] Murphy, K. (2002). Dynamic Bayesian Networks: Representation, Inference and Learning. Doctoral dissertation, University of California, Berkeley. (Cited on page 57.)

[Murphy et al., 2015] Murphy, M. D., Guggenmos, D. J., Bundy, D. T., and Nudo, R. J. (2015). Current Challenges Facing the Translation of Brain Computer Interfaces from Preclinical Trials to Use in Human Patients. Frontiers in cellular neuroscience, 9(January):497. (Cited on page 17.)

[Muthukumaraswamy, 2013] Muthukumaraswamy, S. D. (2013). High-frequency brain activity and muscle artifacts in MEG/EEG: a review and recommendations. Frontiers in Human Neuroscience, 7(April):1–11. (Cited on pages 20, 39 and 40.)

[Nakanishi et al., 2014a] Nakanishi, M., Wang, Y.-T. Y., Mitsukura, Y., and Jung, T.-P. (2014a). A high-speed brain speller using steady-state visual evoked potentials. International journal of neural systems, 24(6):1450019. (Cited on page 2.)

[Nakanishi et al., 2014b] Nakanishi, Y., Yanagisawa, T., Shin, D., Chen, C., Kambara, H., Yoshimura, N., Fukuma, R., Kishima, H., Hirata, M., and Koike, Y. (2014b). Decoding fingertip trajectory from electrocorticographic signals in humans. Neuroscience research, 85:20–7. (Cited on pages 31 and 66.)

[Nakanishi et al., 2013] Nakanishi, Y., Yanagisawa, T., Shin, D., Fukuma, R., Chen, C., Kambara, H., Yoshimura, N., Hirata, M., Yoshimine, T., and Koike, Y. (2013). Prediction of three-dimensional arm trajectories based on ECoG signals recorded from human sensorimotor cortex. PloS one, 8(8):e72085. (Cited on pages 63 and 162.)

[Nakayama and Inagaki, 2006] Nakayama, K. and Inagaki, K. (2006). A Brain Computer Interface Based on Neural Network with Efficient Pre-Processing. Simulation, (L):6–9. (Cited on pages 52 and 54.)

[Naseer et al., 2014] Naseer, N., Hong, M. J., and Hong, K. S. (2014). Online binary decision decoding using functional near-infrared spectroscopy for the development of brain-computer interface. Experimental Brain Research, 232(2):555–564. (Cited on page 4.)

[Navarro et al., 2005] Navarro, I., Hubais, B., and Sepulveda, F. (2005). A Comparison of Time, Frequency and ICA Based Features and Five Classifiers for Wrist Movement Classification in EEG Signals. Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference, 2:2118–2121. (Cited on pages 52, 53 and 54.)

[Nazarpour et al., 2009] Nazarpour, K., Praamstra, P., Miall, R. C., and Sanei, S. (2009). Steady-state movement related potentials for brain-computer interfacing. IEEE transactions on bio-medical engineering, 56(8):2104–2113. (Cited on page 13.)

[Ng and Jordan, 2002] Ng, A. Y. and Jordan, M. I. (2002). On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. Advances in neural information processing systems, pages 841–848. (Cited on pages 50 and 51.)

[Nguyen and Rocke, 2002] Nguyen, D. V. and Rocke, D. M. (2002). Multi-class cancer classification via partial least squares with gene expression profiles. Bioinformatics (Oxford, England), 18(9):1216–1226. (Cited on pages 195 and 199.)

[Nguyen et al., 2013] Nguyen, H. T., Trung, N., Toi, V., and Tran, V. S. (2013). An autoregressive neural network for recognition of eye commands in an EEG-controlled wheelchair. International Conference on Advanced Technologies for Communications, pages 333–338. (Cited on page 2.)

[Nicolas-Alonso and Gomez-Gil, 2012] Nicolas-Alonso, L. F. and Gomez-Gil, J. (2012). Brain computer interfaces, a review. Sensors (Basel, Switzerland), 12(2):1211–79. (Cited on pages 4 and 52.)

[Nicolelis, 2014] Nicolelis, M. a. L. (2014). Brain-to-Brain Interfaces: When Reality Meets Science Fiction. Cerebrum : the Dana forum on brain science, (3):13. (Cited on page 29.)

[Nurse et al., 2015] Nurse, E. S., Freestone, D. R., Oxley, T. J., Ackland, D. C., Vogrin, S. J., Murphy, M., Brien, T. J. O., Cook, M. J., and Grayden, D. B. (2015). Movement Related Directional Tuning from Broadband Electrocorticography in Humans. pages 22–24. (Cited on pages 18, 24, 26, 27, 28, 39 and 115.)

[Obermaier et al., 2001] Obermaier, B., Guger, C., Neuper, C., and Pfurtscheller, G. (2001). Hidden Markov models for online classification of single trial EEG data. Pattern Recognition Letters, 22(12):1299–1309. (Cited on pages 26, 56 and 57.)

[O'Doherty et al., 2011] O'Doherty, J. E., Lebedev, M. a., Ifft, P. J., Zhuang, K. Z., Shokur, S., Bleuler, H., and Nicolelis, M. a. L. (2011). Active tactile exploration using a brain–machine–brain interface. Nature, 479(7372):228–231. (Cited on pages 5 and 30.)

[Ofner and Müller-Putz, 2012] Ofner, P. and Müller-Putz, G. R. (2012). Decoding of velocities and positions of 3D arm movement from EEG. Conference proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference, 2012:6406–9. (Cited on page 24.)

[Onaran et al., 2011] Onaran, I., Ince, N. F., Cetin, a. E., and Abosch, A. (2011). A hybrid SVM/HMM based system for the state detection of individual finger movements from multichannel ECoG signals. 2011 5th International IEEE/EMBS Conference on Neural Engineering, NER 2011, pages 457–460. (Cited on pages 41, 56 and 57.)

[Onose et al., 2012] Onose, G., Grozea, C., Anghelescu, A., , and others Daia, C. (2012). On the feasibility of using motor imagery EEG-based brain–computer interface in chronic tetraplegics for assistive robotic arm control: a clinical test and long-term post trial follow-up. Spinal Cord, 50(9):716–716. (Cited on pages 19, 20, 29, 40, 41 and 45.)

[Ordóñez et al., 2013] Ordóñez, F. J., de Toledo, P., and Sanchis, A. (2013). Activity Recognition Using Hybrid Generative/Discriminative Models on Home Environments Using Binary Sensors. Sensors, 13(5):5460–5477. (Cited on pages 77 and 118.)

[Orsborn et al., 2011] Orsborn, A. L., Dangi, S., Moorman, H. G., and Carmena, J. M. (2011). Exploring time-scales of closed-loop decoder adaptation in brain-machine interfaces. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, pages 5436–5439. (Cited on pages 15 and 16.)

[Orsborn et al., 2012] Orsborn, A. L., Dangi, S., Moorman, H. G., and Carmena, J. M. (2012). Closed-loop decoder adaptation on intermediate time-scales facilitates rapid BMI performance improvements independent of decoder initialization conditions. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 20(4):468–477. (Cited on pages 15 and 16.)

[Ortner et al., 2011] Ortner, R., Allison, B. Z., Korisek, G., Gaggl, H., and Pfurtscheller, G. (2011). An SSVEP BCI to control a hand orthosis for persons with tetraplegia. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 19(1):1–5. (Cited on pages 2, 11, 12, 14 and 29.)

[Ozkan et al., 2012] Ozkan, H., Akman, A., and Kozat, S. S. (2012). A novel training algorithm for HMMs with partial and noisy access to the states. arXiv preprint, (March 2012). (Cited on page 176.)

[Palankar et al., 2008] Palankar, M., De Laurentis, K. J., Alqasemi, R., Veras, E., Dubey, R., Arbel, Y., and Donchin, E. (2008). Control of a 9-DoF wheelchair-mounted robotic arm system using a P300 brain computer interface: Initial experiments. 2008 IEEE International Conference on Robotics and Biomimetics, ROBIO 2008, pages 348–353. (Cited on pages 2, 12, 14 and 42.)

[Paralikar, 2010] Paralikar, K. J. (2010). New approaches to eliminating common-noise artifacts in recordings from intracortical microelectrode arrays: inter-electrode correlation and virtual referencing. Journal of Neuroscience, 181(1):27–35. (Cited on page 39.)

[Parsons et al., 2004] Parsons, L., Parsons, L., Haque, E., Haque, E., Liu, H., and Liu, H. (2004). Subspace clustering for high dimensional data. ACM SIGKDD Explorations Newsletter, 6(1):90–105. (Cited on page 121.)

[Penny and Roberts, 1999] Penny, W. and Roberts, S. (1999). Experiments with an eeg-based computer interface. Technical report, Technical report, Department of Electrical and Electronic Engineering, Imperial College, London. (Cited on page 25.)

[Penny et al., 2000] Penny, W. D., Roberts, S. J., Curran, E. A., and Stokes, M. J. (2000). EEG-based communication: a pattern recognition approach. IEEE transactions on Rehabilitation Engineering, 8(2):214–215. (Cited on pages 13, 14 and 53.)

[Pereira et al., 2015] Pereira, D. G., Afonso, A., and Medeiros, F. M. (2015). Overview of Friedman's test and post-hoc analysis. Communications in Statistics-Simulation and Computation, 44(10):2636–2653. (Cited on page 112.)

[Perruchoud et al., 2016] Perruchoud, D., Pisotta, I., Carda, S., Murray, M. M., and Ionta, S. (2016). Biomimetic rehabilitation engineering: the importance of somatosensory feedback for brain-machine interfaces. Journal of neural engineering, 13(4):041001. (Cited on page 5.)

[Pfurtscheller et al., 2000] Pfurtscheller, G., Guger, C., Müller, G., Krausz, G., and Neuper, C. (2000). Brain oscillations control hand orthosis in a tetraplegic. Neuroscience Letters, 292(3):211–214. (Cited on pages 3, 12, 13, 19, 20, 29, 42, 46, 50 and 52.)

[Pfurtscheller et al., 2010] Pfurtscheller, G., Solis-Escalante, T., Ortner, R., Linortner, P., and Muller-Putz, G. R. (2010). Self-paced operation of an SSVEP-based orthosis with and without an imagery-based "brain switch": A feasibility study towards a hybrid BCI. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 18(4):409–414. (Cited on pages 14, 64 and 75.)

[Phatak and De Hoog, 2002] Phatak, A. and De Hoog, F. (2002). Exploiting the connection between PLS, Lanczos methods and conjugate gradients: Alternative proofs of some properties of PLS. Journal of Chemometrics, 16(7):361–367. (Cited on page 192.)

[Pistohl et al., 2008] Pistohl, T., Ball, T., Schulze-Bonhage, A., Aertsen, A., and Mehring, C. (2008). Prediction of arm movement trajectories from ECoG-recordings in humans. Journal of neuroscience methods, 167(1):105–14. (Cited on pages 24, 25, 26, 27, 28, 42, 43, 62, 63, 115, 116, 120, 201, 206, 207 and 209.)

[Polikov et al., 2005] Polikov, V. S., Tresco, P. a., and Reichert, W. M. (2005). Response of brain tissue to chronically implanted neural electrodes. Journal of neuroscience methods, 148(1):1–18. (Cited on page 16.)

[Pourbakhtiar et al., 2013] Pourbakhtiar, A., Shamsi, M., and Farrokhshad, F. (2013). Neuro-fuzzy classification of brain computer interface data using phase based feature. 13th Iranian Conference on Fuzzy Systems, IFSC 2013, pages 3–6. (Cited on page 45.)

[Qian et al., 2013] Qian, J., Hastie, T., Friedman, J., Tibshirani, R., and Simon, N. (2013). Glmnet for Matlab. https://web.stanford.edu/~hastie/glmnet_matlab/. (Cited on pages 123 and 189.)

[Quandt, 1958] Quandt, R. E. (1958). The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes. Journal of the American Statistical Association, 53(284):873–880. (Cited on page 70.)

[Quandt and Ramsey, 1978] Quandt, R. E. and Ramsey, J. B. (1978). Estimating mixtures of normal distributions and switching regressions. Journal of the American Statistical Association, 73(364):730. (Cited on page 70.)

[Quinn, 2005] Quinn, J. (2005). Switching Kalman Filter Matlab implementation. http://air.ug/~jquinn/downloads/. (Cited on page 204.)

[Rabiner, 1989] Rabiner, L. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE, 77(2):257–286. (Cited on pages 56, 57, 77, 82, 84, 86, 176 and 184.)

[Rabiner and Juang, 1986] Rabiner, L. and Juang, B. (1986). An introduction to hidden Markov models. IEEE ASSP Magazine, 3(January):Appendix 3A. (Cited on page 76.)

[Rak et al., 2012] Rak, R. J., Kołodziej, M., and Majkowski, A. (2012). Brain-computer interface as measurement and control system - the review paper. Metrology and Measurement Systems, XIX(3):427–444. (Cited on page 4.)

[Ramsey et al., 2006] Ramsey, N. F., Heuvel, M. P. V. D., Kho, K. H., and Leijten, F. S. S. (2006). Towards Human BCI Applications Based on Cognitive Brain Systems: An Investi- gation of Neural Signals Recorded From the Dorsolateral Prefrontal Cortex. IEEE Transactions on Neural Systems and Rehabilitation, 14(2):214–217. (Cited on pages 13 and 14.)

[Rebsamen et al., 2006] Rebsamen, B., Burdet, E., Guan, C., Zhang, H., Teo, C. L., Zeng, Q., Ang, M., and Laugier, C. (2006). A brain-controlled wheelchair based on P300 and path guidance. Proceedings of the First IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics, 2006, BioRob 2006, 2006:1101–1106. (Cited on page 2.)

[Renals et al., 1994] Renals, S., Morgan, N., Bourlard, H., Cohen, M., and Franco, H. (1994). Connectionist Probability Estimation in HMM Speech Recognition. IEEE Transactions on Speech and Audio Processing, 2(1):161–174. (Cited on pages 77 and 118.)

[Robotnik, 2016] Robotnik (2016). KINOVA JACO arm. http://www.robotnik.eu/robotics-arms/kinova-jaco-arm/. (Cited on page 29.)

[Roche, 2011] Roche, A. (2011). EM algorithm and variants: an informal tutorial. arXiv preprint arXiv:1105.1476, pages 1–17. (Cited on page 121.)

[Rosipal and Krämer, 2006] Rosipal, R. and Krämer, N. (2006). Overview and Recent Advances in Partial Least Squares. Subspace, Latent Structure and Feature Selection, Saunders, C., et al. (eds.) (Heidelberg: Springer-Verlag, 2006), 3940:34–51. (Cited on pages 195 and 207.)

[Rouse and Moran, 2009] Rouse, A. G. and Moran, D. W. (2009). Neural adaptation of epidural electrocorticographic (EECoG) signals during closed-loop brain computer interface (BCI) tasks. Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference, 2009:5514–7. (Cited on pages 17, 18, 27 and 176.)

[Rouse and Schieber, 2015] Rouse, A. G. and Schieber, M. H. (2015). Advancing brain-machine interfaces: moving beyond linear state space models. Frontiers in Systems Neuroscience, 9(July):1–13. (Cited on pages 69 and 159.)

[Rouse et al., 2013] Rouse, A. G., Williams, J. J., Wheeler, J. J., and Moran, D. W. (2013). Cortical Adaptation to a Chronic Micro- Electrocorticographic Brain Computer Interface. Journal of Neuroscience, 33(4):1326–1330. (Cited on pages 17 and 18.)

[Rouse et al., 2016] Rouse, A. G., Williams, J. J., Wheeler, J. J., and Moran, D. W. (2016). Spatial co-adaptation of cortical control columns in a micro-ECoG brain–computer interface. Journal of Neural Engineering, 13(5):056018. (Cited on pages 17 and 18.)

[Royer et al., 2010] Royer, A. S., Doud, A. J., Rose, M. L., and He, B. (2010). EEG control of a virtual helicopter in 3-dimensional space using intelligent control strategies. IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society, 18(6):581–589. (Cited on pages 19 and 20.)

[Ryali et al., 2010] Ryali, S., Supekar, K., Abrams, D. A., and Menon, V. (2010). Sparse logistic regression foro whole brain classification of fMRI data. NeuroImage, 51(2):752–764. (Cited on page 194.)

[Saa and Çetin, 2013] Saa, J. F. D. and Çetin, M. (2013). Discriminative Methods for Classi fi cation of Asynchronous Imaginary Motor Tasks From EEG Data. Ieee Transactions on Neural Systems and Rehabilitation Engineering, 21(5):716–724. (Cited on pages 57 and 58.)

[Saa et al., 2016] Saa, J. F. D., de Pesters, A., and Cetin, M. (2016). Asynchronous decoding of finger movements from ECoG signals using long-range dependencies conditional random fields. Journal of Neural Engineering, 13(3):36017. (Cited on pages 31, 56, 57, 58, 74, 75 and 165.)

[Sadeghian and Moradi, 2007] Sadeghian, E. B. and Moradi, M. H. (2007). Continuous detection of motor imagery in a four-class asynchronous BCI. Annual

International Conference of the IEEE Engineering in Medicine and Biology - Proceedings, pages 3241–3244. (Cited on pages 39, 41, 52 and 53.)

[Sadtler, 2014] Sadtler, P. T. (2014). Investigating the Neural Basis of Learning Using Brain-Computer Interfaces. PhD thesis, University of Pittsburgh. (Cited on page 207.)

[Sadtler et al., 2014] Sadtler, P. T., Quick, K. M., Golub, M. D., Chase, S. M., Ryu, S. I., Tyler-Kabara, E. C., Yu, B. M., and Batista, A. P. (2014). Neural constraints on learning. Nature, 512(7515):423–426. (Cited on pages 15, 16, 24 and 25.)

[Sanchez et al., 2002] Sanchez, J. C., Erdogmus, D., Principe, J. C., Wessberg, J., and Nicolelis, M. (2002). A comparison between nonlinear mappings and linear state estimation to model the relation from motor cortical neuronal firing to hand movements. Proceedings of SAB Workshop on Motor Control in Humans and Robots: On the Interplay of Real Brains and Artificial Devices, (August):59–65. (Cited on page 60.)

[Sanchez et al., 2008] Sanchez, J. C., Gunduz, A., Carney, P. R., and Principe, J. C. (2008). Extraction and localization of mesoscopic motor control signals for human ECoG neuroprosthetics. Journal of Neuroscience Methods, 167(1):63–81. (Cited on page 17.)

[Schafer, 2011] Schafer, R. (2011). What Is a Savitzky-Golay Filter?: [Lecture Notes]. IEEE Signal Process. Mag., 28(July):111–117. (Cited on page 116.)

[Schalk et al., 2007] Schalk, G., Kubánek, J., Miller, K. J., Anderson, N. R., Leuthardt, E. C., Ojemann, J. G., Limbrick, D., Moran, D., Gerhardt, L. a., and Wolpaw, J. R. (2007). Decoding two-dimensional movement trajectories using electrocorticographic signals in humans. Journal of Neural Engineering, 4(3):264–275. (Cited on pages 17, 26, 27, 28, 41, 42, 44, 48, 63, 78, 95, 115 and 162.)

[Schalk and Leuthardt, 2011] Schalk, G. and Leuthardt, E. C. (2011). Brain-computer interfaces using electrocorticographic signals. IEEE Reviews in Biomedical Engineering, 4:140–154. (Cited on pages 12, 15, 17, 18 and 20.)

[Schalk et al., 2008] Schalk, G., Miller, K. J., Anderson, N. R., Wilson, J. a., Smyth, M. D., Ojemann, J. G., Moran, D. W., Wolpaw, J. R., and Leuthardt, E. C. (2008). Two-dimensional movement control using electrocorticographic signals in humans. Journal of neural engineering, 5(1):75–84. (Cited on pages 17, 18, 26, 27, 31, 44 and 60.)

[Scheffer et al., 2001] Scheffer, T., Decomain, C., and Wrobel, S. (2001). Active hidden markov models for information extraction. Advances in Intelligent Data Analysis, pages 309–318. (Cited on page 176.)

[Scheid et al., 2013] Scheid, M. R., Flint, R. D., Wright, Z. a., and Slutzky, M. W. (2013). Long-term, stable behavior of local field potentials during brain machine interface use. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, pages 307–310. (Cited on page 60.)

[Scherer et al., 2008] Scherer, R., Lee, F., Schl??gl, A., Leeb, R., Bischof, H., and Pfurtscheller, G. (2008). Toward self-paced brain-computer communication: Navigation through virtual worlds. IEEE Transactions on Biomedical Engineering, 55(2):675–682. (Cited on page 48.)

[Schlögl et al., 2005] Schlögl, A., Lee, F., Bischof, H., and Pfurtscheller, G. (2005). Characterization of Four-Class Motor Imagery EEG Data for the BCI-Competition 2005. Journal of Neural Engineering, 2(4):L14–L22. (Cited on pages 44, 51, 52, 53, 54 and 55.)

[Schwartz et al., 2006] Schwartz, A. B., Cui, X. T., Weber, D. J., and Moran, D. W. (2006). Brain-controlled interfaces: movement restoration with neural prosthetics. Neuron, 52(1):205–20. (Cited on page 3.)

[Schwarz et al., 2014] Schwarz, D. a., Lebedev, M. a., Hanson, T. L., Dimitrov, D. F., Lehew, G., Meloy, J., Rajangam, S., Subramanian, V., Ifft, P. J., Li, Z., Ramakrishnan, A., Tate, A., Zhuang, K. Z., and Nicolelis, M. a. L. (2014). Chronic, wireless recordings of large-scale brain activity in freely moving rhesus monkeys. Nature methods, 11(6):670–6. (Cited on page 15.)

[Scott, 2008] Scott, S. H. (2008). Inconvenient truths about neural processing in primary motor cortex. The Journal of physiology, 586(5):1217–24. (Cited on pages 23 and 59.)

[Sellers et al., 2006] Sellers, E. W., Kübler, A., and Donchin, E. (2006). Brain-computer interface research at the University of South Florida cognitive psychophysiology laboratory: The P300 speller. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 14(2):221–224. (Cited on page 2.)

[Shanechi et al., 2013] Shanechi, M. M., Williams, Z. M., Wornell, G. W., Hu, R. C., Powers, M., and Brown, E. N. (2013). A real-time brain-machine interface combining motor target and trajectory intent using an optimal feedback control design. PloS one, 8(4):e59049. (Cited on pages 59 and 189.)

[Shenoy and Carmena, 2014] Shenoy, K. V. and Carmena, J. M. (2014). Combining decoder design and neural adaptation in brain-machine interfaces. Neuron, 84(4):665–80. (Cited on page 23.)

[Shenoy, 2005] Shenoy, P. (2005). Dynamic Bayesian Networks for Brain-Computer Interfaces. Advances in Neural Information Processing Systems, 17:1265–72. (Cited on page 57.)

[Shenoy et al., 2008] Shenoy, P., Miller, K. J., Ojemann, J. G., and Rao, R. P. N. (2008). Generalized features for electrocorticographic BCIs. IEEE Transactions on Biomedical Engineering, 55(1):273–280. (Cited on page 56.)

[Shimoda et al., 2012] Shimoda, K., Nagasaka, Y., Chao, Z. C., and Fujii, N. (2012). Decoding continuous three-dimensional hand trajectories from epidural electrocorticographic signals in Japanese macaques. Journal of Neural Engineering, 9(3):036015. (Cited on pages 17, 19, 24, 39, 43, 44, 47, 59, 63, 64, 89, 91, 115, 116, 162, 188 and 195.)

[Shin et al., 2012] Shin, D., Watanabe, H., Kambara, H., Nambu, A., Isa, T., Nishimura, Y., and Koike, Y. (2012). Prediction of Muscle Activities from Electrocorticograms in Primary Motor Cortex of Primates. PLoS ONE, 7(10):e47992. (Cited on pages 41, 43 and 46.)

[Simeral et al., 2011] Simeral, J. D., Kim, S.-P., Black, M. J., Donoghue, J. P., and Hochberg, L. R. (2011). Neural control of cursor trajectory and click by a human with tetraplegia 1000 days after implant of an intracortical microelectrode array. Journal of neural engineering, 8(2):025027. (Cited on pages 15, 16, 17, 22 and 30.)

[Sing et al., 2007] Sing, C., Tsui, L., and Gan, J. Q. (2007). Asynchronous BCI Control of a Robot Simulator with Supervised Online Training. Intelligent Data Engineering and Automated Learning - IDEAL 2007, pages 125–134. (Cited on pages 49, 50 and 55.)

[Slutzky et al., 2010] Slutzky, M. W., Jordan, L. R., Krieg, T., Chen, M., Mogul, D. J., and Miller, L. E. (2010). Optimal Spacing of Surface Electrode Arrays for Brain Machine Interface Applications. Journal of neural engineering, 7(2):26004. (Cited on page 18.)

[Souza and Araújo, 2014] Souza, F. a. a. and Araújo, R. (2014). Mixture of partial least squares experts and application in prediction settings with multiple operating modes. Chemometrics and Intelligent Laboratory Systems, 130:192–202. (Cited on page 123.)

[Spüler et al., 2016] Spüler, M., Grimm, F., Gharabaghi, A., Bogdan, M., and Rosenstiel, W. (2016). Comparing Methods for Decoding Movement Trajectory from ECoG in Chronic Stroke Patients. In Advances in Neurotechnology, Electronics and Informatics, pages 125–139. Springer. (Cited on pages 39, 44, 48, 60, 66, 115, 188, 189 and 192.)

[Spuler et al., 2015] Spuler, M., Sarasola-Sanz, A., Birbaumer, N., Rosenstiel, W., and Ramos-Murguialday, A. (2015). Comparing metrics to evaluate performance of regression methods for decoding of neural signals. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, volume 2015-Novem, pages 1083–1086. (Cited on page 108.)

[Squire et al., 2013] Squire, L., Berg, D., Bloom, F. E., Du Lac, S., Ghosh, A., and Spitzer, N. C. (2013). Fundamental Neuroscience. Academic Press. (Cited on pages 9 and 16.)

[Srinivasan et al., 2007] Srinivasan, L., Eden, U. T., Mitter, S. K., and Brown, E. N. (2007). General-purpose filter design for neural prosthetic devices. Journal of Neurophysiology, 98(4):2456–2475. (Cited on pages 29, 62, 65, 70, 71, 74, 76, 104 and 172.)

[Stavisky et al., 2015] Stavisky, S. D., Kao, J. C., Sorokin, J. M., Ryu, S. I., and Shenoy, K. V. (2015). System identification of brain-machine interface control using a cursor jump perturbation. International IEEE/EMBS Conference on Neural Engineering, NER, 2015-July:643–647. (Cited on page 78.)

[Suk and Lee, 2010] Suk, H. I. and Lee, S. W. (2010). Two-layer hidden Markov models for multi-class motor imagery classification. Proceedings - Workshop on Brain Decoding: Pattern Recognition Challenges in Neuroimaging, WBD 2010 - In Conjunction with theInternational Conference on Pattern Recognition, ICPR 2010, pages 5–8. (Cited on pages 47 and 195.)

[Suminski et al., 2010] Suminski, a. J., Tkach, D. C., Fagg, a. H., and Hatsopoulos, N. G. (2010). Incorporating Feedback from Multiple Sensory Modalities Enhances Brain-Machine Interface Control. Journal of Neuroscience, 30(50):16777–16787. (Cited on pages 5, 6, 30, 59 and 189.)

[Sutton and Mccallum, 2012] Sutton, C. and Mccallum, A. (2012). An Introduction to Conditional Random Fields. Foundations and Trends® in Machine Learning in Machine Learning, 4(4):267–373. (Cited on pages 51 and 57.)

[Suway et al., 2013] Suway, S. B., Tien, R. N., Jeffries, S. M., Zohny, Z., Clanton, S. T., McMorland, A. J. C., and Velliste, M. (2013). Resting state detection for gating movement of a neural prosthesis. 2013 6th International IEEE/EMBS Conference on Neural Engineering (NER), pages 665–668. (Cited on pages 65, 77, 78 and 175.)

[Swinnen and Wenderoth, 2004] Swinnen, S. P. and Wenderoth, N. (2004). Two hands, one brain: cognitive neuroscience of bimanual skill. Trends in Cognitive Sciences, 8(1):18–25. (Cited on pages 31 and 177.)

[Tangermann et al., 2012] Tangermann, M., Müller, K. R., Aertsen, A., Birbaumer, N., Braun, C., Brunner, C., Leeb, R., Mehring, C., Miller, K. J., Müller-Putz, G. R., Nolte, G., Pfurtscheller, G., Preissl, H., Schalk, G., Schlögl, A., Vidaurre, C., Waldert, S., and Blankertz, B. (2012). Review of the BCI competition IV. Frontiers in Neuroscience, 6(JULY):1–31. (Cited on page 95.)

[Taylor et al., 2002] Taylor, D. M., Tillery, S. I. H., and Schwartz, A. B. (2002). Direct cortical control of 3D neuroprosthetic devices. Science (New York, N.Y.), 296(5574):1829–32. (Cited on pages 15, 16, 30, 59 and 60.)

[Teplan, 2002] Teplan, M. (2002). Fundamentals of EEG measurement. Measurement Science Review, 2:1–11. (Cited on pages 4, 19 and 20.)

[Thomas et al., 2013] Thomas, E., Dyson, M., and Clerc, M. (2013). An analysis of performance evaluation for motor-imagery based BCI. Journal of Neural Engineering, 10(3):031001. (Cited on page 107.)

[Tillery et al., 2003] Tillery, S. I. H., Taylor, D. M., and Schwartz, A. B. (2003). Training in cortical control of neuroprosthetic devices improves signal extraction from small neuronal ensembles. Reviews in the neurosciences, 14(1-2):107–19. (Cited on page 22.)

[Tomioka et al., 2007] Tomioka, R., Aihara, K., and Müller, K.-R. (2007). Logistic regression for single trial EEG classification. Analysis, 19:1377–1384. (Cited on pages 52 and 53.)

[Townsend et al., 2004] Townsend, G., Graimann, B., and Pfurtscheller, G. (2004). Continuous EEG classification during motor imagery - Simulation of an asynchronous BCI. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 12(2):258–265. (Cited on pages 41, 63, 64 and 75.)

[Tranquillo, 2013] Tranquillo, J. V. (2013). Biomedical Signals and Systems. Synthesis Lectures on Biomedical Engineering, 8(3):1–233. (Cited on page 108.)

[Trejo et al., 2006] Trejo, L. J., Rosipal, R., and Matthews, B. (2006). Brain-computer interfaces for 1-D and 2-D cursor control: Designs using volitional control of the EEG spectrum or steady-state visual evoked potentials. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 14(2):225–229. (Cited on pages 11, 19, 20, 40 and 42.)

[Turkmen and Billor, 2012] Turkmen, A. and Billor, N. (2012). Partial least squares classification for high dimensional data using the PCOUT algorithm. Computational Statistics, 28(2):771–788. (Cited on page 195.)

[Tusell, 2011] Tusell, F. (2011). Kalman filtering in R. Journal Of Statistical Software, 39(2). (Cited on page 172.)

[Unicare, 2017] Unicare (2017). Unicare EEG cap. http://www.eeg-sensor.com/show.php?id=56. (Cited on page 20.)

[Urigüen and Garcia-Zapirain, 2015] Urigüen, J. A. and Garcia-Zapirain, B. (2015). EEG artifact removal-state-of-the-art and guidelines. Journal of Neural Engineering, 12(3):31001. (Cited on page 40.)

[Valstar and Pantic, 2007] Valstar, M. F. and Pantic, M. (2007). Combined Support Vector Machines and Hidden Markov Models for Modeling Facial Action Temporal Dynamics. In International Workshop on Human-Computer Interaction, pages 118–127. (Cited on pages 77 and 118.)

[van Gerven et al., 2012] van Gerven, M. a. J., Chao, Z. C., and Heskes, T. (2012). On the decoding of intracranial data using sparse orthonormalized partial least squares. Journal of Neural Engineering, 9(2):026017. (Cited on pages 47, 59 and 195.)

[Vansteensel et al., 2010] Vansteensel, M. J., Hermes, D., Aarnoutse, E. J., Bleichner, M. G., Schalk, G., Van Rijen, P. C., Leijten, F. S. S., and Ramsey, N. F. (2010). Brain-computer interfacing based on cognitive control. Annals of Neurology, 67:809–816. (Cited on pages 13, 14, 17, 18, 30 and 31.)

[Velliste et al., 2014] Velliste, M., Kennedy, S. D., Schwartz, A. B., Whitford, A. S., Sohn, J.-W., and McMorland, A. J. C. (2014). Motor cortical correlates of arm resting in the context of a reaching task and implications for prosthetic control. The Journal of neuroscience : the official journal of the Society for Neuroscience, 34(17):6011–22. (Cited on pages 29, 52, 62, 64, 65, 66, 70, 102 and 159.)

[Velliste et al., 2008] Velliste, M., Perel, S., Spalding, M. C., Whitford, A. S., and Schwartz, A. B. (2008). Cortical control of a prosthetic arm for self-feeding. Nature, 453(7198):1098–101. (Cited on pages 12, 15, 16, 29, 59 and 60.)

[Vidaurre et al., 2016] Vidaurre, C., Klauer, C., Schauer, T., Ramos-Murguialday, A., and Müller, K.-R. (2016). EEG-based BCI for the linear control of an upper-limb neuroprosthesis. Medical Engineering & Physics, 0:1–10. (Cited on pages 29, 40, 45, 50, 52 and 55.)

[Vidaurre et al., 2009] 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. (Cited on page 45.)

[Vigário, 1997] Vigário, R. N. (1997). Extraction of ocular artefacts from EEG using independent component analysis. Electroencephalography and Clinical Neurophysiology, 103(3):395–404. (Cited on page 40.)

[Vouga et al., 2017] Vouga, T., Zhuang, K., Olivier, J., Lebedev, M., Nicolelis, M., Bouri, M., and Bleuler, H. (2017). EXiO - A Brain-Controlled Lower Limb Exoskeleton for Rhesus Macaques. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 25(2):1–1. (Cited on page 31.)

[Waldert et al., 2009] 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):244–54. (Cited on pages 2, 3, 13, 14, 15, 18, 19, 20, 23, 25, 26, 27 and 41.)

[Waldert et al., 2008] 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. (Cited on pages 23, 24 and 27.)

[Wallstrom et al., 2004] Wallstrom, G. L., Kass, R. E., Miller, A., Cohn, J. F., and Fox, N. A. (2004). Automatic correction of ocular artifacts in the EEG: A comparison of regression-based and component-based methods. International Journal of Psychophysiology, 53(2):105–119. (Cited on page 40.)

[Wang et al., 2009a] Wang, B., Wong, C. M., Wan, F., Mak, P. U., Mak, P. I., and Vai, M. I. (2009a). Comparison of different classification methods for EEG-based brain computer interfaces: A case study. 2009 IEEE International Conference on Information and Automation, ICIA 2009, pages 1416–1421. (Cited on pages 53, 54, 55 and 56.)

[Wang et al., 2015] Wang, J. J., Xue, F., and Li, H. (2015). Simultaneous channel and feature selection of fused EEG features based on Sparse Group Lasso. BioMed Research International, 2015. (Cited on page 48.)

[Wang et al., 2016] Wang, P. T., King, C. E., McCrimmon, C. M., Lin, J. J., Sazgar, M., Hsu, F. P. K., Shaw, S. J., Millet, D. E., Chui, L. a., Liu, C. Y., Do, A. H., and Nenadic, Z. (2016). Comparison of decoding resolution of standard and high-density electrocorticogram electrodes. Journal of Neural Engineering, 13(2):026016. (Cited on pages 18, 50, 51, 173 and 175.)

[Wang et al., 2013a] Wang, P. T., King, C. E., Schombs, A., Lin, J. J., Sazgar, M., and Hsu, F. P. K. (2013a). Electrocorticographic Gamma Band Power Encodes the Velocity of Upper Extremity Movements. International BCI Meeting. (Cited on pages 42 and 43.)

[Wang et al., 2013b] Wang, P. T., Puttock, E. J., King, C. E., Schombs, A., Lin, J. J., Sazgar, M., Hsu, F. P. K., Shaw, S. J., Millett, D. E., Liu, C. Y., Chui, L. a., Do, A. H., and Nenadic, Z. (2013b). State and trajectory decoding of upper extremity movements from electrocorticogram. International IEEE/EMBS Conference on Neural Engineering, NER, pages 969–972. (Cited on pages 43, 62, 63, 65, 66, 102, 103, 165, 201 and 207.)

[Wang et al., 2007] Wang, W., Chan, S. S., Heldman, D. a., and Moran, D. W. (2007). Motor cortical representation of position and velocity during reaching. Journal of neurophysiology, 97(6):4258–70. (Cited on page 61.)

[Wang et al., 2013c] Wang, W., Collinger, J. L., Degenhart, A. D., Tyler-Kabara, E. C., Schwartz, A. B., Moran, D. W., Weber, D. J., Wodlinger, B., Vinjamuri, R. K., Ashmore, R. C., Kelly, J. W., and Boninger, M. L. (2013c). An electrocorticographic brain interface in an individual with tetraplegia. PloS one, 8(2):e55344. (Cited on pages 3, 10, 11, 12, 13, 17, 18, 21, 23, 26, 27, 29, 30, 31, 44, 60, 61, 63, 162, 176 and 177.)

[Wang et al., 2009b] Wang, W., Degenhart, a. D., Collinger, J. L., Vinjamuri, R., Sudre, G. P., Adelson, P. D., Holder, D. L., Leuthardt, E. C., Moran, D. W., Boninger, M. L., Schwartz, a. B., Crammond, D. J., Tyler-Kabara, E. C., and Weber, D. J.

(2009b). Human motor cortical activity recorded with Micro-ECoG electrodes, during individual finger movements. Conference proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference, 2009:586–9. (Cited on pages 18, 31, 47, 53, 54 and 195.)

[Wang et al., 2013d] Wang, X., Mueen, A., Ding, H., Trajcevski, G., Scheuermann, P., and Keogh, E. (2013d). Experimental comparison of representation methods and distance measures for time series data. Data Mining and Knowledge Discovery, 26(2):275–309. (Cited on page 109.)

[Wang et al., 2006] Wang, Y., Hong, B., Gao, X., and Gao, S. (2006). Phase synchrony measurement in motor cortex for classifying single-trial EEG during motor imagery. Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings, pages 75–78. (Cited on page 45.)

[Wang et al., 2012] Wang, Z., Gunduz, A., Brunner, P., Ritaccio, A. L., Ji, Q., and Schalk, G. (2012). Decoding onset and direction of movements using Electrocorticographic (ECoG) signals in humans. Frontiers in neuroengineering, 5(August):15. (Cited on pages 42, 44, 48 and 57.)

[Wang et al., 2011] Wang, Z., Ji, Q., Miller, K. J., and Schalk, G. (2011). Prior knowledge improves decoding of finger flexion from electrocorticographic signals. Frontiers in neuroscience, 5(November):127. (Cited on pages 63, 64, 66, 67, 102, 103, 160 and 171.)

[Waser and Garn, 2013] Waser, M. and Garn, H. (2013). Removing cardiac interference from the electroencephalogram using a modified Pan-Tompkins algorithm and linear regression. Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference, 2013:2028–2031. (Cited on page 40.)

[Waterhouse, 1997] Waterhouse, S. R. (1997). Classification and Regression using Mixtures of Experts. PhD thesis. (Cited on pages 72, 73, 75 and 123.)

[Webb et al., 2012] Webb, J., Xiao, Z. G., Aschenbrenner, K. P., Herrnstadt, G., and Menon, C. (2012). Towards a portable assistive arm exoskeleton for stroke patient rehabilitation controlled through a brain computer interface. Proceedings of the IEEE RAS and EMBS International Conference on Biomedical Robotics and Biomechatronics, pages 1299–1304. (Cited on pages 19, 20, 29 and 40.)

[Wei et al., 2006] Wei, Q., Gao, X., and Gao, S. (2006). Feature extraction and subset selection for classifying single-trial ECoG during motor imagery. Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings, pages 1589–1592. (Cited on pages 45 and 48.)

[Wei and Tu, 2008] Wei, Q. and Tu, W. (2008). Channel selection by genetic algorithms for classifying single-trial ECoG during motor imagery. Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference, 2008:624–7. (Cited on page 41.)

[Wei et al., 2007] Wei, Q., Wang, Y., Gao, X., and Gao, S. (2007). Amplitude and phase coupling measures for feature extraction in an EEG-based brain-computer interface. Journal of neural engineering, 4(2):120–129. (Cited on page 45.)

[Weigend et al., 1995] Weigend, a. S., Mangeas, M., and Srivastava, a. N. (1995). Nonlinear gated experts for time series: discovering regimes and avoiding overfitting. International journal of neural systems, 6(4):373–99. (Cited on page 81.)

[Welch and Bishop, 2006] Welch, G. and Bishop, G. (2006). An Introduction to the Kalman Filter. https://www.cs.unc.edu/~welch/media/pdf/kalman_intro.pdf, 7(1):1–16. (Cited on page 202.)

[Whitham et al., 2007] Whitham, E. M., Pope, K. J., Fitzgibbon, S. P., Lewis, T., Clark, C. R., Loveless, S., Broberg, M., Wallace, A., DeLosAngeles, D., Lillie, P., Hardy, A., Fronsko, R., Pulbrook, A., and Willoughby, J. O. (2007). Scalp electrical recording during paralysis: Quantitative evidence that EEG frequencies above 20 Hz are contaminated by EMG. Clinical Neurophysiology, 118(8):1877–1888. (Cited on page 39.)

[Willett et al., 2013] Willett, F. R., Suminski, A. J., Fagg, A. H., and Hatsopoulos, N. G. (2013). Improving brain-machine interface performance by decoding intended future movements. Journal of neural engineering, 10(2):026011. (Cited on pages 8, 59, 109 and 189.)

[Williams et al., 2013] Williams, J. J., Rouse, A. G., Thongpang, S., Williams, J. C., and Moran, D. W. (2013). Differentiating closed-loop cortical intention from rest: building an asynchronous electrocorticographic BCI. Journal of neural engineering, 10(4):046001. (Cited on pages 17, 18, 27, 29, 53, 59, 63, 65, 70, 71, 73, 74, 103, 159, 161, 162, 165, 171 and 189.)

[Wilson et al., 2006] Wilson, J. A., Felton, E. a., Garell, P. C., Schalk, G., and Williams, J. C. (2006). ECoG factors underlying multimodal control of a brain-computer interface. IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society, 14(2):246–50. (Cited on pages 17 and 18.)

[Wilson et al., 2012] Wilson, J. A., Walton, L. M., Tyler, M., and Williams, J. (2012). Lingual electrotactile stimulation as an alternative sensory feedback pathway for brain-computer interface applications. Journal of neural engineering, 9(4):045007. (Cited on page 5.)

[Wissel et al., 2013] Wissel, T., Pfeiffer, T., Frysch, R., Knight, R. T., Chang, E. F., Hinrichs, H., Rieger, J. W., and Rose, G. (2013). Hidden Markov model and support vector machine based decoding of finger movements using electrocorticography. Journal of Neural Engineering, 10(5):056020. (Cited on pages 31, 48, 56 and 57.)

[Wodlinger et al., 2015] Wodlinger, B., Downey, J. E., Tyler-Kabara, E. C., Schwartz, a. B., Boninger, M. L., and Collinger, J. L. (2015). Ten-dimensional anthropomorphic arm control in a human brain-machine interface: difficulties, solutions, and limitations. Journal of neural engineering, 12(1):016011. (Cited on pages 5, 10, 11, 12, 15, 16, 23, 24, 25, 29, 30, 31, 35, 42, 49, 59, 60, 61, 64, 167, 173 and 177.)

[Wold et al., 2001] Wold, S., Sjöström, M., and Eriksson, L. (2001). PLS-regression: A basic tool of chemometrics. In Chemometrics and Intelligent Laboratory Systems, volume 58, pages 109–130. (Cited on page 189.)

[Wolpaw et al., 2002] Wolpaw, J. R., Birbaumer, N., Mcfarland, D. J., Pfurtscheller, G., and Vaughan, T. M. (2002). Brain – computer interfaces for communication and control. Clinical neurophysiology, 113:767–791. (Cited on pages 3, 4, 6 and 8.)

[Wolpaw and McFarland, 2004] Wolpaw, J. R. and McFarland, D. J. (2004). Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. Proceedings of the National Academy of Sciences of the United States of America, 101(51):17849–54. (Cited on pages 3, 10, 11, 12, 13, 19, 26 and 30.)

[Wong et al., 2013] Wong, Y. T., Putrino, D., Weiss, A., and Pesaran, B. (2013). Utilizing movement synergies to improve decoding performance for a brain machine interface. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, pages 289–292. (Cited on page 46.)

[Wood et al., 2005] Wood, F., Prabhat, Donoghue, J., and Black, M. (2005). Inferring attentional state and kinematics from motor cortical firing rates. Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference, 1:149–152. (Cited on pages 65 and 70.)

[Wu et al., 2003a] Wu, W., Black, M. J., Gao, Y., Bienenstock, E., Serruya, M., Shaikhouni, A., and Donoghue, J. P. (2003a). Neural Decoding of Cursor Motion Using a Kalman Filter. Advances in Neural Information Processing Systems 15: Proceedings of the 2002 Conference, (1):133–140. (Cited on pages 61, 62, 70 and 104.)

[Wu et al., 2002] Wu, W., Black, M. J., Gao, Y., Bienenstock, E., Serruya, M. D., and Donoghue, J. P. (2002). Inferring hand motion from multi-cell recordings in motor cortex using a Kalman Filter. SAB'02- Workshop on Motor Control

in Humans and Robots: On the Interplay of Real Brains and Artificial Devices, pages 66–73. (Cited on pages 61, 62, 70, 201, 206 and 207.)

[Wu et al., 2003b] Wu, W., Black, M. J., Mumford, D., Gao, Y., Bienenstock, E., and Donoghue, J. P. (2003b). A Switching Kalman Filter Model for the Motor Cortical Coding of Hand Motion. In Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE, pages 2083–2086. (Cited on pages 47, 71, 72, 172 and 207.)

[Wu et al., 2004] Wu, W., Black, M. J., Mumford, D., Gao, Y., Bienenstock, E., and Donoghue, J. P. (2004). Modeling and decoding motor cortical activity using a switching Kalman filter. IEEE transactions on bio-medical engineering, 51(6):933–42. (Cited on pages 67, 70, 71, 74, 104, 159 and 172.)

[Wu et al., 2008] Wu, W., Gao, X., Hong, B., and Gao, S. (2008). Classifying single-trial EEG during motor imagery by iterative spatio-spectral patterns learning (ISSPL). IEEE Transactions on Biomedical Engineering, 55(6):1733–1743. (Cited on page 45.)

[Xu et al., 1994] Xu, L., Jordan, M. I., and Hinton, G. E. (1994). An alternative model for mixtures of experts. Nips, (7):633–640. (Cited on page 123.)

[Yanagisawa et al., 2012a] Yanagisawa, T., Hirata, M., Saitoh, Y., Kishima, H., Matsushita, K., Goto, T., Fukuma, R., Yokoi, H., Kamitani, Y., and Yoshimine, T. (2012a). Electrocorticographic control of a prosthetic arm in paralyzed patients. Annals of Neurology, 71(3):353–361. (Cited on pages 8, 13, 17, 18, 26, 30, 44, 51 and 53.)

[Yanagisawa et al., 2012b] Yanagisawa, T., Hirata, M., Saitoh, Y., Kishima, H., Matsushita, K., Goto, T., Fukuma, R., Yokoi, H., Kamitani, Y., and Yoshimine, T. (2012b). Electrocorticographic control of a prosthetic arm in paralyzed patients. Annals of Neurology, 71. (Cited on pages 21, 42, 43, 49, 50, 51 and 52.)

[Yin et al., 2015] Yin, E., Zhou, Z., Jiang, J., Yu, Y., and Hu, D. (2015). A dynamically optimized SSVEP brain-computer interface (BCI) speller. IEEE Transactions on Biomedical Engineering, 62(6):1447–1456. (Cited on page 2.)

[Yu et al., 2007] Yu, B. M., Kemere, C., Santhanam, G., Afshar, A., Ryu, S. I., Meng, T. H., Sahani, M., and Shenoy, K. V. (2007). Mixture of trajectory models for neural decoding of goal-directed movements. Journal of neurophysiology, 97(5):3763–3780. (Cited on pages 67, 70, 71, 159, 164 and 172.)

[Yu, 2010] Yu, S.-Z. (2010). Hidden semi-Markov models. Artificial Intelligence, 174(2):215–243. (Cited on page 174.)

[Yuan et al., 2007] Yuan, H., Doud, A. J., Gururajan, A., and He, B. (2007). Localization of Event-related (de) Synchronization of Cerebral Cortex during Online Control of Brain-Computer Interface Using Minimum-norm Estimates in the

Frequency Domain. International Journal of Bioelectromagnetism, 9(2):109–110. (Cited on pages 19, 20 and 30.)

[Yuan and He, 2014] Yuan, H. and He, B. (2014). Brain-Computer Interfaces Using Sensorimotor Rhythms: Current State and Future Perspectives. IEEE Transactions on Biomedical Engineering, 61(5):1425–1435. (Cited on pages 4, 5, 8, 12, 17, 23 and 27.)

[Yuksel et al., 2012] Yuksel, S. E., Wilson, J. N., and Gader, P. D. (2012). Twenty years of mixture of experts. IEEE Transactions on Neural Networks and Learning Systems, 23(8):1177–1193. (Cited on pages 73, 75, 79 and 82.)

[Zhang et al., 2007] Zhang, C., Fu, H., Jiang, Y., and Yu, T. (2007). High-dimensional pseudo-logistic regression and classification with applications to gene expression data. Computational Statistics and Data Analysis, 52(1):452–470. (Cited on page 196.)

ECoG signal processing for Brain Computer Interface with multiple degrees of freedom for clinical application

Brain-Computer Interfaces (BCI) are systems that allow severely motor-impaired patients to use their brain activity to control external devices, for example upper-limb prostheses in the case of motor BCIs. The user's intentions are estimated by applying a decoder on neural features extracted from the user's brain activity. Signal processing challenges specific to the clinical deployment of motor BCI systems are addressed in the present doctoral thesis, namely asynchronous mono-limb or sequential multi-limb decoding and accurate decoding during active control states. A switching decoder, namely a Markov Switching Linear Model (MSLM), has been developed to limit spurious system activations, to prevent parallel limb movements and to accurately decode complex movements. The MSLM associates linear models with different possible control states, e.g. activation of a specific limb, specific movement phases. Dynamic state detection is performed by the MSLM, and the probability of each state is used to weight the linear models. The performance of the MSLM decoder was assessed for asynchronous wrist and multi-finger trajectory reconstruction from electrocorticographic signals. It was found to outperform previously reported decoders for the limitation of spurious activations during no-control periods and permitted to improve decoding accuracy during active periods.

Keywords: Asynchronous brain-computer-interface, Hidden Markov Models, Mixtures of Experts

Traitement du signal ECoG pour Interface Cerveau Machine à grand nombre de degrés de liberté pour application clinique

Les Interfaces Cerveau-Machine (ICM) sont des systèmes qui permettent à des patients souffrant d'un handicap moteur sévère d'utiliser leur activité cérébrale pour contrôler des effecteurs, par exemple des prothèses des membres supérieurs dans le cas d'ICM motrices. Les intentions de mouvement de l'utilisateur sont estimées en appliquant un décodeur sur des caractéristiques extraites de son activité cérébrale. Des challenges spécifiques au déploiement clinique d'ICMs motrices ont été considérés, à savoir le contrôle mono-membre ou séquentiel multi-membre asynchrone et précis. Un décodeur, le Markov Switching Linear Model (MSLM), a été développé pour limiter les activations erronées de l'ICM, empêcher des mouvements parallèles des effecteurs et décoder avec précision des mouvements complexes. Le MSLM associe des modèles linéaires à différents états possibles, e.g. le contrôle d'un membre spécifique ou une phase de mouvement particulière. Le MSLM réalise une détection d'état dynamique, et les probabilités des états sont utilisées pour pondérer les modèles linéaires. La performance du décodeur MSLM a été évaluée pour la reconstruction asynchrone de trajectoires de poignet et de doigts à partir de signaux electrocorticographiques. Il a permis de limiter les activations erronées du système et d'améliorer la précision du décodage du signal cérébral.

Mots-clefs: Interface Cerveau-Machine Asynchrone, Modèles de Markov Cachés, Mélanges d'Experts