EEG signal analysis for brain-computer interfaces for large public applications

Yuan Yang

▶ **To cite this version:**

Yuan Yang. EEG signal analysis for brain-computer interfaces for large public applications. Signal and Image processing. Télécom ParisTech, 2013. English. NNT : 2013ENST0043 . tel-01234955

HAL Id: tel-01234955

https://pastel.hal.science/tel-01234955v1

Submitted on 27 Nov 2015

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.

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

TELECOM ParisTech

2013-ENST-0043

EDITE - ED 130

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

TELECOM ParisTech

Spécialité « SIGNAL et IMAGES »

présentée et soutenue publiquement par

Yuan YANG

le 08 juillet 2013

Analyse de signaux EEG pour des applications grand-public des interfaces cerveau-machine

Directeurs de thèse : **Isabelle BLOCH et Joe WIART**

Jury

M. Olivier BERTRAND, Directeur de recherche, INSERM, CNRL	Président
M. Christian JUTTEN, Professeur, GIPSA Lab, Université Joseph Fourier	Rapporteur
Mme. Michèle SEBAG, Professeur, LRI, Université Paris-Sud	Rapporteur
M. Jamal ATIF, Maître de Conférences, LRI, Université Paris-Sud	Examinateur
M. Anatole LECUYER, Directeur de recherche, INRIA Rennes	Examinateur
Mme. Isabelle BLOCH, Professeur, CNRS LTCI, Télécom ParisTech	Directrice de thèse
M. Joe WIART, Ingénieur en chef, R&D unit WAVE, France Télécom	Directeur de thèse

TELECOM ParisTech
école de l'Institut Mines-Télécom - membre de ParisTech
46 rue Barrault 75013 Paris - (+33) 1 45 81 77 77 - www.telecom-paristech.fr

T H È S E

Analyse de signaux EEG pour des applications grand-public des interfaces cerveau-machine

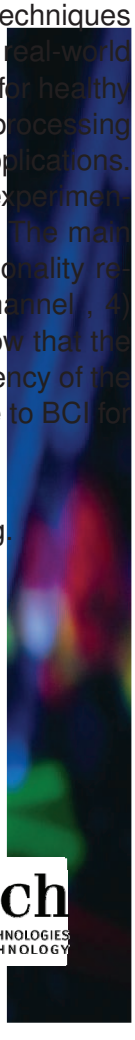
Yuan YANG

RESUME : Les interfaces cerveau-machine (ICM) utilisent les signaux émis par le cerveau pour contrôler des machines ainsi que des appareils (claviers, voitures, neuro-prothèses). Après plusieurs décennies de développement, les techniques de ICM modernes montrent une maturité relative par rapport aux dernières décennies et reçoivent de plus en plus d'attention dans les applications grand public du monde réel, en particulier dans le domaine des interactions homme-machine pour personnes en bonne santé, par exemple les neuro-jeux. L'objectif de cette thèse est de développer un modèle d'ICM et des algorithmes de traitement de signaux EEG pour relever ces défis, donc conduire à une ICM non-invasive, portable et facile à utiliser, exploitant des rythmes EEG pour les applications grand public (non médicales). Pour atteindre cet objectif, un examen de l'état de l'art (prototypes existants et produits commerciaux, configurations expérimentales, algorithmes) a d'abord été effectué pour acquérir une bonne compréhension de ce domaine. Les contributions de cette thèse comprennent : 1) un paradigme ICM hybride avec peu d'électrodes, 2) la réduction de la dimensionnalité pour l'ICM multi-canal (avec un nombre élevé d'électrodes), 3) la réduction et la sélection de canal, 4) l'amélioration de la classification pour l'ICM avec des électrodes prédéterminées. Les résultats expérimentaux montrent que les méthodes proposées dans cette thèse peuvent améliorer les performances de classification et/ou augmenter l'efficacité du système (par exemple, réduire le temps d'apprentissage, réduire le coût du matériel), de manière à contribuer à des ICM pour des l'applications générales.

MOTS-CLEFS : Traitement du signal, Automatique, STIC, Robotique, Signal biomédical,Ingénierie.

ABSTRACT : Brain-computer interfaces (BCIs) use signals from the brain to control machines and devices (keyboards , cars, neuro- prostheses) . After several decades of development, modern BCI techniques show a relative maturity compared to the past decades and receive more and more attention in real-world general public applications, in particular in the domain of BCI-based human-computer interactions for healthy people, such as neuro-games The aim of this thesis is to develop an experimental setup and signal processing algorithms for non-invasive, portable and easy-to-use BCI systems for large public (non-medical) applications. To achieve this goal, a review of the state of the art (existing prototypes and commercial products, experimental setup, algorithms) is first performed to get a full scope and a good understanding in this field. The main contributions of this thesis include : 1) a hybrid BCI paradigm with a few electrodes , 2) dimensionality reduction for multi-channel BCI (with a high number of electrodes), 3) reduction and selection channel , 4) improved classification for BCI with a few predetermined electrodes. The experimental results show that the methods proposed in this thesis can improve classification performance and / or increase the efficiency of the system (for example, reduce the learning time, reduce the cost of equipment) , so as to contribute to BCI for the general applications.

KEY-WORDS : Signal Processing,STIC Automatics, Robotics, Biomedical Signal, Engineering.



TELECOM ParisTech

EDITE

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

ABSTRACT

Brain-computer interfaces (BCIs) bring us the possibility of using our mind to control machines or devices (e.g. keyboards, cars, neuro-prosthetics). The traditional application of this technique mainly focuses on personal assistance for improving the life of disabled people. After several decades of development, modern BCI techniques show a relative maturity compared to the past decades and receive more and more attention in real-world general public applications, in particular in the domain of BCI-based human-computer interactions for healthy people, such as neuro-games. However, several obstacles still impede the widespread adoption of BCI. On the one hand, the traditional definition of BCI, which excludes non-brain-based controls in the system, hinders the freedom of healthy people to realize the control through all possible physiological functions in BCI experiences. On the other hand, the difficulties of signal acquisition and processing, such as electrode redundancy, reduce the portability and practicability of BCI.

The research contribution of this thesis includes two aspects according to the challenges in BCI definition, and in signal acquisition and processing:

A hybrid BCI (hBCI) is first proposed based on one scalp electroencephalography (EEG) electrode placed close to the visual cortex and a few electrooculographic (EOG) electrodes around the orbit. It provides the users with a possibility to communicate through ocular activities and to intentionally decide the timing of communication through brain signals (EEG). As both EOG and EEG data can be detected by an EEG recording system, this design allows using only one recording system with few electrodes, and exploiting this multi-function hBCI, in real-world applications, providing users with more freedom and less cost.

Secondly, the main work of this thesis is to develop machine learning methodologies for addressing the challenges in signal acquisition and processing for easy-to-use motor imagery BCI systems.

The essential task of a motor imagery BCI is to extract the spatial relevant event-related desynchronization (ERD) and/or synchronization (ERS) patterns from EEG signals for identifying subject's motor intentions. However, raw scalp EEG signals have a poor spatial resolution due to the volume conduction effect in EEG measuring. Common spatial patterns (CSP) algorithm is believed to be the most effective spatial filtering method for solving this problem. However, the performance of CSP usually depends on preprocessing (EEG time segment and frequency band selection), channel selection and number of paired filters. Until now, most studies focused on channel selection and improving a preprocessing procedure for CSP, in particular for selecting the optimal frequency band, but not on selecting the number of paired filters in CSP. We propose an automatic method relying on *Rayleigh quotient* to estimate the optimal number of filters for each subject. Based on an existing dataset (BCI competition IV dataset IIa), we study the effect of this parameter on the classification performances. The evaluation on testing data shows that the estimated subject-specific optimal values yield better performances than the recommended value in

the literature.

Note that the CSP algorithm typically needs a multi-channel setting, which may be a drawback for daily-use BCIs in signal acquisition. Thus, we further develop some non-CSP based methods to select the channels and reduce their number for motor imagery BCIs. Unlike existing methods choosing channels without optimization of the time segment, we propose a novel subject-specific channel selection method based on a criterion derived from *Fisher's* discriminant analysis, called F score, to realize the parametrization of both time segment and channel positions. The experimental results show that the method can efficiently reduce the number of channels (from 118 channels to no more than 11), and shorten the training time, without a significant decrease of classification accuracy on a 118-channel dataset (BCI competition III dataset IVa). Compared to existing commercial BCI products (e.g. Emotiv), the number of electrodes we used is reasonable and acceptable for general applications (e.g. in a game environment).

However, algorithm-based channel selection methods usually need a full EEG cap and additional computational time for finding the optimal subset of channels. To solve this problem, some researchers may simply use three bipolar or Laplacian channels located around C3, Cz and C4 (according to the standard 10-20 EEG recording system) to record motor imagery data. This simple placement can easily reduce the number of electrodes but may not always yield good classification performances. To improve classification performances, we propose a novel method to identify subject-specific time-frequency characteristics, so as to extract effective band power features. We adapt the proposed F score to time-frequency selection problems. We also propose a novel alternative criterion based on domain specific knowledge (such as location of brain activity during a motor imagery) to address the specific problem of the time-frequency selection. Both criteria lead to good performances on two standard datasets (BCI competition IV dataset IIa and IIb) for the discrimination between imagination of right and left hand movements, using less electrodes (only two bipolar channels C3 and C4) than other methods.

The time-frequency selection method based on F score is further extended to multi-class problems through one-versus-rest (OVR) strategy. The method is applied with three Laplacian channels C3, Cz and C4 for a four-class BCI problem (left hand, right hand, foot and tongue) and tested on two standard multi-class datasets (BCI competition IV dataset IIa and BCI competition III dataset IIIa). The experimental results show that our method is robust to unpredictable signal changes in session-to-session transfer and artifacts. It can also be used to help reducing the number of electrodes in multi-class problems.

In conclusion, the methods proposed in this thesis can improve the classification performances and/or increase the efficiency of system (e.g. shorten the training time, reduce the hardware cost), so as to contribute to BCIs for general application purposes.

Acknowledgments

First of all, I would like to express my deepest gratitude to my thesis advisors, Professor Isabelle Bloch and Dr. Joe Wiart, for their guidance, support and help during my PhD study. Professor Bloch is a very nice and excellent supervisor. Her comments and suggestions on my thesis were very helpful, which greatly improved the quality of my PhD work. In the past three years, she not only provided me the useful guidance in the research, but also gave me a lot of support and help in other aspects, such as housing, job hunting, etc. I am so lucky that I can be her PhD student. Dr. Wiart is also very nice. He offered me an opportunity to cooperate with Orange Labs, one of the largest driving forces in IT innovation. He gave me a great support in accessing to various EEG devices (BCI products), such as Emotiv, Neurosky, Micromed. His comments and suggestions also benefit my research.

Second, I would like to show my thanks to Dr. Sylvain Chevallier and Dr. Olexiy Kyrgyzov. I had very closed cooperation with them, when they were the postdoctoral researcher in Télécom ParisTech. Like my advisors, they also gave a lot of support and help during my PhD study, not only in research but also in some personal affairs.

I very appreciate the useful discussions with Professor Michèle Sebag, Dr. Jamal Atif, Dr. Slim Essid during my PhD, and the useful comments and suggestions from all jury members.

This work has been performed in the Whist Lab, and partially supported by Chinese Scholarship Council and Orange Labs. Thus, I want to show my sincerely thanks to them.

Last but not least, I would like to express my great gratitude to my mother, Ms. Qing Yang, for her continuous support in my life. Without her encouragement, I could not finish the thesis so smoothly. The same gratitude to other family members and my girlfriend, Miss. Hong Li for all their understanding, support and help. The final thanks are given to all of my friends, who have given me support, help and suggestion during my PhD study.

Table of Contents

Resumé en français

Les interfaces cerveau-machine (ICM, en anglais *brain-computer Interface* ou BCI) utilisent les signaux émis par le cerveau pour contrôler des machines ainsi que des appareils (claviers, voitures, neuro-prothèses) [203]. Une ICM est un système composé de quatre étapes principales : (1) la mesure de l'activité cérébrale, (2) l'extraction de caractéristiques des signaux, (3) la classifiation des signaux, (4) la traduction en une commande envoyée à l'ordinateur. Certains systèmes comprennent également un retour perceptif afin que l'utilisateur puisse voir le résultat du contrôle mental, permettant une rétroaction.

L'application traditionnelle de cette technique se concentre principalement sur l'assistance personnelle pour améliorer la vie des personnes handicapées. Après plusieurs décennies de développement, les techniques ICM modernes montrent une maturité relative par rapport aux dernières décennies et reçoivent de plus en plus d'attention dans les applications grand-public du monde réel [36], en particulier dans le domaine des interactions homme-machine pour des personnes en bonne santé, par exemple les neuro-jeux.

Les systèmes ICM existants utilisent différents types de signaux du cerveau, comme l'électroencéphalographie (EEG) [52, 226], l'électrocorticographie (ECoG) [104], la magnétoencéphalographie (MEG) [123], la spectroscopie proche infrarouge (NIRS) [188] et l'imagerie par résonance magnétique fonctionnelle (IRMf) [221]. Comme elle est peu coûteuse, non-invasive et portable, l'EEG de surface est la plus prometteuse [104]. Aujourd'hui, certaines entreprises, comme Emotiv, Neurosky, OCZ, ont développé des produits commerciaux fondés sur l'EEG pour certaines applications grand public (par exemple, de divertissement, de navigation).

Cependant, plusieurs obstacles entravent encore l'adoption généralisée des ICM. De manière générale, les difficultés proviennent de deux aspects. D'une part, la définition traditionnelle des ICM, qui exclut les contrôles non basés sur le cerveau dans le système, les possibilités de contrôle à travers toutes les fonctions physiologiques possibles. D'autre part, de nombreuses en recherches ICM utilisent un grand nombre d'électrodes (par exemple 64) pour les mesures EEG afin de recueillir suffisamment d'informations pour le décodage précis. Ce paramètre réduit la portabilité et la praticabilité de l'ICM. Pour diminuer le nombre d'électrodes, au moins trois défis doivent être abordés : (1) le choix du nombre d'électrodes, (2) le placement du nombre réduit d'électrodes, et (3) le développement d'une méthode efficace d'apprentissage pour les ICM utilisant seulement quelques électrodes.

L'objectif de cette thèse est de développer un modèle d'ICM et des algorithmes de traitement de signaux EEG pour relever ces défis, donc conduire à une ICM non-invasive, portable et facile à utiliser, exploitant des rythmes EEG pour les applications grand public (non médicales). Pour atteindre cet objectif, un examen de l'état de l'art (prototypes existants et produits commerciaux, configurations expérimentales, algorithmes) a d'abord été effectué pour acquérir une bonne compréhension de ce domaine. Les contributions de cette thèse comprennent : 1) un paradigme ICM hybride avec peu d'électrodes, 2) la réduction de la dimensionnalité pour l'ICM multi-canal (avec un nombre élevé d'électrodes),

3) la réduction et la sélection de canal, 4) l'amélioration de la classification pour l'ICM avec des électrodes prédéterminées.

ICM hybride

Dans la première partie, l'ICM hybride (ICMh) est proposée sur la base d'une électrode EEG placée sur le cuir chevelu à proximité du cortex visuel et d'un petit nombre d'électrodes électrooculographiques (EOG) autour de l'orbite. Elle fournit aux utilisateurs la possibilité de communiquer à travers des activités oculaires et de décider volontairement du moment de la communication par signaux du cerveau (EEG).

Dans le projet d'ICMh, l'EEG a été enregistrée dans la région occipitale (Oz), et les données EOG ont été enregistrées au-dessus (E1, E2) et en-dessous (E3, E4) des yeux gauche et droit, respectivement, et sur le bord extérieur des yeux gauche (E5) et droit (E6) (voir figure 1). Les données sont enregistrées par le système $g.tec^{TM}$ utilisant le lobe de l'oreille droite comme référence (avec un taux d'échantillonnage de 512Hz). Ainsi, une seule électrode EEG et six électrodes EOG sont nécessaires dans l'ensemble du système. Les signaux EEG sont traités pour identifier l'état de l'utilisateur. Si un α-rythme actif (8-13Hz EEG) est détecté, ce qui signifie que l'utilisateur est éveillé, au reps avec les yeux fermés, le système entier est en état d'attente. Quand il détecte que l'α-rythme actif disparaît, le système sera activé et le traitement EOG est déclenché. L'utilisateur peut contrôler la direction du mouvement de l'objet dans un environnement réel ou virtuel par saccades (mouvements simultanés des deux yeux). Quand une saccade est détectée, le système reconnaît sa direction. Ensuite, un message demandera à l'utilisateur si la détection est correcte et intentionnelle. L'utilisateur peut confirmer en clignant des yeux ou infirmer en gardant les yeux ouverts dans le temps imparti.

L'activité de α-rythme est estimée par l'écart-type Σ du signal dans des fenêtres glissantes de 4 secondes. Ainsi, un seuil T_α pour détecter l'α-rythme actif est fixé :

$$T_\alpha = \frac{1}{2}[\min_i(\Sigma_{ferm}) + \max_i(\Sigma_{ouvert})], \tag{1}$$

où Σ_{ferm} et Σ_{ouvert} représentent le Σ des données d'apprentissage dans l'état d'éveil avec les yeux fermés et les yeux ouverts, respectivement. Lorsque la valeur Σ est inférieure à T_α, le système considère que l'α-rythme actif a disparu et déclenche le traitement EOG.

Des expériences ont été effectuées sur quatre personnes en bonne santé, avec: 1) une session d'apprentissage, où les sujets étaient dans des états d'éveil avec les yeux fermés et ouverts, 2min, respectivement, et répétaient 20 fois des clignements et les différents saccades à 10°, 15°, 26°, 30°, 45° à gauche, à droite, vers le haut et le bas pour apprendre les seuils spécifiques des sujets; 2) une session de test, où les données ont été enregistrées pour les mêmes sujets qui ont exécuté 2 min l'état d'éveil avec les yeux fermés, 1min avec les yeux ouverts, puis 6 différentes saccades guidées par des signaux sonores pendant 1,5 min. Un dignement d'œil a été réalisé après chaque saccade pour servir de signal de confirmation. Le seuil spécifique de chaque sujet T_α est estimé par l'équation 1 (voir le tableau 1) et permet avec succès au système de passer d'un mode à un autre en fonction de l'état mental du sujet avec quelques secondes de retard (D). Une relation linéaire a été observée entre les amplitudes EOG et les angles de saccades pour toutes les données

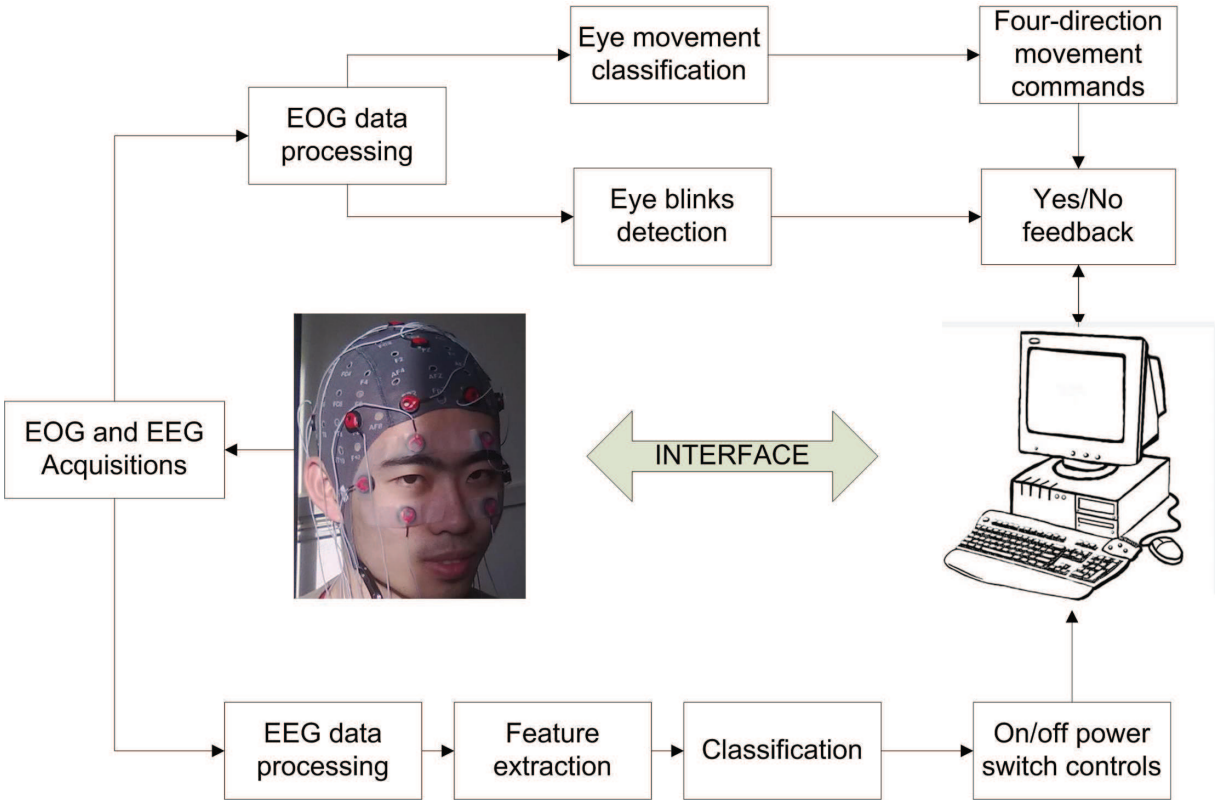


Fig. 1: Système ICM hybride proposé.

d'apprentissage, et permet de définir des plages d'amplitude pour détecter différentes saccades. Les plages (unité : μV) varient en fonction des sujets (par exemple, saccade 15° gauche : [328, 445] pour le sujet 01, mais [144, 195] pour le sujet 03). Le système a détecté à tort deux saccades verticales comme des clignements d'yeux pour le sujet 02 en raison de ses mouvements de paupières pendant les saccades. Ce problème peut être résolu en entraînant les sujets à éviter les mouvements des paupières pendant les saccades.

Table 1: Résultats de l'"interrupteur" du cerveau et de saccade, la détection de dignement d'oeil pour tous les sujets pendant le test. VP/FP indiquent les nombres de vrais/faux positifs.

Sujet	L'interrupteur du cerveau		Saccades	clignements des yeux
	T_α	D(s)	VP/FP	VP/FP
1	31,1	2,11	1,00/0,00	1,00/0,00
2	27,0	2,07	0,67/0,00	1,00/0,33
3	58,8	1,81	1,00/0,00	1,00/0,00
4	16,0	2,58	0,67/0,00	1,00/0,33
5	9,9	3,13	1,00/0,00	1,00/0,00

Ainsi, cette ICMh exploite les activités oculaires comme source de contrôle dans le système, donnant une solution possible aux trois défis mentionnés dans l'introduction. Par ailleurs, elle nous apporte aussi un concept d'utilisation d'un seul système d'enregistrement avec peu d'électrodes, permettant une ICMh multifonctions, pour des applications du monde réel, offrant aux utilisateurs plus de liberté et moins de coût.

Nous devons mentionner que les saccades verticales sont difficiles à détecter lors-que des

mouvements de paupières se produisent au cours de la saccade. Atténuer les effets de l'activité oculaire non volontaire est un défi pour cette ICM hybride. Ainsi, il peut être intéressant d'utiliser un *eye-tracker* pour améliorer le contrôle oculaire dans l'avenir. D'autre part, combiner cette technique avec plus de communication exploitant les signaux du cerveau est également un axe de recherche important qui rendra le système plus puissant pour répondre aux divers besoins d'un utilisateur.

Estimation du nombre optimal de filtres CSP pour l'imagination motrice

Le mouvement réel ou l'imagination d'un mouvement suscitent généralement un ERD (en anglais *event-related desynchronization*), une baisse de puissance dans les bandes de fréquences μ et β dans le cortex sensori-moteur concerné. La tâche essentielle de l'imagerie motrice ICM est d'extraire les modèles ERD de signaux EEG pour identifier les intentions motrices du sujet. Cependant, les signaux EEG acquis sur le cuir chevelu ont une résolution spatiale médiocre en raison de l'effet de conduction de volume lors de la mesure EEG. L'algorithme de configurations spatiales communes (en anglais *Common Spatial Patterns* ou CSP) est une des méthodes de filtrage spatial les plus efficaces pour résoudre ce problème. Cependant, la performance du CSP dépend généralement du prétraitement (les sélections du segment de temps et de la bande de fréquence), des canaux et du nombre de filtres. Jusqu'à présent, la plupart des études ont porté sur la sélection de canaux ou l'amélioration d'une procédure de pré-traitement pour les CSP, en particulier pour sélectionner la bande de fréquence optimale, mais pas sur le choix du nombre de filtres. Nous proposons une méthode automatique s'appuyant sur le *quotient de Rayleigh* pour estimer le nombre optimal de filtres pour chaque sujet. A partir des données de quatre classes de l'imagination motrice (issues du concours BCI IV, données IIa), nous allons vérifier l'efficacité de la méthode proposée en comparant les résultats de la classification en utilisant les valeurs optimales estimées à ceux obtenus avec la valeur fixe recommandée dans la littérature, d'abord dans un problème à deux-classes (discrimination entre l'imagination des mouvements des mains gauche et droite), puis dans un problème à quatre-classes (main gauche, main droite, le pied et la langue).

La méthode proposée pour estimer le nombre de filtres comporte deux étapes :

1. un critère basé sur le *quotient de Rayleigh* est appliqué pour la pré-sélection de la gamme de ce paramètre,
2. un algorithme basé sur la validation croisée est ensuite utilisé pour l'estimation plus précise de la valeur optimale de ce paramètre dans la plage présélectionnée.

L'algorithm CSP est une approche guidée par les données pour construire des filtres spatiaux, $W = [w_1, ..., w_N]$, qui décomposent le signal EEG à N canaux $X = [x_1, x_2, ..., x_N]^T$ en signaux non corrélées $Z = [z_1, z_2, ..., z_N]^T$ par la transformation $z_j = w_j^T X$, $(j = 1, 2, ..., N)$, de manière à calculer les caractéristiques, $f_j^{CSP} = \log(\text{var}(z_j)) = \log(\text{var}(w_j^T X))$, utilisés ensuite pour la classification.

Dans la transformation, w_j est un vecteur propre généralisé qui satisfait les relations :

$$\begin{aligned} w_j^T C^L w_j &= \lambda_j^L \\ w_j^T C^R w_j &= \lambda_j^R \\ \lambda_j^L + \lambda_j^R &= 1 \end{aligned} \tag{2}$$

où C^L, $C^R \in \mathbb{R}^{N\times N}$ sont les matrices de covariance estimées de deux classes, L and R, des signaux EEG. Techniquement, cette procédure de diagonalisation simultanée peut être simplement réalisée en résolvant le problème aux valeurs propres généralisé suivant :

$$C^L w_j = \lambda_j C^R w_j \tag{3}$$

où $\lambda_j = \lambda_j^L/\lambda_j^R$. Comme $\lambda_j^L + \lambda_j^R = 1$, une grande valeur de λ_j^L (λ_j^R) indique que le filtre spatial correspondant w_j tend à donner une grande variance de signal pour la classe L (R) et une petite variance de signal pour la classe R (L). Ces effets contraires de w_j sur deux classes décorrèle les signaux, de manière à contribuer à la discrimination.

Nous trions les valeurs de λ_j^L dans l'ordre décroissant. Ainsi, w_1 correspond à la plus grande valeur propre λ_j^L, et w_N correspond à la plus grande valeur propre λ_j^R (la plus petite λ_j^L). Ils composent la première paire de filtres CSP. De même, w_i et w_{N-i+1} correspondant au i-ème λ^L et λ^R, respectivement, et composent la i-ème paire de filtres CSP.

L'activité discriminatoire S_d et une activité commune S_c entre deux classes sont définies comme :

$$S_d = C^L - C^R \tag{4}$$

$$S_c = C^L + C^R \tag{5}$$

Ainsi, le rapport entre l'activité discriminatoire et l'activité commune projetée sur le filtre spatial w_j est le quotient de Rayleigh $R(w_j)$ [37] et est obtenu par:

$$R(w_j) = w_j^T S_d w_j / w_j^T S_c w_j = \left|\lambda_j^L - \lambda_j^R\right| / (\lambda_j^L + \lambda_j^R) = \left|2\lambda_j^L - 1\right| \tag{6}$$

Pour le i-ème paire de filtres, la quantité $FD(i) = R(w_i) + R(w_{Ni})$ témoigne de son efficacité à extraire les composants discriminants du signal original [110]. Habituellement, les m premières paires de filtres spatiaux ayant les m plus grandes valeurs de $FD(i)$ sont utilisées. De trop petites ou trop grandes valeurs de m mèneront à de mauvaises performances de classification, de sorte que la valeur optimale de m doit être estimée pour chaque sujet. Une trop petite valeur de $FD(i)$ (typiquement $FD(i) < 0,1$, dans la pratique) indique que la i-ème paire de filtres a une très faible capacité d'extraction de composants discriminatoires, et ne peut pas améliorer les résultats de classification. Comme tous les filtres appariés sont classés par ordre de $FD(i)$ décroissant, les valeurs $FD(i)$ sont utilisées comme un critère de présélection pour réduire la gamme de recherche de la valeur optimale de m.

En supposant que le nombre de paires de filtres spatiaux avec $FD(i) \geq 0,1$ est M, le nombre optimal de filtres spatiaux appariés est évalué en vérifiant chaque valeur possible de m afin de voir si la valeur du *coefficient kappa* κ correspondant est sensiblement plus grande que d'autres obtenues pour des valeurs plus petites de m. Un test T est utilisé pour l'analyse de la significativité [231]. Dans ce cas, si plusieurs valeurs de m donnent

Algorithme A: Sélection du nombre optimal de paires de filtres spatiaux

Soit M le nombre de paires de filtres spatiaux avec $FD(i) \geq 0,1$, et $m \leq M$; $\kappa(m)$ est un ensemble de κ pour un m donné évalué avec 100 répétitions de 10 fois la validation croisée ($\kappa(m) \in \mathbb{R}^{1\times 100}$), $\bar{\kappa}(m)$ est la valeur moyenne (sur les 100 composantes), $t(a,b)$ représente les valeurs du test T apparié entre les vecteurs a et b

$m_i \leftarrow 1$; $m_j \leftarrow 2$
while $m_j \leq M$ **do**
  if $\bar{\kappa}(m_j) > \bar{\kappa}(m_i) + \delta$ **and** $t(\kappa(m_i), \kappa(m_j)) < 0,05$ **then**
    $m_i \leftarrow m_j$
  endif
  $m_j \leftarrow m_j + 1$
endwhile
$m_{opt} \leftarrow m_i$
return le paramètre optimal, m_{opt}

des résultats de classification égalux ou comparables, la plus petite valuer sera choisie comme la valeur optimale m_{opt}. Cette procédure est décrite dans l'**Algorithme A**.

Le paramètre optimal m_{opt} est estimé hors ligne à partir des données d'apprentissage pour chaque sujet, puis appliqué aux données de tests ou d'applications en ligne pour le même sujet. Cette stratégie peut être étendue à des problèmes multi-classes [60, 233].

La méthode proposée a été testée sur les données IIa du concours BCI IV [39]. Une large bande de fréquences de 8-30Hz (bandes μ et β) et le segment de 0,5 à 2,5 s de données de l'EEG après le début ont été utilisées dans ce travail pour le calcul de la matrice de transformation W par CSP, les valeurs des $FD(i)$, et l'apprentissage du classifieur. Analyse linéaire discriminante de Fisher (LDA), qui est classiquement utilisé avec le CSP, a été employée ici pour la classification [37]. L'effet du nombre de filtres spatiaux a été étudié sur les données d'apprentissage à l'aide de 100 répétitions de 10 validations. Les performances de classification ont été mesurées par la valeur κ. Les algorithmes de CSP, l'apprentissage et l'évaluation (y compris le calcul de la valeur κ) sont effectués avec la boîte à outils Biosig[1].

Les effets du paramètre de m sur les résultats de la classification et la valeur de $FD(i)$ de chaque paire de filtres spatiaux pour tous les sujets sont illustrés dans la figure 2. De la figure 2, nous pouvons voir que les performances de classification par CSP ne sont pas proportionnelles à m mais présentent des variations importantes selon m pour tous les sujets. Ainsi, il est essentiel de choisir une valeur de m adaptée à chaque individu.

Pour examiner l'efficacité du seuil $FD = 0,1$ à rétrécir l'éventail des m, nous énumérons les valeurs de M pour les différents sujets dans le tableau 2. Dans la figure 2 et le tableau 2, nous pouvons voir que l'ajout de filtres spatiaux avec $FD(i) < 0,1$ n'améliore pas les résultats de la classification (par exemple, pour le sujet 1, $M = 6$, et la valeur de κ diminue si $m > 6$ est utilisé). Ainsi, il est raisonnable d'estimer m_{opt} dans la plage $[1, M]$ en utilisant le seuil $FD = 0,1$.

Le tableau 3 répertorie les valeurs de m_{opt} estimées, obtenues par 100 répétitions de 10 validations croisées dans les données d'apprentissage de deux classes (main gauche et main

[1] http://biosig.sourceforge.net/

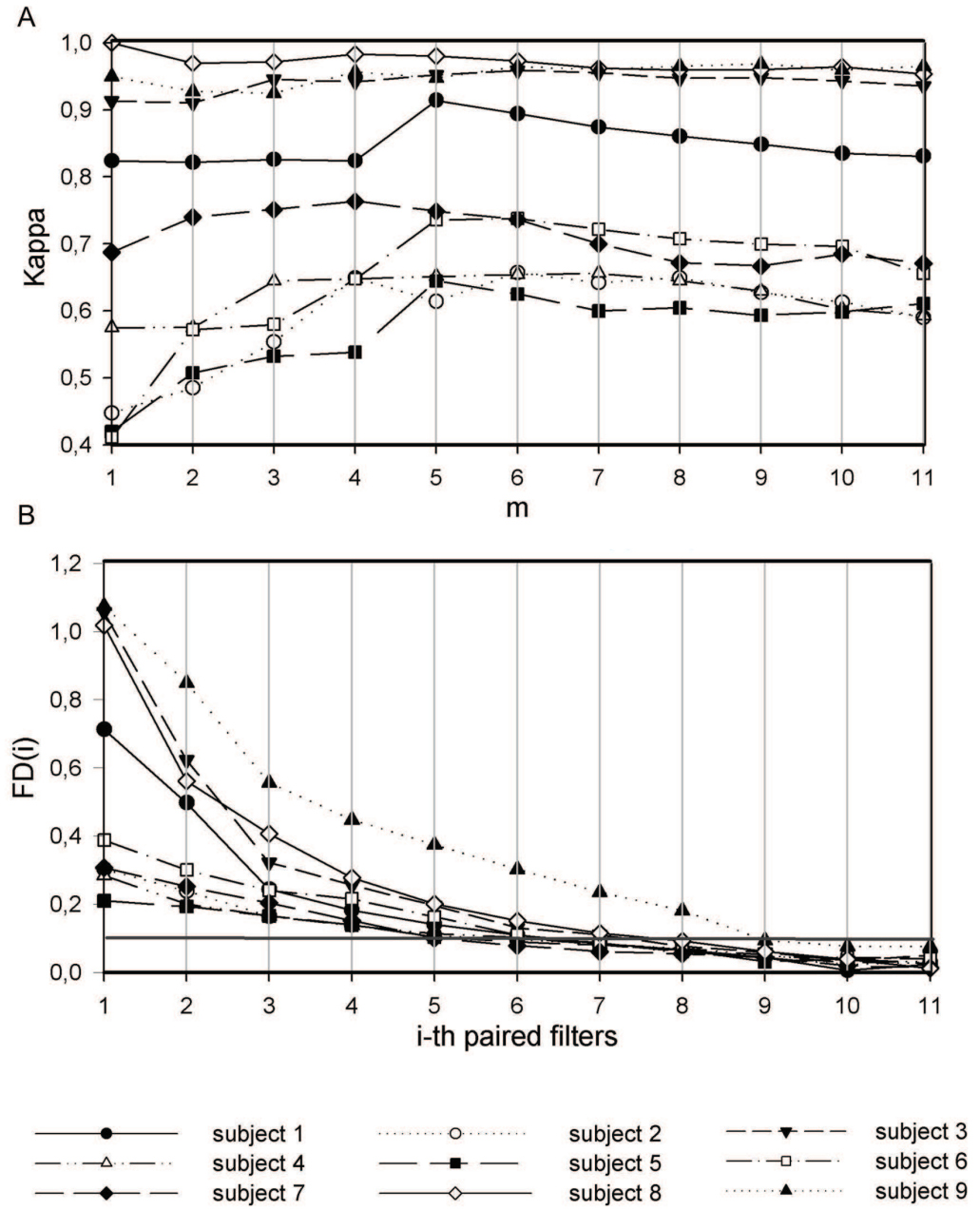


Fig. 2: Effet du paramètre m sur la classification gauche droite, et $FD(i)$ de chaque paire de filtres spatiaux pour tous les sujets dans les données IIa du concours BCI IV. La ligne horizontale sur le second tracé indique $FD = 0,1$.

Table 2: Valeurs de M pour différents sujets.

	Sujets								
	1	2	3	4	5	6	7	8	9
M	6	4	7	6	5	6	5	7	8

droite) et fournit une comparaison des résultats de l'évaluation sur les données de test en utilisant la valeur optimale m_{opt} et la valeur classique $m = 3$ [37, 135]. Dans ce tableau, nous pouvons voir que les performances de classification obtenues avec la valeur estimée sont meilleures que celles obtenues en utilisant la valeur classique ($P < 0,01$).

Table 3: Valeurs estimées, m_{opt}, et comparaison des résultats de l'évaluation indépendante entre l'utilisation de la valeur estimée, m_{opt}, et la valeur classique, $m = 3$. Les meilleures performances sont en gris.

	Sujets									Moyenne
	1	2	3	4	5	6	7	8	9	
m_{opt}	5	4	3	3	5	5	4	1	1	
κ (m_{opt})	0,75	0,22	0,96	0,40	0,11	0,35	0,70	0,94	0,86	0,59
κ ($m = 3$)	0,67	0,13	0,96	0,40	0,09	0,25	0,69	0,93	0,82	0,55

Le CSP un contre les autres (en anglais *one-versus-rest*, ou OVR) est une approche CSP multi-classes qui calcule W pour chaque classes contre toutes les autres, puis projette les signaux EEG sur tous les $2m \times P$ filtres spatiaux choisis (P est le nombre de classes, ici $P = 4$), extrait les caractéristiques, puis effectue une classification LDA multi-classes [60]. Sur la base de la procédure de pré-sélection décrite précédemment, chaque W génère une valeur de M, donc P valeurs de M sont obtenues. La plus grande valeur M (M_{max}) est choisie comme limite supérieure de la valeur possible de m_{opt}. Alors m_{opt} est estimée sur la base des résultats de classification dans la plage [1, M_{max}] en utilisant l'**Algorithme A**, puis appliquée aux données de test. Sans perte de généralité, le cadre de notre méthode pour un problème à quatre classes est illustré dans la figure 3.

La comparaison des résultats obtenus avec m_{opt} et $m = 3$ est présentée dans le tableau 4 pour le problème des quatre classes du concours BCI IV, données IIa. Utiliser m_{opt} conduit à de meilleures performances que d'utiliser $m = 3$. Comme nous avons utilisé une large bande de fréquences (8-30Hz) du signal EEG dans ce travail, il est difficile de faire une comparaison avec le gagnant du concours BCI IV, qui extrait les caractéristiques de plusieurs bandes étroites et a rapporté des résultats basés sur la recherche du plus grand κ sur toute la plage de temps des données de test à l'aide d'une fenêtre glissante de 2s [45]. Cependant, il est plus logique de comparer avec le participant placé deuxième, qui utilise la même bande de fréquence, afin de valider l'intérêt de l'utilisation de m_{opt} adapté au sujet avec OVR CSP. La comparaison a montré que notre méthode nécessite moins de classifieurs (un seul LDA multi-classes) et génère une meilleure performance moyenne.

Table 4: Estimation de m_{opt} et évaluation indépendante dans un problème à quatre classes en utilisant m_{opt} et $m = 3$, et comparaison avec le deuxième gagnant dans le concours BCI IV. Les meilleures performances sont en gris.

	Sujets									Moyenne
	1	2	3	4	5	6	7	8	9	
m_{opt}	4	3	1	2	3	3	2	4	2	
κ (m_{opt})	0,72	0,30	0,74	0,48	0,20	0,25	0,75	0,71	0,68	0,53
κ ($m = 3$)	0,69	0,30	0,71	0,47	0,20	0,25	0,74	0,71	0,50	0,51
κ (2^e, m=4)	0.69	0,34	0,71	0,44	0,16	0.21	0,66	0,73	0,69	0,52

La méthode proposée fournit une solution pour estimer le nombre optimal de filtres spatiaux, et par conséquent réduit également le nombre de dimensions des données de la classification. Cependant, les méthodes par CSP nécessitent généralement une configuration multi-canaux, ce qui présente un inconvénient majeur pour les applications grand public.

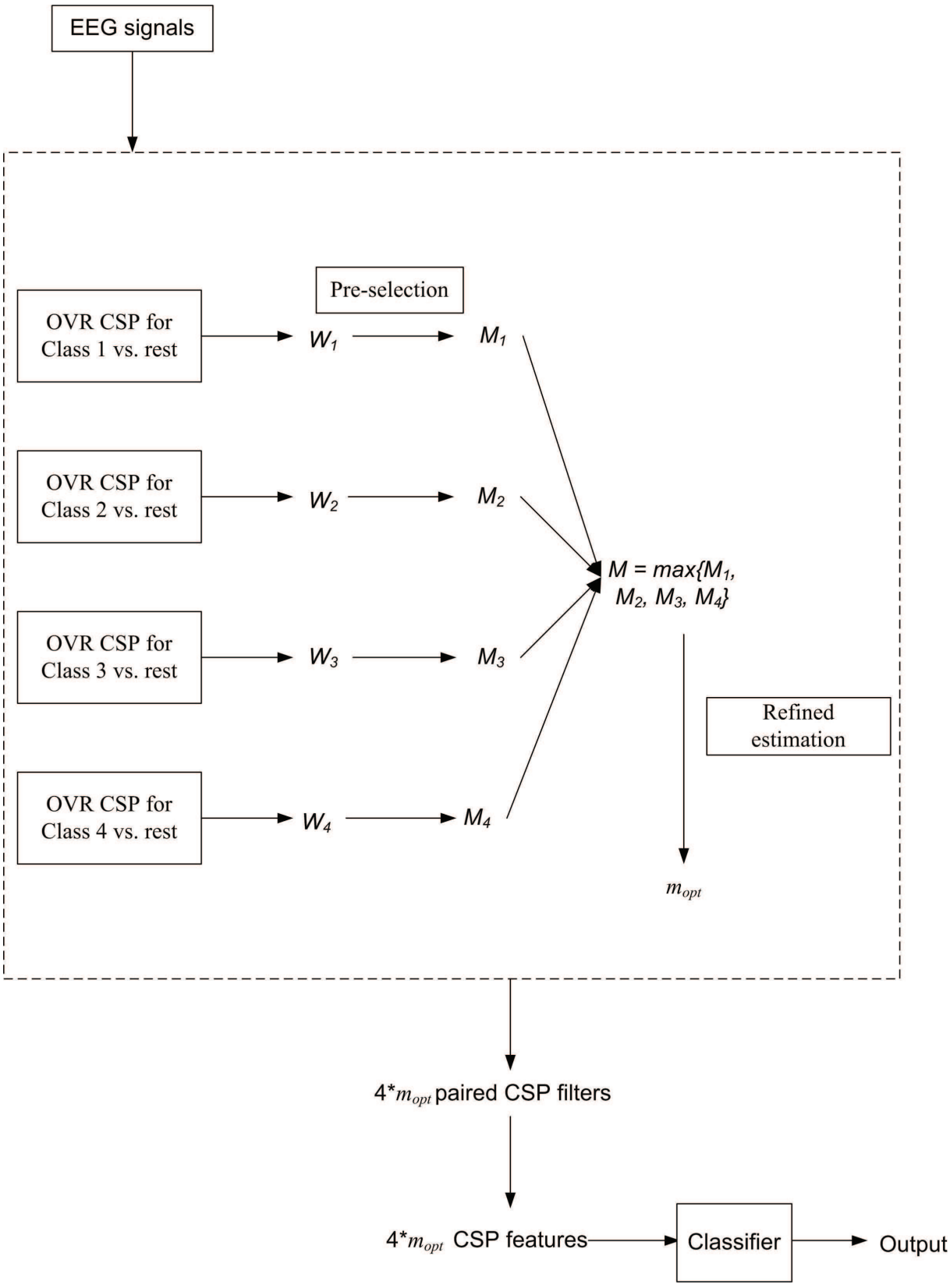


Fig. 3: Méthode proposée pour un problème à quatre classes.

Réduction du nombre d'électrodes

Deux stratégies peuvent être utilisées pour réduire le nombre d'électrodes :

1. Chercher les électrodes principales grâce à des méthodes d'apprentissage automatique.

2. Placer directement quelques électrodes sur certains endroits clés en exploitant des connaissances neurophysiologiques.

Sur la base de la première stratégie, plusieurs algorithmes ont été proposés pour réduire le nombre de canaux dans les ICM [17, 97, 220]. Cependant, ils n'ont abordé la question de la sélection des canaux que par l'information spatiale, sans tenir compte de l'impact potentiel du temps et des fréquences. En conséquence, la combinaison optimale de positions de temps, de fréquences et de canaux ne peut pas être atteinte dans un design ICM. Bien qu'une étude récente a montré qu'une large bande de fréquences (8-30Hz) qui couvre à la fois les bandes μ (8-12Hz) et β (18-25Hz) peut généralement être utilisée pour calculer les paramètres dans le domaine temporel (en anglais *time domain parameter*, ou TDP), les méthodes de sélection de canaux existantes travaillent principalement avec la puissance de la bande (BP), qui est sensible à la bande de fréquences et au segment de temps.

Ici, nous proposons une méthode de sélection des canaux en utilisant les TDP. Contrairement aux méthodes existantes, notre méthode considère l'effet de l'intervalle de temps utilisé sur la sélection des canaux, afin de trouver la combinaison optimale de segment de temps et de sous-ensemble de canaux. Le principle général de la méthode (CSTI) est illustré à la figure 4.

Tout d'abord, nous calculons les TDP caractéristiques (voir l'équation 7) pour chaque canal dans une série d'intervalles de temps de largeur se chevauchant T $[t_n, t_n + T - 1]$ $(n = 1, ..., N)$, $t_{n+1} = t_n + T_s$ (T_s est le pas), pendant la durée de l'imagerie motrice $[T_0, T_e]$, où T_0 est l'heure de début de l'imagerie motrice et T_e est l'heure de fin.

$$TDP^{(p)} = \log\left(\underset{t\in[t_0,t_0+T-1]}{var}\left(\frac{d^p x(t)}{dt^p}\right)\right), p = 0, 1, 2, ... \tag{7}$$

Les TDP sont un groupe de fonctions, dont le pouvoir discriminant n'est pas apte à être évalué par un critère de Fisher classique :

$$FC = \frac{(\mu^h - \mu^f)^2}{(\sigma^h)^2 + (\sigma^f)^2} \tag{8}$$

où μ^h et μ^f sont les valeurs moyennes de la fonction sur l'ensemble des essais pour les classes h et f, respectivement, et $(\sigma^h)^2$ et $(\sigma^f)^2$ sont les variances de la fonction.

Pour évaluer le pouvoir discriminant des TDP, nous avons proposé le F-score:

$$\hat{F} = \frac{\left\|\vec{\mu}^h - \vec{\mu}^f\right\|_2^2}{tr(\Sigma^h) + tr(\Sigma^f)} \tag{9}$$

où Σ désigne la matrice de covariance du vecteur de caractéristiques, $\vec{\mu}$ la moyenne du vecteur de caractéristiques, $\|\cdot\|_2$ la norme $L2$ (norme euclidienne), et $tr(\cdot)$ la trace d'une matrice. Dans le F-score, $\left\|\vec{\mu}^h - \vec{\mu}^f\right\|_2^2$ reflète la différence entre deux classes. La trace de la matrice de covariance pour chaque classe est la distance euclidienne moyenne entre les échantillons et le centre de la classe, qui reflète la dispersion intra-classe.

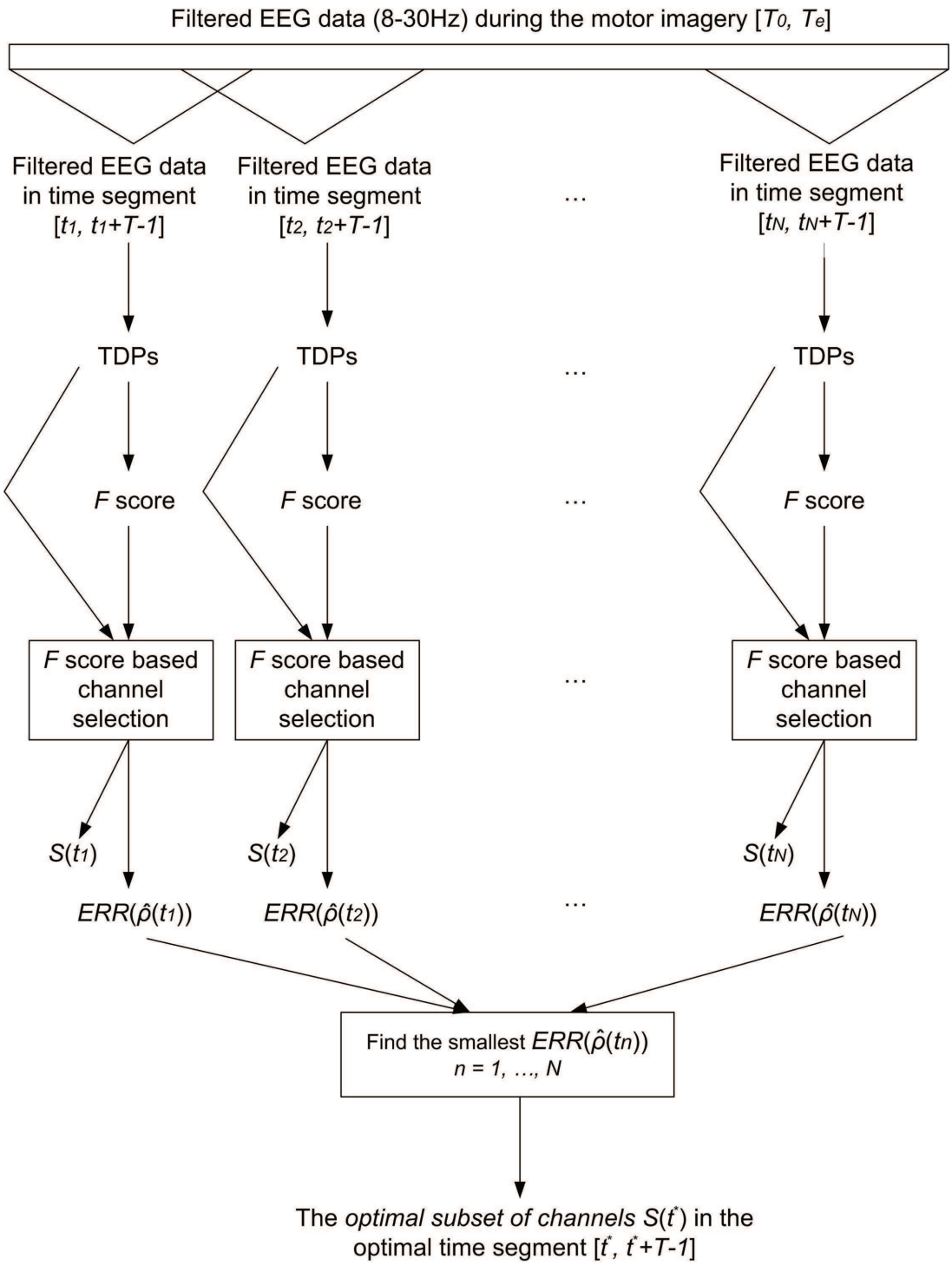


Fig. 4: Schéma général du CSTI.

Ensuite, nous proposons une approche automatique pour déterminer le sous-ensemble de canaux sélectionnés. Soit $\hat{F}_m$ la plus grande valeur du F-score parmi tous les canaux :

$$\hat{F}_{max} = \max\left\{\hat{F}_e \mid e \in \{1, ..., 118\}\right\} \tag{10}$$

Le pouvoir discriminant relatif de chaque canal e est défini comme:

$$\rho_F(e) = \frac{\hat{F}_e}{\hat{F}_{max}} \tag{11}$$

La valeur de $\rho_f(e)$ est comprise entre 0 et 1. Une plus grande valeur de $\rho_f(e)$ indique une puissance de discrimination relative plus importante. Ainsi, un seuil $\hat{\rho}$ peut être défini pour extraire les canaux avec $\rho_f(e) > \hat{\rho}$ qui seront utilisés pour la classification. Une valeur inférieure de $\hat{\rho}$ à tendance à choisir d'autres canaux. En pratique, les données d'apprentissage devraient avoir plusieurs fois la dimension des caractéristiques pour garantir une bonne performance du classifieur. Sur la base de cette connaissance, la gamme de $\hat{\rho}$ peut être réduite à $[P, 1]$ pour nourrir le classificateur, où P est obtenue par :

$$\begin{gathered} \min_P \mathrm{Num}(P) \quad tel\ que \\ P \in [0,1], \mathrm{Num}(P) \geq K/3R \end{gathered} \tag{12}$$

où $Num(P)$ est le nombre de canaux sélectionnés avec $\rho_f(e) > P$, K est le nombre de données d'apprentissage, et R est le rapport entre le nombre de données et le nombre de caractéristiques pour un classificateur spécifique. Notez que nous avons ici $\mathrm{Num}(P) \geq K/3R$ car chaque canal donne trois traits. Comme un classifieur linéaire, tel que le *Fisher's* LDA, a généralement besoin de 5-10 fois plus de données d'apprentissage que la dimension des caractéristiques, nous prenons $R = 5$ pour avoir une gamme lâche de $\hat{\rho}$ pour une optimisation plus poussée. Différents sous-ensembles de canaux en fonction de différents $\hat{\rho} \in [P, 1]$ sont utilisés pour entraîner le classifieur. La valeur optimale de $\hat{\rho}$ est obtenue en cherchant le sous-ensemble avec l'erreur (ERR) la plus faible lors de l'apprentissage du classifieur. L'erreur d'apprentissage est définie comme le désaccord global observé entre les sorties de classification et les véritables classes. La valeur optimale $\hat{\rho}^*$ de $\hat{\rho}$ est donc obtenue par :

$$ERR(\hat{\rho}^*) = \min\{ERR(\hat{\rho}) \mid \hat{\rho} \in [P, 1]\} \tag{13}$$

S'il y plus d'un $\hat{\rho}^*$ obtenu par l'équation 13, nous utilisons le plus grand $\hat{\rho}^*$ comme la solution optimale, afin de trouver les sous-ensembles optimaux des chaînes. L'erreur correspondante est $ERR(\hat{\rho}^*(t_n))$. Le segment de temps optimal $[t^*, t^* + T - 1]$ est alors trouvé en cherchant l'erreur la plus faible $ERR(\hat{\rho}^*(t_n))$ entre tous les segments :

$$ERR(\hat{\rho}^*(t^*)) = \min_{t_n}\{ERR(\hat{\rho}^*(t_n))\} \tag{14}$$

de manière à obtenir le sous-ensemble optimal de canaux $S(t^*)$ dans le segment de temps optimal $[t^*, t^* + T - 1]$.

Cette méthode a été testée sur les données IVa du concours BCI III (pour les informations sur les données, voir [33]). La répartition spatiale de la partition de F-score et les électrodes sélectionnées dans différents segments de temps sont présentées à la figure 5, où les segments de temps sélectionnés sont repérés par des carrés. Les résultats obtenus lors de l'utilisation des électrodes sélectionnées dans différents segments temporels de longueur 2s sont fournis dans le tableau 5, et les résultats des segments de temps sélectionnés sont en gris. Dans la figure 5, nous pouvons voir que les sous-ensembles d'électrodes sélectionnées varient en fonction des segments de temps pour chaque sujet, indiquant que le segment

de temps est un facteur important qui doit être pris en compte dans la sélection des électrodes. Parmi toutes les combinaisons possibles de segment de temps et sous-ensemble d'électrodes, la combinaison estimée donne la précision de classification la plus élevée (ACC) sur les données de tests. Ce résultat montre que le méthode CSTI est efficace pour trouver la combinaison optimale de segment de temps et sous-ensemble d'électrodes. Toutefois, elle a un coût de calcul qui est au moins N fois (N est le nombre des différents segments de temps, ici $n = 5$) celui des méthodes qui effectuent la sélection de canal sur un segment de temps unique. Dans nos expériences, le temps de calcul pour CSTI était de 11 secondes, tandis que pour les méthodes utilisant un segment de temps unique, il était de 2 secondes (Matlab 7.10.0, Windows 7 64bits professionnels, CPU 2.66GHz, RAM 2.0Go). Néanmoins, ce temps supplémentaire reste acceptable pour de nombreuses applications, telles que les neuro-jeux.

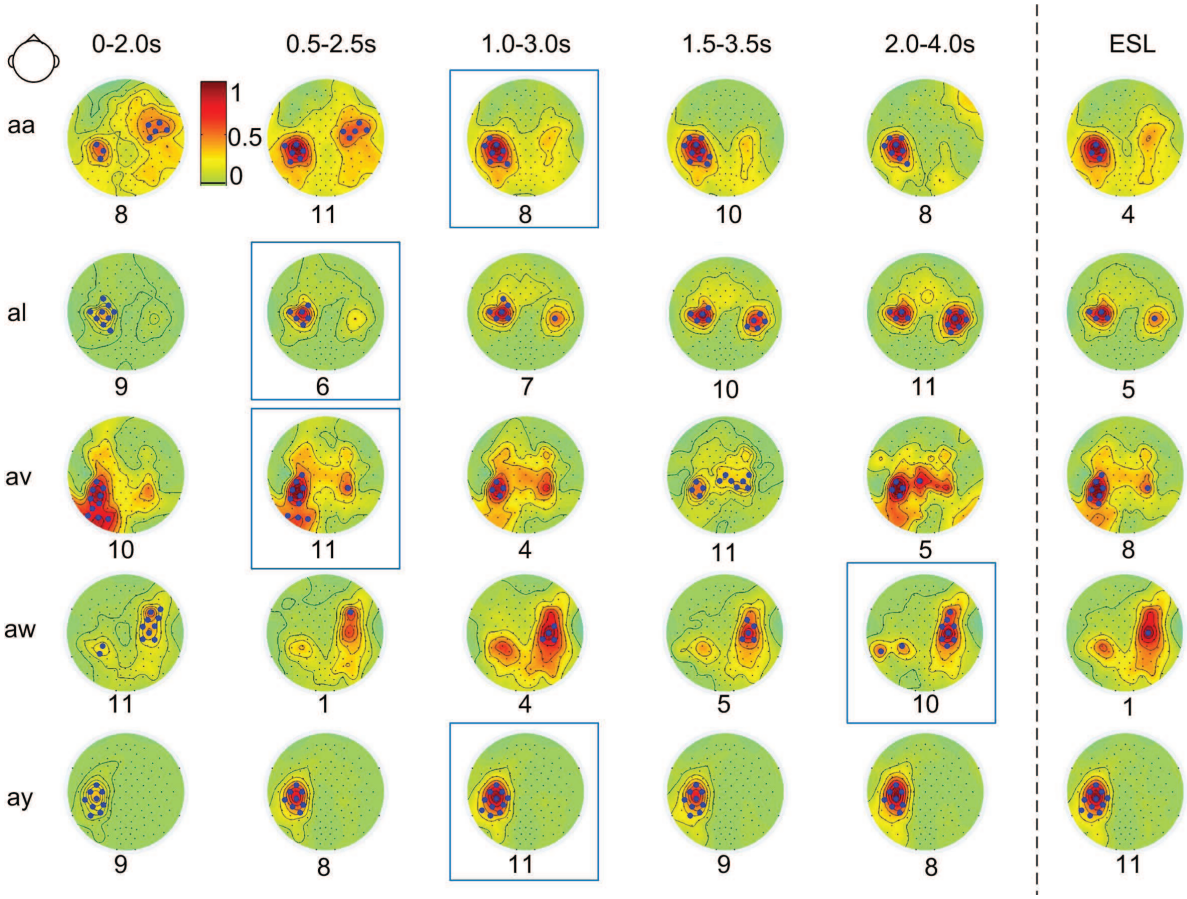


Fig. 5: Cartes topographiques du F-score (échelle de couleurs) et des électrodes sélectionnées (marquées par des points noirs) dans les différents segments de temps. Le nombre d'électrodes sélectionnées est donné en-dessous de chaque carte. Les segments de temps sélectionnés sont repérés par des carrés. Les résultats obtenus par CSL dans un segment de temps long sont fournis dans la dernière colonne.

Des comparaisons entre la sélection des canaux en utilisant les informations de temps (CSTI), la sélection de canaux basée sur le segment de longue durée (CSL), la configuration 3C (C3, Cz, C4), et le CSP classique en utilisant tous les canaux sont effectuées afin d'évaluer la contribution de notre méthode (CSTI). Les résultats sont montrés dans le tableau 6. D'après ce tableau, nous pouvons voir que les résultats obtenus à l'aide de TDP sont meilleurs que ceux utilisant les fonctions de BP pour la plupart des sujets (même

Table 5: Résultats mesurés par ACC pour l'utilisation du sous-ensemble sélectionné d'électrodes dans différents segments de temps (le segment de temps sélectionné est grisé).

Sujet	0-2,0s	0,5-2,5s	1,0-3,0s	1,5-3,5s	2,0-4,0s
aa	0,53	0,67	0,67	0,67	0,64
al	0,78	0,89	0,87	0,87	0,88
av	0,49	0,61	0,61	0,61	0,58
aw	0,66	0,51	0,71	0,75	0,81
ay	0,73	0,90	0,92	0,84	0,80

Table 6: Résultats des tests (ACC) pour les différentes méthodes. La meilleure performance pour chaque sujet est grisée.

	CSP (118)		3C setup (3)		CSTI (< 11)
Subject	BP	TDPs	BP	TDPs	TDPs
aa	0,46	0,47	0,64	0,59	0,67
al	0,94	0,94	0,79	0,81	0,88
av	0.68	0.69	0.58	0.59	0.61
aw	0,94	0,94	0,73	0,79	0,81
ay	0,75	0,84	0,81	0,82	0,92
mean	0,76	0,78	0,71	0,72	0,78

si la différence n'est pas significative $p > 0,05$), ce qui indique l'intérêt d'utiliser TDP dans l'imagerie motrice ICM. Les résultats obtenus avec notre méthode ($\overline{ACC} = 0,78$) sont meilleurs que ceux obtenus avec les trois canaux couramment utilisés C3, Cz, C4 ($\overline{ACC} = 0,72$ lors de l'utilisation de TDP, $\overline{ACC} = 0,71$ lors de l'utilisation de BP, tous p <0,05). Par rapport aux résultats obtenus en utilisant tous les canaux avec CSP, la précision moyenne de la classification par notre méthode est meilleure que celle utilisant tous les canaux avec CSP et les caractéristique BP ($\overline{ACC} = 0,76$, non significatif avec $p > 0,05$) et en utilisant tous les canaux avec CSP et les caractéristiques TDP ($\overline{ACC} = 0,78$ lors de l'utilisation TDP). Pour certains sujets ("aa" et "ay"), notre méthode donne une valeur d'ACC encore plus élevée qu'en utilisant tous les canaux avec CSP. Ainsi, elle répond à l'objectif de réduire largement le nombre d'électrodes (de 118 à pas plus de 11), sans perte majeure de qualité de la classification. De plus, nous utilisons un intervalle de temps relativement plus court (de longueur 2s) que les autres méthodes (longueur 3,5s). Pour la plupart des sujets (sauf "aw"), les sorties de classification ont été obtenues avant la fin de l'acquisition, ce qui indique que moins de temps (ici moins de 3,5s) est nécessaire pour l'enregistrement de données d'apprentissage.

Selection temps-fréquence

Cependant, les méthodes ont généralement besoin d'un casque EEG complet et de temps de calcul supplémentaire pour trouver le sous-ensemble optimal de canaux. Pour résoudre ce problème, certains chercheurs peuvent simplement utiliser trois canaux bipolaires ou laplaciens situés autour de C3, C4 et Cz (selon le système d'enregistrement EEG standard 10-20) pour enregistrer les données d'imagerie motrice. Ce simple placement peut facilement réduire le nombre d'électrodes, mais ne peut pas toujours donner de bonnes

performances de classification. Pour améliorer les performances, nous proposons une nouvelle méthode pour identifier des caractéristiques temps-fréquence spécifiques à chaque sujet, de manière à extraire les caractéristiques de puissance de bande efficaces. Nous adaptons le F-score proposé à des problèmes de sélection temps-fréquence. Nous proposons également un critère basé sur des connaissances spécifiques du domaine (comme l'emplacement de l'activité cérébrale au cours d'une imagerie motrice) pour résoudre le problème de la sélection temps-fréquence.

Le critère proposé pour trouver les régions temps-fréquence est basé sur deux principes neurophysiologiques :

1. L'imagination d'un côté du mouvement de la main génère généralement un ERD dans le côté opposé du cerveau, de sorte qu'il est possible de discriminer entre l'imagination des mouvements de la main droite et de la main gauche en utilisant des électrodes bipolaires placées sur des zones de représentation de la main correspondante, C3 et C4 [156]. Pour obtenir de bonnes performances de classification, 1) la différence de configuration entre imagination des mouvements de la main gauche et de la main droite doit exister dans la région temps-fréquence sélectionnée (en anglais *region of interest*, ou ROI) sur chaque canal; 2) et la différence entre C3 et C4 devrait également exister dans la ROI pour les deux imageries motrices.

2. Les études électrophysiologiques ont souligné le rôle de la conduction de volume, de sorte que les activités neuronales dans un domaine sont réparties sur plusieurs positions d'électrodes [150]. En raison de cet effet, les signaux de certains rythmes de l'EEG indésirables (par exemple des composantes communes) sont également enregistrés, et mélangés avec les signaux spécifiques des différents mouvements de la main, ce qui pourrait affecter les résultats de la classification [37]. Bien que l'enregistrement bipolaire puisse éliminer cet effet dans une certaine mesure, il ne peut pas supprimer complètement l'ensemble de ces éléments communs. Ainsi, nous devrions considérer l'influence de ces composantes communes dans la sélection de la ROI.

Les signaux EEG en C3 et C4 sont décomposés, dans une série de régions temps-fréquence se chevauchant $(\omega_m \times \tau_n)$, $m \in \{1, ..., M\}$, $n \in \{1, ..., N\}$ avec les différentes bandes de fréquence $\omega_m = [f_m, f_m + F - 1], f_{m+1} = f_m + F_s$ (F est la bande passante, F_s est le pas de fréquence) et les différents segments de temps $\tau_n = [t_n, t_n + T - 1], t_{m+1} = t_m + T_s$ (T est la largeur de l'intervalle, T_s est le pas de temps).

Comme les caractéristiques ERD sont souvent mesurées par le logarithme de la variance normalisée du signal EEG filtré par un filtre passe-bande dans un intervalle de temps spécifique, que l'on appelle estimateur de puissance de bande logarithmique (BP) [156, 213], le BP global du segment de l'EEG (pour K essais) dans une région temps-fréquence (ω_m, τ_n) pour chaque chasse ($\chi = L, R$) et chaque canal (e = C3, C4) est estimé comme suit :

$$\widetilde{BP}_e^{\chi} = \log\left(\tilde{v}_e^{\chi}\right) \tag{15}$$

où $\tilde{v}_e^{\chi}$ représente la médiane des écarts types des données $v_e(i)$ sur toutes les données de la classe χ.

Ainsi, la différence de configuration (PD_e) entre deux conditions (gauche vs droite) dans une région temps-fréquence ($\omega_m \times \tau_n$) dans chaque canal est exprimée comme suit :

$$PD_{C3}(\omega_m, \tau_n) = \widetilde{BP}^L_{C3}(\omega_m, \tau_n) - \widetilde{BP}^R_{C3}(\omega_m, \tau_n) \tag{16}$$

$$PD_{C4}(\omega_m, \tau_n) = \widetilde{BP}^L_{C4}(\omega_m, \tau_n) - \widetilde{BP}^R_{C4}(\omega_m, \tau_n) \tag{17}$$

Le signe de PD_e reflète la tendance (augmentation ou diminution) de la modulation BP de la condition L (imagination de mouvement de la main gauche) à la condition R (imagination de mouvement de la main droite) dans le canal e.

L'imagination des mouvements de gauche et de droite produisent habituellement une dominance controlatérale contraire de l'ERD dans les canaux C3 et C4 [156, 152]. Ces modulations spatiales discriminantes liées aux tâches peuvent être mesurées par la force discriminative $Fd(\omega_m, \tau_n) = |PD_{C3}(\omega_m, \tau_n) - PD_{C4}(\omega_m, \tau_n)|$, pour estimer cette contribution positive dans une région temps-fréquence ($\omega_m \times \tau_n$). Une grande valeur de $Fd(\omega_m, \tau_n)$ indique que de grandes modulations discriminantes se produisent dans la région temps-fréquence ($\omega_m \times \tau_n$).

D'autre part, il a été prouvé que d'autres sources (sources de l'imagerie motrice non ciblées) vont générer des signaux (par exemple α-rythme du cortex visuel) à la même fréquence que l'ERD au cours de l'imagerie motrice (pour plus de détails, voir [37, 108]). Par exemple, les sujets sont devant un écran pendant deux tâches d'imagerie motrice, ce qui peut générer des modulations communes visuelles liées à C3 et C4. Bien que ces sources ne soient pas près de C3 et C4, leurs signaux se propagent à travers le cuir chevelu et seront mélangés avec des composantes discriminantes en raison de la conduction de volume [144]. Pendant ce temps, les activités neuronales en C3 et C4 affecteront également les canaux contralatéralux à cause de la conduction de volume. Ce sont ce que nous appelons des composantes communes. Elles se chevauchent avec les modulations discriminantes, ce qui présente un effet négatif sur la classification.

Ainsi, nous définissons le *blurring force*, $Fb(\omega_m, \tau_n) = |PD_{C3}(\omega_m, \tau_n) + PD_{C4}(\omega_m, \tau_n)|$, pour estimer ces modulations communes dans la région temps-fréquence ($\omega_m \times \tau_n$). Une petite valeur de $Fb(\omega_m, \tau_n)$ indique que les petites modulations communes se produisent dans la région temps-fréquence ($\omega_m \times \tau_n$).

Enfin, le critère *Time-frequency Discrimination Factor*, $TFDF(\omega_m, \tau_n)$, est défini comme la différence entre $Fd(\omega_m, \tau_n)$ et $Fb(\omega_m, \tau_n)$ pour évaleur la contribution globale des données dans la région temps-fréquence (ω_m, τ_n) à partir des électrodes C3 et C4 pour discriminer deux classes :

$$TFDF(\omega_m, \tau_n) = Fd(\omega_m, \tau_n) - Fb(\omega_m, \tau_n) \tag{18}$$

Une région temps-fréquence idéale pour la classification devrait avoir de grandes modulations discriminantes (grande valeur de $Fd(\omega_m, \tau_n)$) et petites modulations communes (petite valeur de $Fb(\omega_m, \tau_n)$), de sorte que la ROI ($\omega^* \times \tau^*$) est estimée par la recherche de la valeur maximale de $TFDF(\omega_m, \tau_n)$ entre toutes les régions temps-fréquence :

$$TFDF(\omega^*, \tau^*) = \max \{TFDF(\omega_m, \tau_n) \mid m \in \{1, 2, ..., M\}, n \in \{1, 2, ..., N\}\} \tag{19}$$

La méthode d'optimisation temps-fréquence a été évaluée sur les données IIa [39] et IIb [100] des concours BCI IV. Les résultats sur les données IIb sont présentés dans le

tableau 7. Le $TFDF$ génère la meilleure valeur moyenne de κ ($\bar{\kappa} = 0,62$) parmi toutes les méthodes de l'évaluation indépendante. La valeur moyenne de κ obtenue par le F-score ($\bar{\kappa} = 0,60$) est relativement plus faible que celle obtenue par $TFDF$ (non significativement, $p = 0,29$), mais comparable à celle du gagnant, et supérieure à celle du 2^e (non significativement, $p = 0,52$). Il doit être mentionné que le F-score donne les meilleures valeurs de κ pour la plupart des sujets (4 sujets) parmi toutes les méthodes. Ainsi, les deux critères sont prometteurs pour la recherche des motifs temps-fréquence optimaux pour améliorer les performances de classification à partir de quelques canaux bipolaires seulement.

Table 7: Résultats sur les données IIb [100] des concours BCI IV (évaluées par la valeur κ). La meilleure performance pour chaque sujet est grisée. La dernière colonne donne la performance moyenne sur tous les sujets.

	Sujets									
	1	2	3	4	5	6	7	8	9	Moyenne
$TFDF$	0,44	0,24	0,25	0,93	0,86	0,70	0,55	0,85	0,75	0,62
F score	0,39	0,25	0,13	0,93	0.88	0,63	0,55	0,88	0.78	0,60
sans CSP	0,40	0,24	0,18	0,94	0,39	0,66	0,52	0,81	0,68	0,53
avec CSP	0,28	0,13	0,11	0,47	0,56	0,13	0,58	0,76	0,67	0,42
FBCSP(1^{er}) [5]	0,40	0,21	0,22	0,95	0,86	0,61	0,56	0,85	0,74	0,60
CSSD(2^e) [194]	0,43	0,21	0,14	0,94	0,71	0,62	0,61	0,84	0,78	0,58
NTSPP(3^e) [194]	0,19	0,12	0,12	0,77	0,57	0,49	0,38	0,85	0,61	0,46

Les résultats sur les données IIa sont présentés dans le tableau 8. Le jeu de données IIa dispose de 22 canaux. Nous avons comparé les résultats obtenus par notre méthode avec ceux obtenus par FBCSP [5], CSP parcimonieux (SCSP) [10] et CSP classique, pour évaluer la contribution de notre méthode de réduction du nombre d'électrodes. Notez que FBCSP est considérée comme une méthode efficace qui résout l'optimisation de la fréquence et/ou du temps [8], qui a réalisé la meilleure performance de classification sur au moins deux ensembles de données, y compris les données IIa du concours BCI IV [194]. SCSP est un DSP optimisé qui sélectionne le plus petit nombre de canaux dans la classification par CSP sous une contrainte de précision de la classification. SCSP a généré de meilleures performances que les autres méthodes de réduction des canaux (en fonction du ratio de Fisher, information mutuelle, SVM, coefficients CSP) et le CSP régularisé sur les données IIa du concours BCI IV pour le problème de la main droite vs gauche (pour plus de détails, voir [10]). Ici, les comparaisons des résultats de la classification et le nombre d'électrodes ($\#E$) utilisées dans la classification pour les différentes méthodes sont données. Pour notre méthode reposant sur le F-score ($\overline{ACC} = 79,67$), on obtient des résultats légèrement meilleurs que FBCSP ($\overline{ACC} = 79,17$) et SCSP ($\overline{ACC} = 79,07$), mais en utilisant beaucoup moins d'électrodes sur cet ensemble de données : notre méthode utilise seulement deux canaux bipolaires C3 et C4 (équivalents à 4 canaux mono-polaires), FBCSP utilise tous les 22 canaux mono-polaires, et SCSP utilise 8,55 canaux mono-polaires en moyenne [10]. Un examen plus approfondi des résultats individuels montre que notre méthode par F Score génère le meilleur ACC pour la plupart des sujets (4 sujets), ce qui indique qu'elle est la plus efficace sur cette base de données. Bien que le résultat moyen de classification de notre méthode par $TFDF$ ($\overline{ACC} = 78,00$) soit légèrement inférieur à celui des FBCSP et SCSP, les différences ne sont pas statistiquement significative ($p > 0,05$). En comparant les performances individuelles, notre méthode par

$TFDF$ surpasse FBCSP et SCSP pour 5 des 9 sujets. En outre, notre méthode par $TFDF$ emploie également moins d'électrodes que FBCSP et SCSP. Ainsi, la méthode par $TFDF$ répond toujours à l'objectif de réduction du nombre d'électrodes sans baisse significative de la précision de la classification.

Table 8: Comparaison des résultats de la classification (ACC) et des nombres d'électrodes ($\#E$) utilisé dans la classification entre notre méthode, FBCSP [5], SCSP [10] et CSP classique sur les données IIa du concours BCI IV. La meilleure performance pour chaque sujet est grisée.

		Sujets									
Methode	$\#E$	1	2	3	4	5	6	7	8	9	Moyenne
$TFDF$	4	87,23	66,20	97,81	68,97	64,44	69,44	68,57	96,27	83,08	78,00
F score	4	89,36	69,01	97,81	66,38	66,67	72,22	68,57	97,01	90,00	79,67
FBCSP[1]	22	94,44	52,77	93,05	65,97	88,19	60,41	70,13	94,44	93,05	79,17
SCSP [10]	8,55	91,66	60,41	97,14	70,83	63,19	61,11	78,47	95,13	93,75	79,07
CSP [235]	22	83,51	56,53	97,50	70,00	54,50	62,49	84,50	95,57	90,77	77,26

La méthode de sélection temps-fréquence basée sur le F- score peut également être étendue à des problèmes multi-classes à travers une stratégie une contre tous (en anglais *one-versus-rest*, ou OVR). Elle a été appliquée avec trois canaux de Laplace C3, Cz et C4 pour un problème BCI à quatre classes (main gauche, main droite, pied et langue), et peut donc être utilisée pour réduire le nombre d'électrodes dans des problèmes multi-classes.

En conclusion, les méthodes proposées dans cette thèse peuvent améliorer les performances de classification et/ou augmenter l'efficacité du système (par exemple, réduire le temps d'apprentissage, réduire le coût du matériel), de manière à contribuer à des ICM pour des l'applications générales.

Chapter 1

Introduction

Brain-computer interfaces are systems that measure specific brain activities (e.g. attention level, motor imagery) and then translate them into commands to build a direct communication between brains and computers (see Figure 1.1) [203]. They bring us the possibility of using our mind to control machines or devices (e.g. keyboards, cars, neuroprosthetics). The first brain-computer interface (BCI) research began in the 1970's at University of California Los Angeles [211]. The traditional application of this technique mainly focuses on personal assistance for improving the life of disabled people. After several decades of development, modern BCI techniques show a relative maturity compared to the past decades and receive more and more attention in real-world general public applications [36], in particular in the domain of BCI-based human-computer interactions for healthy people, such as neuro-games.

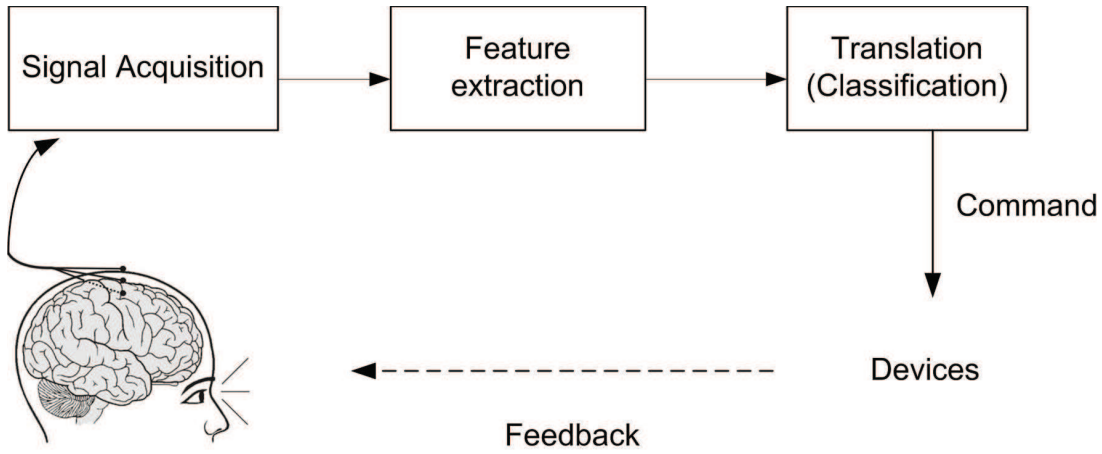


Fig. 1.1: Schematic illustration of the main basic BCI components: signal acquisition, feature extraction, feature-to-command translation, command output pathway, and optional – feedback.

Existing BCI systems can be driven based on various types of brain signals, such as electroencephalography (EEG) [52, 226], electrocorticography (ECoG) [104], magnetoencephalography (MEG) [123], near-infrared spectroscopy (NIRS) [188], and functional magnetic resonance imaging (fMRI) [221]. As it is inexpensive, non-invasive and portable, scalp EEG-based BCI seems to be the most promising type among them [104].

Recently, both large general IT companies, such as IBM, Microsoft, and specialized commercial BCI companies, such as Emotiv, Neurosky, OCZ, have been developing scalp EEG-based commercial BCI products for some general public applications (e.g. entertainment, navigation) [140]. Although many EEG patterns, such as EEG rhythms (α, β,

μ rhythms) and evoked potentials (P300), can be used in BCI systems, most commercial BCI products (e.g. Neurosky, Emotiv and NIA) are based on EEG rhythms. Due to the commercial secret, those companies are unwilling to reveal which rhythms are used in their BCI products. However, the functions for measuring attention levels indicate that at least the α rhythm (8-13Hz) is employed in Neurosky's products and Emotiv systems, since this rhythm is known to be linked to subject's attention levels in EEG study. The other two EEG rhythms often used in BCI are μ (8-12Hz)[1] and β rhythms (13-30Hz). They are known as sensorimotor rhythms and associated with body movements or motor imagery [119]. Motor imagery BCI is one typical scalp EEG-based BCI, which predicts subject's motor intensions through sensorimotor rhythms. Several factors indicate that motor imagery BCI is quite promising. On the one hand, this type of BCI employs the signals from the sensorimotor cortex regions, which are most directly linked to the motor output pathway in the brain. Thus, motor imagery BCI can realize motor tasks without involving muscle activities. On the other hand, motor imagery BCI can be driven by voluntary brain activities without any cues from conditioning protocols, so that the control can be independent and self-paced [174, 225].

One typical two-class motor imagery BCI is to interpret human brain activities corresponding to the imagination (motor imagery) of hand movement. The user is asked to imagine the movement of either the right or the left hand, so as to elicit specific patterns of brain activity in the EEG signals. Artificial intelligent methods are employed to interpret the signals for identifying which hand the user intended to move. This technique has been widely used for helping disabled people, from controlling prosthesis to stoke rehabilitation [54, 155]. As usual games require left and right hand movements, this BCI is now also applied to neuro-games to realize a hand-free game control (see Figure 1.2) [94, 237]. Although two-class motor imagery BCI is helpful for hand-free control, only hand movement discrimination is monotonous for the user. Thus, decoding more than two-class motor imagery data is also required [116]. Several brain signal datasets for classifying motor imagery (for two-class and multi-class problems) are available in BCI competitions[2], so as to encourage researchers to develop algorithms for this purpose [116].

Although the main task of a BCI system is to classify brain signals, developing BCI systems are more complex than a simple classification problem. Even for BCI systems based on scalp EEG, the experimental protocols, and methods for preprocessing (signal enhancement and filtering) and feature extraction can be quite different due to different EEG patterns used. For example, the "peak picking" algorithm [22] is often used in P300-based BCI, since P300 typical has a detectable waveform with a positive peak [160]. However, this algorithm cannot be used in motor imagery BCI, since sensorimotor rhythms do not have a detectable waveform and are often measured by the variance of the signal in a specific time segment and frequency band [88].

In this thesis, we only focus on solving the challenges of developing BCI systems based on EEG rhythms, in particular for motor imagery BCI, for general public (non-medical) applications.

Generally speaking, the challenges come from two aspects. On the one hand, the tradi-

[1]The μ rhythm is different from the α rhythm, though they share the similar frequency band [161]. The μ rhythm occurs in the sensorimotor cortex, and shows activities of motor neurons, while the α rhythm mainly happens in the visual cortex, and is associated with inhibition control. Although the α rhythm is not a new EEG rhythm, its functions and sources are still under exploration until now.

[2]http://www.bbci.de/competition/

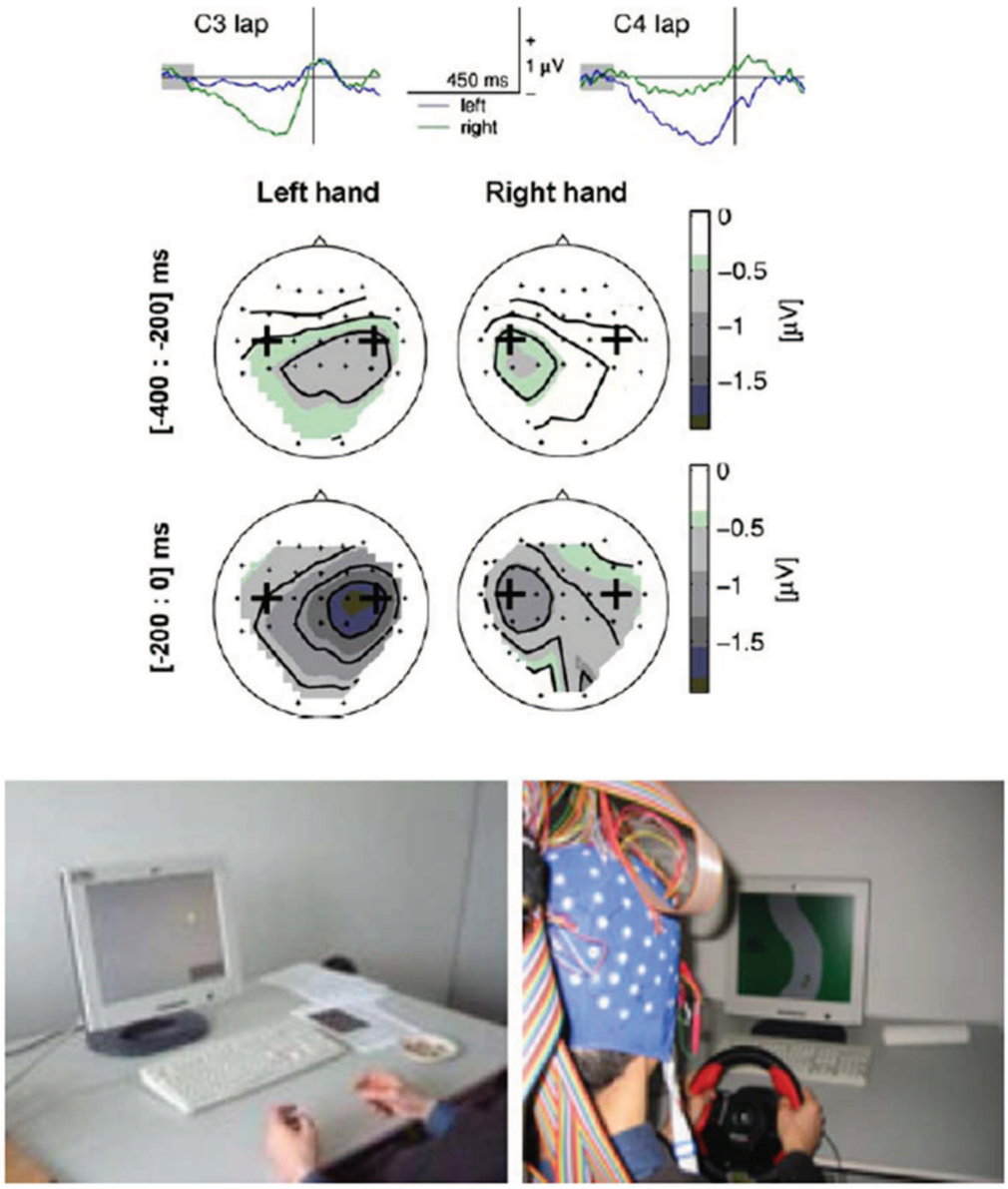


Fig. 1.2: An example of the application of hand motor imagery BCI to games. The figure is modified from [94].

tional definition of BCI, which excludes non-brain-based controls in the system, hinders the freedom of healthy people to realize the control through all possible physiological functions in BCI experiences. On the other hand, many BCI researches use a large number of electrodes (e.g. 64) for EEG measurements in order to gather enough information for precise decoding. This setting reduces the portability and practicability of BCI. To decrease the number of electrodes, at least three challenges have to be addressed: (1) the choice of the number of electrodes, (2) the optimal placement of the reduced number of electrodes and (3) the development of effective machine learning algorithms for BCI systems based on only few electrodes.

The aim of this thesis is to develop a BCI paradigm and EEG processing algorithms for addressing these challenges and therefore lead to non-invasive, portable and easy-to-use BCI systems based on EEG rhythms for large public (non-medical) applications.

To achieve this goal, a review of the state of the art (existing prototypes and commercial products, experimental setup, algorithms) is first performed to get a full scope and a good understanding in this field. The research contribution of this thesis includes two aspects according to the challenges in BCI definition, and in signal acquisition and processing:

A hybrid BCI (hBCI) is proposed based on one scalp EEG electrode (sensor) placed close to the visual cortex and a few electrooculographic (EOG) electrodes around the orbit. It provides the users with a possibility to communicate through ocular activities and to intentionally decide the timing of communication through brain signals (EEG). As both EOG and EEG data can be detected by an EEG recording system, this design brings us a concept of using only one recording system with few electrodes allowing a multi-function hBCI, which can be applied to real-world applications, providing users with more freedom and less cost in BCI experience. It is a contribution to the concept of daily-use BCI.

The main work of this thesis is to develop machine learning methodologies for addressing the challenges in signal acquisition and processing for easy-to-use BCI systems. The main contributions include an automatic selection of spatial filters for improving signal-to-noise ratio (SNR), a novel method for selecting the most useful electrodes for portable BCI, and a new subject-specific optimization for extracting essential time-frequency information for classification. All methods have been evaluated on standard datasets to make comparisons with existing methods. The experimental results show that our methods can improve the classification performances and/or increase the efficiency of system (e.g. shorten the training time, reduce the hardware cost).

The whole thesis is organized as follows: Chapter 2 reviews the state of the art in development of BCI systems for general public applications. Chapter 3 proposes a hybrid BCI (hBCI) system based on a self-paced brain switch and ocular activities. Chapter 4 presents an automatic selection of spatial filters for improving signal-to-noise ratio (SNR) of EEG rhythms in the motor imagery BCI, and also a few preliminary work on electrode (channel) selection. Chapter 5 provides solutions and discussions in selecting the most useful electrodes for portable BCI. Chapter 6 proposes new subject-specific optimization methods for extracting essential time-frequency information from a few interesting electrodes for improving the BCI performances. The time-frequency optimization method is further extended to multi-class problems in Chapter 7. Finally, a conclusion concerning BCI definitions and intrinsic relationship between the number and positions of sensors, preprocessing strategy (spatial filtering), and feature selection, is given, so as to propose the future research directions.

This thesis was performed within the Whist Lab (Institut Mines-Telecom and France Telecom joint Laboratory[3]), and started on 1st October 2010. The work was also performed in collaboration with Dr. Sylvain Chevallier, while he was a post-doctoral researcher at Telecom ParisTech (he is now an Associate Professor at Université de Versailles St-Quentin), and Dr. Olexiy Kyrgyzov, who was a post-doctoral researcher at Whist Lab, and benefitted from useful discussions with Dr. Slim Essid (Associate Professor in Telecom ParisTech) and TAO group at LRI, INRIA (Prof. Michèle Sebag and Dr. Jamal Atif, Associate Professor). Some parts of work also benefitted from suggestions and helps from Brain-Computer Interface (BCI) Laboratory (Dr. Kai Keng Ang and Ms Mahnaz Arvaneh) at Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore, Dr. Gan Huang, University of Hong Kong, China, and EEGLAB

[3]http://whist.institut-telecom.fr/

group, Swartz Center for Computational Neuroscience, USA.

Chapter 2

BCI systems for general public applications: state of the art

2.1 Introduction

In this chapter, we first provide a general review of EEG-based BCI techniques and existing BCI products, and then give detail surveys and discussions on several specific topics about how to bring motor imagery BCI into daily life.

The chapter is organized as follows: Section 2.2 presents the definition of BCI. A brief review of existing scalp EEG-based BCI (e.g. types, products and softwares) is provided in Section 2.3. Section 2.4 introduces the general applications and challenges of motor imagery BCI. Section 2.5 lists the standard datasets and classification evaluation criteria often used in motor imagery BCI research. The signal acquisition and spatial filtering techniques are surveyed in Section 2.6. In Section 2.7, we describe the typical features used in classification and time-frequency optimization methods for improving the performance of BCI systems. The algorithms and methods for addressing the problems of electrode reduction are given in Section 2.8. The conclusion in Section 2.9 leads to the motivation and research direction of this thesis.

2.2 What is BCI?

2.2.1 Traditional definition of BCI

The traditional definition describes BCI as a system that supports a direct communication between brain and computer without any use of peripheral nerves and muscle movements [4, 225]. The basic structure of a BCI typically includes four essential parts: brain signal acquisition, feature extraction, feature-to-command translation, command output pathway. Some systems may contain a feedback. The brain signal can be recorded by various techniques, some are invasive [121, 171], some are non-invasive [193, 228, 229].

BCIs that measure brain activity without surgery are termed non-invasive BCIs, such as scalp EEG-based BCIs, which record EEG signal with electrodes placed on the surface of the scalp [4]. The advantage of scalp EEG-based BCI is that it is inexpensive, of low-risk and portable. However, due to volume conduction through the scalp, skull and other layers of the brain, the EEG recorded by a scalp sensor is a "blurred" copy of multi-source activities [68, 138], which increases the difficulty of signal decoding.

Invasive BCIs acquire signals from sensors implanted in or on the cortex or other brain tissues through neurosurgery, such as ECoG-based BCI [104]. This type of BCI may produce a more detailed view of the brain activity than non-invasive systems [225]. However, expensive surgery, risk of infection, and unclear long-term stability limit its range of applications.

Besides electrical signals (e.g. EEG and ECoG), some nonelectrical signals are also able to drive a BCI system. Techniques that acquire nonelectrical BCI signals include but are not limited to magnetoencephalography (MEG) [123], near-infrared spectroscopy (NIRS) [188], and functional magnetic resonance imaging (fMRI) [221]. Most of them are non-invasive. However, fMRI and fNIR have poor temporal resolution compared to EEG/ECoG [4]. Moreover, fMRI and MEG are very expensive and require highly controlled environment.

Although BCIs measure brain activity, not all devices that record brain signals are BCIs. Systems that acquire brain signals to study cognitive neural activity, mental disorder, sleep stage, etc. are not BCIs, because there is no interface in these systems. Systems that employ non-brain signals (e.g. muscle signals, ocular activities) to realize the interface between human and computer are not BCIs either. Systems that include at least one channel brain-based control but involve other non-brain based channels are not BCIs in the traditional definition; however, some researchers define them as hybrid BCIs [151].

A system that requires some peripheral nerve and/or muscle activity to generate the brain activities which carry the control signal is a dependent BCI. For example, steady-state visual evoked potential (SSVEP) based BCI is a typical dependent BCI, which uses SSVEP as control signal, but the generation of SSVEP depends on the gaze direction. A BCI that does not need activities from peripheral nerves and/or muscles to generate control signals is an independent BCI, such as P300 BCI (see Section 2.3.1) or motor imagery BCI (see Sections 2.3.1 and 2.4).

2.2.2 Hybrid BCI

Unlike traditional BCIs, a hybrid BCI (hBCI) allows users to communicate with computers based on multi brain-based control approaches (e.g. using both event-related desynchronization (ERD) and steady state visually evoked potential (SSVEP) signals) or a combination of brain-based controls and other bio-signals based controls (e.g. electromyography (EMG), heart rate) or assistive techniques (e.g. eye tracking system) [1, 151]. The EU Project "Tools for Brain-computer Interaction" (TOBI)[1] improved the concept of hBCI with the explanation that a BCI channel should be used only if the user needs it [151]. In other words, a TOBI hybrid BCI should choose the most reliable signal (maybe muscle-based signal) from multiple input channels and/or switch between input channels to achieve the best performance of interaction [101]. This control model has been validated in several hBCI models, such as a BCI combining ocular activities (i.e. gazes) and motor imagery control [192], and a SSVEP BCI with an on/off switch based on heart rate [173].

[1] http://www.tobi-project.org/

2.2.3 Self-paced BCI

Most BCIs were designed in the past with an externally-paced (cue-based) mode of operation, where the timing and speed of communication are predetermined by the paradigm instead of the user [172]. On the contrary, a self-paced BCI is able to separate the intentional control activities from the ongoing spontaneous brain activities without external cues [174], so users can intentionally decide the timing and speed of communication by keeping or changing their mental states. This control paradigm has been validated in both mono-modal BCIs (e.g. motor-imagery based BCI, SSVEP BCI) [149, 174] and hybrid BCIs [157, 173]. The key points for realizing a self-paced BCI are (1) to train the user to induce distinctive brain activity patterns, and (2) to train the system to identify those patterns from the on-going EEG [173]. The most successful example is self-paced BCI based on motor imagery, since unlike P300 that needs external stimulus to evoke, a motor imagery task can be performed in a self-paced way without external cue [24]. The typical way toward self-paced operation proposed by the Graz BCI group is to use two classifiers in the system: one for separating motor imagery modulated mental state from non-control state, and the other for further classification of different motor imagery controls (for details, see [157, 173] and Section 2.3.1).

2.2.4 Discussion on BCI definitions

The traditional definition of BCI excludes non-brain based control in the communication. To some extent, this impacts the users' freedom of using all possible physiological functions to realize the communication. The definition of hybrid BCI breaks this restriction. Moreover, assistive interfaces based on non-brain control may complement the current BCI techniques in the communication.

A self-paced BCI provides users the freedom to determine the timing and speed of communication. This control paradigm enables users to control the BCI system in a more natural way than before. Therefore, both hybrid BCI and self-paced BCI should benefit to the users in their BCI experiments and make BCI interesting to a wider audience than before. Thus, the study on hybrid BCI and self-paced BCI will be a part of this thesis (See Chapter 3).

2.3 Scalp EEG-based BCI

Although many brain signals can be used in BCI systems, the most popular BCI systems are based on scalp EEG signals due to their low cost, and to the fact that they are non-invasive and easy to use. In this section, we describe several popular EEG patterns used in present-day BCIs and commercial products based on scalp EEG.

2.3.1 Popular scalp EEG patterns for present-day BCIs

As mentioned in Chapter 1, many scalp EEG patterns can be used in present-day BCI systems to realize different types of controls. They can be categorized into four types

depending on the signal patterns they employ: BCIs based on slow cortical potentials (SCPs), P300, steady-state visual evoked potentials (SSVEPs) and sensorimotor rhythms.

Slow cortical potentials

Slow cortical potentials (SCPs) are potential shifts of the cerebral cortex, which occur over the frequency range below 1-2Hz and can last for a few seconds [26]. This electrophysiological response can be measured by the scalp EEG or MEG. The amplitudes of SCP usually vary from 10 to 100μV. SCPs are generated by the synchronous discharge of afferent excitation of the apical dendrites of cortical neurons [73]. These dendrites are located in the cortex layer. SCPs can be negative potential shift (i.e. signal amplitude decrease), or positive potential shift (i.e. signal amplitude decrease). Negative SCPs are usually related to the mobilization or readiness, while positive SCPs are associated with inhibition of neuronal activity [27]. For example, a negative SCP can be measured when a driver is expecting the green light and is about to start the car; and a positive SCP can be recorded when the brain is quite busy with information processing. Figure 2.1 (A) shows both positive and negative SCPs in a BCI experiment [29].

Although people are not aware of these potential shifts in daily life, they are able to control the SCP amplitude after a training of few months [26]. Thus, SCP has been used as a signal source for EEG-based BCI. A typical example is "Thought Translation Device (TTD)", which already enabled several completely paralyzed patients to communicate through self-regulation of SCPs [28, 29]. Although both healthy and disabled people are able to use SCP-based BCIs, several months training may lower enthusiasm of healthy people for using these BCIs for general purposes (e.g. playing video games). Moveover, SCP-based control is relatively slow, because detectable changes in SCPs need at least several seconds to occur [4].

P300 evoked potentials

P300 is an event-related potential (ERP) component, which is typically evoked by the perceived and discriminating stimuli [160]. It usually occurs over the parietal cortex with a positive peak at approximately 300 ms after the given stimulus [111]. A typical example of P300 is shown in Figure 2.1 (B). The first P300-based BCI system was developed by Farwell and Donchin in 1988 [63]. P300-based BCIs usually present to the subject a matrix of letters or symbols. Each row and each column of the matrix are flashed rapidly (typically 100ms) in a random sequence. The subject is required to focus his/her attention on the letter or symbol that he/she wants to choose. To help the subject to participate with the best performance, researchers often instruct the subject to silently count the flashes that include the desired target while ignoring the others [2]. The flashes of rows and columns containing the desired targets will elicit a P300 in the subject's parietal cortex, since they constitute a rare but perceived event in the context of all other flashes [162]. The typical model for collecting the training data is copy spelling. In this process, the target letters are presented on the top of the screen. The subject is required to gaze on the pre-spelling letter in the matrix during flashes. The flashes are repeated in multiple sequences, so as to elicit a detectable P300 [53, 58]. The performance of P300-based BCI can be affected by several factors, such as the matrix size [2, 182], the inter stimulus

interval (ISI) [63, 122, 182], the flash patterns [3] and the amount of training data [69]. The optimal combination of those parameters is still under analysis. A current research indicates that the online classification accuracy is the highest for a 3×3 matrix and 175ms inter stimulus interval, while the communication speed is the fastest for a 6×6 matrix and 175-ms inter stimulus interval [182]. P300 could also be elicited under auditory stimuli [181], and now auditory P300-based BCIs have also been developed for blind users [65, 95]. Recent studies showed that P300-based BCI systems yield better performance (i.e. faster communication speed and higher classification accuracy) than other EEG signal patterns in many subjects [4, 69, 183]. Moreover, a remarkable advantage of P300-based BCI is that it does not require initial user training. However, it is difficult to develop a self-paced BCI based on P300, since most P300-based BCIs require the user to observe flashes that are presented at a predetermined rhythm. Another problem is that the amplitude of P300 is likely to change over time [165, 190], so that the BCI performance deteriorates after a long time use. It is a big challenge for a long-term-use system [225].

Steady-state visual evoked potentials

When the retina is exposed to a visual stimulus that flickers at rates ranging from 3.5 Hz to 75 Hz, the brain generates neural responses at the same frequency as the one of the visual stimulus, or at a multiple of this frequency [129]. Those responses are known as steady-state visually evoked potentials (SSVEPs) [167] (see Figure 2.1 (C)). The SSVEPs are generated in the visual cortex and could be detected at each stimulus frequency through Fourier analysis of corresponding EEG signals [57]. Thus, it is possible for the system to identify which stimulus the subject is gazing at, in case of stimuli with different flickering frequencies [129, 133]. A typical SSVEP-based BCI system presents to the subject several symbols that flicker at different frequencies on the computer screen. The subject is instructed to focus his/her attention on the target flickering symbol to generate the SSVEP signals for BCI communication. The feedback is often needed, especially in the training procedure, to help the subject to control/keep the amplitude of the brain wave corresponding to the frequency of the stimulus at which the subject is gazing [25]. SSVEP-based BCI could be used in many general public applications, such as robot control [202] and BCI speller [42]. Recently, researchers from Neurosky Inc. introduced a user-friendly SSVEP-based BCI, which uses single-channel EEG recorded by a low-noise dry electrode and a novel stimulus-locked inter-trace correlation (SLIC) method for SSVEP classification [109]. In SSVEP studies, traditional methods often use the information from the frequency domain for classification. SLIC analyzes EEG data in the time domain by computing the correlation between single-trial event-related potentials, which are time-locked to a visual stimulus. Compared to traditional methods, this method does not require any prior knowledge about the frequency of the flashing stimulus or the response waveform of a subject.

Some researchers believe that this system might have a brilliant future in general public applications due to its high robustness and user-friendly features [109, 210]. However, the performances of most SSVEP-based BCI largely depend on the user's ability to control the gaze direction, because SSVEP amplitudes are associated with this direction. Most SSVEP-based BCIs are dependent BCIs, which require some efforts from muscle (in particular ocular muscle) to produce brain signals (i.e. SSVEP) for communication [225]. Some researchers think that dependent BCIs are not "pure BCI", since they cannot be

used without motor control [4]. A SSVEP can also reflect the subject's attention, and some independent SSVEP BCIs were recently proposed based on attention driven control, so as to exclude any motor control [90].

Sensorimotor rhythms

Sensorimotor rhythms are EEG rhythms typically at μ (8-12Hz) and β (12-30Hz) bands recorded over the sensorimotor cortex [119]. Existing studies show that imaginary movements of different body parts can cause an attenuation of sensorimotor rhythms, termed as event-related desynchronization (ERD), at corresponding "active" cortex areas; meanwhile, an enhancement of sensorimotor rhythms called event-related synchronization (ERS) may be observed at other "idling" areas (see Figure 2.1 (D)) [152, 154].

Several factors suggest that sensorimotor rhythms should be good signal patterns for EEG-based BCI. First of all, they contain information from the cortical areas most directly linked to the normal motor output channels. Thus, some precise control may be possible for subjects to control movement of a robot arm or a computer cursor [121, 228]. Furthermore, subjects can voluntarily evoke ERD/ERS by imagining a movement corresponding to any muscle activity, so that the control mode can be self-paced and independent of muscle involvement [99, 172]. Several training sessions are necessary for most users to master the use of BCI based on motor imagery [118]. Typically, they may need five to ten times training for one dimensionality of movement (e.g. moving the cursor from left to right), and each training session lasts less than 30 minutes [4]. Several studies showed that two dimensional movement control is also possible with some additional training [93, 227]. Thus, BCIs based on sensorimotor rhythms, also called motor imagery BCIs, are quite promising to help users to control the movement of output devices through their minds.

2.3.2 BCI products

After several decades of research improvement, BCI research has shown a relative maturity through the demonstration of different prototypes. Several companies adapted BCI systems and the related technologies for real-world applications, to bring them to a broad audience. Here, we briefly describe four most popular BCI products: Neurosky's Mindset, Emotiv EPOC, G.tec's IntendiX and OCZ Technology's Neural impulse actuator.

Neurosky's products

NeuroSky[2] is a manufacturer of inexpensive, portable EEG-based BCI systems for general public applications. It adapted hybrid BCIs based on EEG and EMG to answer a market demand, only for healthy people, within a number of fields, such as entertainment (toys and games), education, sport, research and wellness [240]. NeuroSky has several independent products, such as MindWave Mobile, Necomimi, MindWave and MindSet. Most Neurosky products look like wireless earphones; but all of them include a dry sensor that touches at the forehead. The typical one is Mindset (see Figure 2.2 (A)). MindSet contains ThinkGear technology [130, 131], which measures the EEG signals at a location around

[2] http://www.neurosky.com/

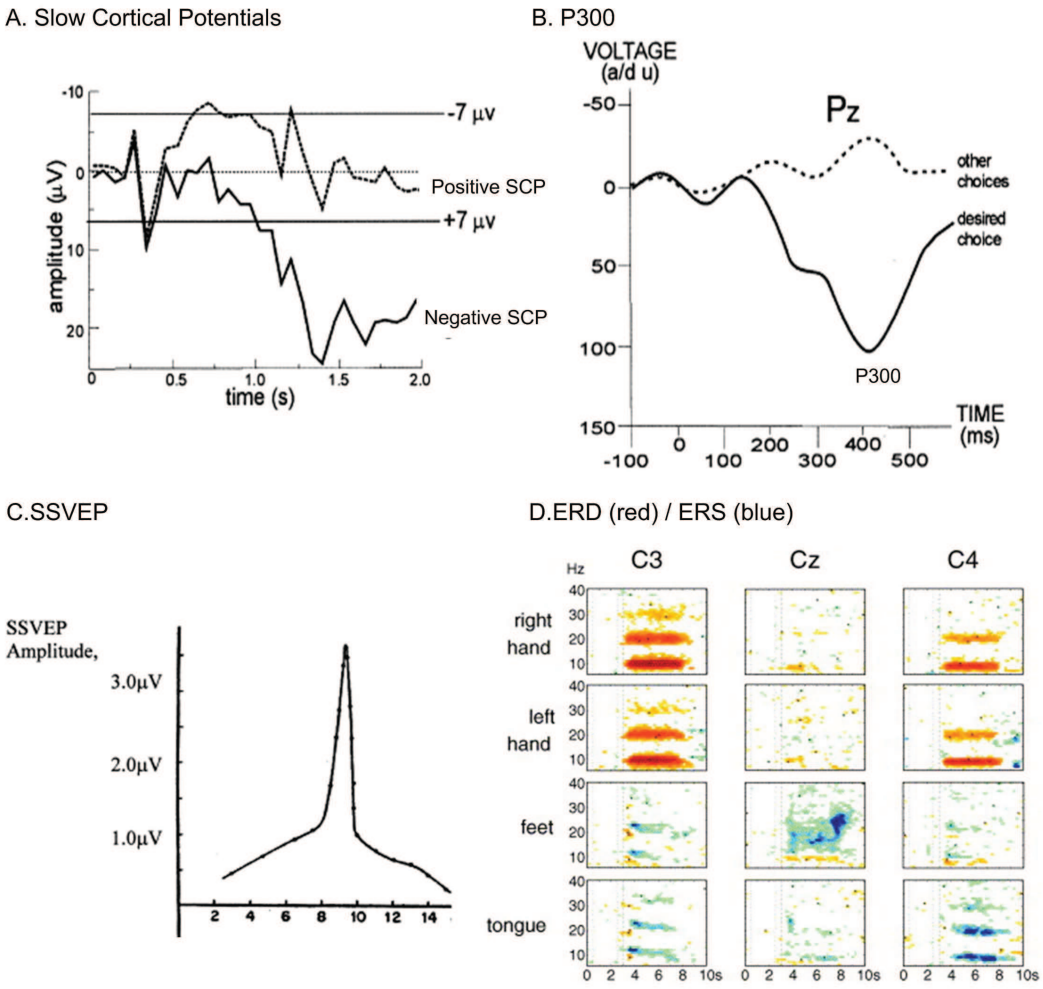


Fig. 2.1: Different EEG patterns for actual BCIs: (A) Typical slow cortical potentials measured in an BCI experiment [29]. (B) P300 obtained in [63]. (C) A SSVEP elicited in an BCI experiment (source from [187]). (D) Time-frequency maps of ERD (areas with red color) and ERS (areas with blue color) for four different classes of motor imagery (right hand, left hand, both feet and tongue) [152]. All illustrations are reproduced from the cited papers.

the pre-frontal cortex through one dry sensor (with reference points at the left ear pad) to access the attention and meditation level measurements for BCI applications [166]. It can record, separate and display various frequencies of EEG simultaneously, detect EOG automatically and also measure attention and meditation levels separately [76]. Here, the meditation level indicates the level of a user's mental calmness [79]. It has two leading techniques: one is the small, low power ThinkGear chip, the other is the low-noise dry sensor. Traditional EEG electrodes are typically made of metal (e.g. Ag/AgCl). A good contact between electrodes and skin needs an electrically-conductive gel. This gel often takes a lot of time to apply when the electrodes are many in the EEG cap (more than 1 hour for 64 channels system). It may diffuse through the hair creating a short circuit between electrodes, or may dry out during the experiment resulting in considerable noise in the recording signals. NeuroSky's low-noise dry sensor technology tries to get rid of those problems [86]. Only one dry sensor is used. This dry sensor touches the skin directly without any gel. However, our experiment showed that the Mindset system is not robust

when there are hairs between the dry sensor and the skin. This drawback may be the reason why all Neurosky's products measure EEG signals from forehead.

Neurosky provides two Mindset toolboxes: Mindset Development Tools (MDT) and Mindset Research Tools (MRT), for development and research purposes, respectively. MDT is a communication socket, which provides the possibility for some self-developed applications/softwares to connect to and communicate with the ThinkGear Chip inside MindSet headsets. It can work with both Windows and Mac OS systems. MRT enables BCI researchers to use the MindSet headsets as a raw EEG data collection device. It includes two softwares: NeuroView and NeuroSkyLab. NeuroView provides the users with an easy way to receive, present, view and store Mindset data in real time, while NeuroSkyLab allows researchers to analyze the data through the free software EEGLAB in the Matlab environment. But MRT only works with Windows System.

The main advantage of Neurosky is its inexpensive, portable and one dry sensor based communication, which makes BCI experiences easier for healthy users. However, the functions of Neurosky products are quite limited. Only a measurement of attention and meditation levels and some simple additional functions, such as eye-blink detection, may not be enough to meet the needs of the users.

Emotiv EPOC

Emotiv Systems[3] is an Australian company developing EEG-based BCI for entertainment and research purposes. Emotiv EPOC is its only current product, which only works with Windows System. It contains 14 wet electrodes, AF3, AF4, F3, F4, F7, F8, FC5, FC6, P7, P8, T7, T8, O1, O2, with 2 reference electrodes located at P3 (CMS) and P4 (DRL), for EEG measurement (see Figure 2.2 (B)). The montage of electrodes obeys a standard EEG recording montage – the 10-20 system (see Figure 2.3) [84]. It also has a two-axis gyroscope to measure the direction of head movement. The sampling rate for EEG recording is around 128Hz using a sequential sampling. Emotiv Systems claim that users can play some games using their mind and facial expressions after a training process. Thus, Emotiv is also a hybrid BCI based on both EEG (from mind controls) and EMG (from facial expressions). It can measure the subject's mental levels (e.g. levels of excitement) through analyzing EEG data and reconstruct a facial expression in the screen by EMG data. However, in our tests we found that this system is sensitive to environment noise and artifacts from body movement, which deteriorates its performance. Raw EEG data from Emotiv EPOC headset are accessible for research purposes but a licence fee has to be paid.

The advantage of this system is that it can fulfill more functions than Neurosky's Mindset with its 14 electrodes. But its wet electrodes and its sensitiveness to noise might be main drawbacks for the end users.

G.tec's IntendiX

G.tec[4] is a biomedical engineering company which produces biosignal amplifiers and electrodes. It shows intensive interest for BCI research, and has supported several years the

[3] http://www.emotiv.com/
[4] http://www.gtec.at/

annual BCI research award. IntendiX is its current BCI product for patients' everyday use (see Figure 2.2 (C)) [206]. IntendiX contains two applications now: IntendiX Speller and IntendiX Screen-Overlay Control Interface (SOCI).

IntendiX Speller is a personal EEG-based typing system. The system works based on visual evoked EEG potentials (VEP/P300) and enables the user to sequentially typing characters using a keyboard-like matrix on the screen just by paying attention on the target character for several seconds [70]. It requires only a few minutes of training. Most subjects can use it after only 10 minutes with a reasonable performance, spelling 5 to 10 characters per minute [43].

SOCI is based on Steady-State Visual Evoked Potential (SSVEP). It presents to the users several icons on the screen. Each icon flickers at a certain frequency. The users can select a icon on the computer screen by paying attention to one of the icons [56]. G.tec claims that SOCI can achieve an accuracy of up to 98% for some subjects[5].

G.tec's Intendix is more like a mature prototype of BCI rather than a real commercial product. For the end-user, it may present several drawbacks, such as wired electrodes, relatively expensive price, requirement of professional aids. It is aimed to help patients, so that it is unavoidable to lose the market of healthy users. However, this system is quite suitable for research purpose or second development. For example, the positions of the electrodes in this system can be changed, so that researchers can record the data at different positions to seek the optimal recording positions for different subjects.

Neural Impulse Actuator

The Neural Impulse Actuator (NIA)[6] is a hybrid brain-computer interface (BCI) developed by OCZ Technology [72] (see Figure 2.2 (D)). It has three dry electrodes on forehead, and mainly serves for entertainment purposes, such as BCI games [40]. NIA is not purely driven by electrical activity from the brain. What NIA captures is a mixture of several bio-signals from muscles, skin and neural activities. NIA can interpret two EEG rhythms (i.e. α and β rhythms), signals from facial muscles and eye movements. It allows users to choose using brain, muscle or ocular signals to control a paddle moving in a game. This design agrees with the new concept of hybrid BCI proposed by Graz BCI (see Section 2.2.2). A test showed that it is easy to use muscle signals to realize the control in this system but quite hard to use brain and ocular signals (for details, see [240]).

The advantages of this system are its dry sensors and its low price. However, it is not always easy to use even though it still has limited functions.

2.4 Motor imagery BCI for general public applications and its challenges

The original and still prevalent application of BCI is to provide an assistive technology for disabled people [128]. However, despite the evolution of the technology, several types

[5] http://www.intendix.com/
[6] http://www.ocztechnology.com

Fig. 2.2: Illustration of different BCI products. (A) Neurosky's Mindset. (B) Emotiv devices and its montage of electrodes. (C) G.tec's IntendiX. (D) Neural Impulse Actuator.

of BCI systems are still designed and tested for healthy people only. Motor imagery BCI is one typical Scalp EEG-based BCI. The general public applications of motor imagery BCI like other types of BCI are mainly focused on developing next generation entertainments and intelligent systems [141], such as neuro-games [48] and navigation in virtual environment [102]. A research group in Berlin has developed motor imagery BCIs to play games [94], such as Pacman, Pong and Tetris, and also a BCI-controlled web browser [36]. A recent user-satisfaction survey revealed that the motor imagery BCI-based version of a game was more engaging and interesting for users than its physical keyboard version [204], which indicates that motor imagery BCI is promising for neuro-games. An application in virtual reality is proposed in [153]. The subject "walks" in the virtual world by imagination of foot movements, and "touches" objects by imagination of hand movements.

The advantage of motor imagery BCI is that it can identify users' motor motivations, so that it may further have some potential applications in the modern industry and military domain. For example, people might use it to control robotic arms to realize some tasks in dangerous environments in the future, such as grabbing a very hot bottle.

The underlying physiological phenomenon is that motor imagery of a specific body part (e.g. left hand, right hand) induces an event-related desynchronization (ERD) in the μ and β bands over the corresponding functional area in the sensorimotor cortex [154]. Thus, the essential task of a motor-imagery based BCI is to extract the spatial relevant ERD patterns from EEG signal for identifying the subject's motor intensions. However, ERD/ERS patterns are typically short-lasting (half to few seconds) and their frequency range may vary with subjects [234]. How to find the right frequency band and the optimal time segment of EEG is one of the major problems in motor imagery BCIs.

Moveover, due to volume conduction through the scalp, skull and other layers of the brain, the EEG signal recorded by a sensor is a "blurred" copy of multi-source activities [68, 138].

As a result, other undesirable signal patterns (e.g. strong α-rhythm and other on-going background EEG components) may submerge ERDs, which leads to a poor signal-to-noise ratio (SNR) and increases the difficulty of extracting ERDs. To overcome this problem, multi-channel EEG recording has been used in many BCI systems to acquire more than enough information for further feature extraction and selection [163]. Spatial filters such as common average reference (CAR) [170], orthogonal source derivation (i.e. Laplacian derivation) [14] and common spatial pattern (CSP) [135] are often used in multi-channel EEG data classification for increasing the SNR and enhancing the patterns of interest. However, the multi-channel setting reduces the portability and practicability of BCI, which presents a major drawback for end-users.

We will review the existing methods for solving those challenges in Sections 2.6, 2.7 and 2.8, so as to seek for the possible way to bring motor imagery BCIs out of labs into general public applications (e.g. entertainments, industry applications). Before reviewing those methods, we introduce in the next section some standard datasets, which are often used to test the methods, and several criteria for evaluating their performances.

2.5 Standard datasets and classification evaluation

2.5.1 Standard datasets

BCI competitions provide many standard datasets of motor imagery BCI. From Table 2.1, we can see that more than 80% datasets are motor imagery BCI datasets in available databases, indicating that people pay more attention to the challenges of motor imagery BCIs than other types of BCI. This section only describes some popular datasets. Information on other datasets can be found on the websites of BCI competition II[7], III[8] and IV[9].

Table 2.1: Number of different types of BCI datasets in BCI competitions.

Competitions	Motor imagery	P300	SCP	SSVEP
I	unavailable now			
II	3	1	1	0
III	7	1	0	0
IV	5	0	0	0
total	15	2	1	0

- **Datasets for channel reduction**:

  BCI competition III dataset IVa: As the most frequently used database, BCI competition III dataset IVa [33] has been used as the experimental data in [16, 17, 97, 124, 143, 220, 239]. This dataset includes 118-channel cue-driven motor imagery EEG data with two classes (right hand v.s. right foot) from five subjects, which is suitable to test the ability of algorithms to reduce and select channels from a

[7]BCI competition II: http://bbci.de/competition/ii/
[8]BCI competition III: http://bbci.de/competition/iii/
[9]BCI competition IV: http://bbci.de/competition/iv/

large number of EEG recording channels in binary-class problems. In the competition, this dataset is aimed to pose the challenge of reducing the time needed for the initial training. For some subjects, only a few training data are available to teach the classifier (see Table 2.2), so as to encourage players to develop algorithms that can work on small training sets. However, the classification accuracies in Table 2.2 show that for some subjects, like subject "aw", only a few initial training data are enough to teach the classifier, so whether a short time training is possible depends not only on the machine learning method but also on the subject. Thus, after this competition, this dataset is often used for the research on channel reduction instead of its original purpose.

Table 2.2: Number of training (labeled) trials "tr" and test (unlabeled) trials "te" for each subject, and classification accuracies (see Section 2.5.2) of three top players (listed by rank) in the BCI competition III [35].

Subject	tr	te	1	2	3
al	224	56	1.00	0.98	0.95
aa	168	112	0.96	0.89	0.82
av	84	196	0.81	0.77	0.70
aw	56	224	1.00	.92	0.88
ay	28	252	0.98	0.81	0.88

BCI competition II dataset IV [32]: It contains data from a single subject with two finger movements (fingers from left and right hands), recorded using 28 EEG channels, which have been used in [110] as experimental data for reducing the number of electrodes. This dataset was recorded from a normal subject during a no-feedback session. The real finger movements are performed in a self-chosen order and a self-paced timing, thus it poses the challenge of algorithm design for self-paced BCI systems. Meanwhile, as finger movements unavoidably cause artifacts to EEG signals, this dataset can also be used to test the robustness of a method to noise.

BCI competition III dataset I [96]: This dataset contains two-class cue-driven motor imagery (left small finger v.s. tongue) data from a single subject (epileptic patient) but with 64 ECoG channels, which is an invasive BCI system. ECoG data usually have higher SNR than scalp EEG data. As the training and testing data are recorded in different days, this dataset poses the challenge of using a classifier that was trained on the first day to classify the data recorded during the following days (without retraining). It is a common but tough challenge due to at least two reasons: 1) the patient might be in a different state concerning motivation, fatigue, etc., so that his brain will show a different electrical activity, 2) the recording system might have undergone slight changes concerning electrode positions and impedances. Thus, this dataset can be used to test the robustness of a method to data evolution for a long-term-use invasive BCI system. This dataset has also served as experimental data with BCI competition II dataset IV in [110] for electrode reduction.

- **Datasets for multi-class problems**:

  BCI competition III dataset IIIa[10]: This dataset contains four-class cue-driven motor imagery (left hand, right hand, foot and tongue) data recorded from three subjects

[10] http://www.bbci.de/competition/iii/desc_IIIa.pdf

(ranging from quite good to fair performance) using 60 EEG channels. In this dataset, the trials contaminated by artifacts are marked. As there are no EOG or EMG channels used for artifact correction, a small Laplacian derivation (see Section 2.6.3) is performed to remove muscle and ocular artifacts. Thus, the signal in this dataset is relatively clean. The challenge is to develop an effective method to improve the multi-class BCI performance (evaluated by kappa coefficient, see Section 2.5.2).

BCI competition III dataset V [127]: It is a three-classes (left hand, right hand, and mental imagery of word) recorded from three normal subjects using 32 EEG channels. The raw EEG potentials were spatially filtered by a small Laplacian derivation (see Section 2.6.3) to enhance SNR. This dataset can be used to evaluate machine learning algorithms for the multi-class self-pace BCI.

BCI competition IV dataset IIa [137]: These four-class (left hand, right hand, foot and tongue) data are recorded from nine health subjects with 22 EEG channels and 3 EOG channels. As many trials contain eye movement artifacts (detected by EOG channels), the main challenge for this dataset is to classify of EEG signals affected by ocular artifacts. The problem can be solved by applying an on-line EOG correction algorithm (see Section 2.7.1) before extracting the features.

- **Bipolar recording datasets**:

  Bipolar recording (see Section 2.6.1) is a popular recording technique in BCI research, since it can improve the SNR and reduce the number of electrodes. BCI competition III dataset IIIb [214] and BCI competition IV dataset IIb [100] contain three bipolar channels data with two different classes (left vs. right hand), which can serve for developing methods based on bipolar recording (see Section 2.6.1). BCI competition III dataset IIIb are recorded from three subjects. The main challenge of this dataset is the non-stationarity problem. BCI competition IV dataset IIb are recorded from nine subjects. The main challenge is to classify the data contaminated by ocular artifacts.

To sum up, most of datasets listed above are recorded from healthy people, which are suitable for testing new methods for BCI systems for large public (non-medical) applications. They pose different challenges, such as channel (electrode) reduction, artifact removal, etc. Many of them can be used to evaluate the robustness of methods to artifacts and data evolution. Moreover, a lot of researchers tested their methods on BCI competition data and presented their results on the competition websites or in the literature, thus allowing an easy comparison of methods. Although testing methods on self-recorded data might be also interesting, evaluation and comparison would be more difficult. However, most BCI competition data are recorded on a few subjects only (usually no more than nine subjects in a dataset), so testing methods on several different BCI competition datasets should be interesting for achieving a good evaluation.

2.5.2 Classification evaluation

Several criteria can be used for classification evaluation. This subsection presents three most popular criteria used in BCI research: confusion matrix, classification accuracy (ACC)/error rate (ERR) and kappa coefficient (κ).

- **Confusion matrix** [178]: For a M-class problem, the confusion matrix shows the relationship between the classification output and the true classes. An element c_{ij} in the confusion matrix indicates how many samples of class i have been found in class j by the classification algorithm. The diagonal elements c_{ii} represent the number of correctly classified samples for class i. The off-diagonal elements c_{ij} $(i \neq j)$ represent how many samples of class i have been incorrectly classified in class j. The total number of samples is $\sum_{i=1}^{M}\sum_{j=1}^{M} c_{ij}$. The confusion matrix is seldom used in BCI evaluation, because it is difficult to compare two confusion matrices when the number of classes is larger than 2 [178].

- **Classification accuracy and error rate** [178]: The classification accuracy (ACC) and the error rate ($ERR = 1-ACC$) are the most popular evaluation criteria in BCI research. The classification accuracy is defined as the observed overall agreement between classification outputs and true classes. It can be derived from the confusion matrix:

$$ACC = \frac{\sum_{i=1}^{M} c_{ii}}{\sum_{i=1}^{M}\sum_{j=1}^{M} c_{ij}} \tag{2.1}$$

  As it is a statistical summary of agreement, the classification accuracy has some limitations: 1) it does not consider the chance level for agreement in the classification; 2) it cannot reflect the performance of a classifier for classes with few samples, since they have a small contribution to ACC. Thus, this criterion is often used when all classes have comparable numbers of samples.

- **Kappa coefficient** [47]: To overcome the limitations of the classification accuracy, researchers proposed to use kappa coefficient κ for BCI evaluation. The chance level for agreement P_e is defined as:

$$P_e = \frac{\sum_{i=1}^{M}(\sum_{j=1}^{M} c_{ji} \cdot \sum_{j=1}^{M} c_{ij})}{(\sum_{i=1}^{M}\sum_{j=1}^{M} c_{ij})^2} \tag{2.2}$$

  By denoting P_o the observed agreement between classification output and true classes ($P_o = ACC$). The kappa coefficient is defined as:

$$\kappa = (P_o - P_e)/(1 - P_e) = (ACC - P_e)/(1 - P_e) \tag{2.3}$$

  A larger κ value indicates a better classification performance. The kappa coefficient has several advantages: 1) it considers the effect of chance agreement, 2) classes with few samples get the same weight as classes with many samples. Thus, in BCI competition IV, the kappa coefficient was chosen instead of classification accuracy as evaluation criterion.

2.6 Signal acquisition and spatial filters

EEG signals are usually recorded against a common reference, so called monopolar recording [191]. The electrodes for recording each channel of EEG are called "active" electrodes [196]. They are placed according to the 10-20 recording system[11] (see Figure 2.3). The reference can be located at an ear lobe, mastoid or nose. The data are therefore reference-dependent for monopolar recording [154].

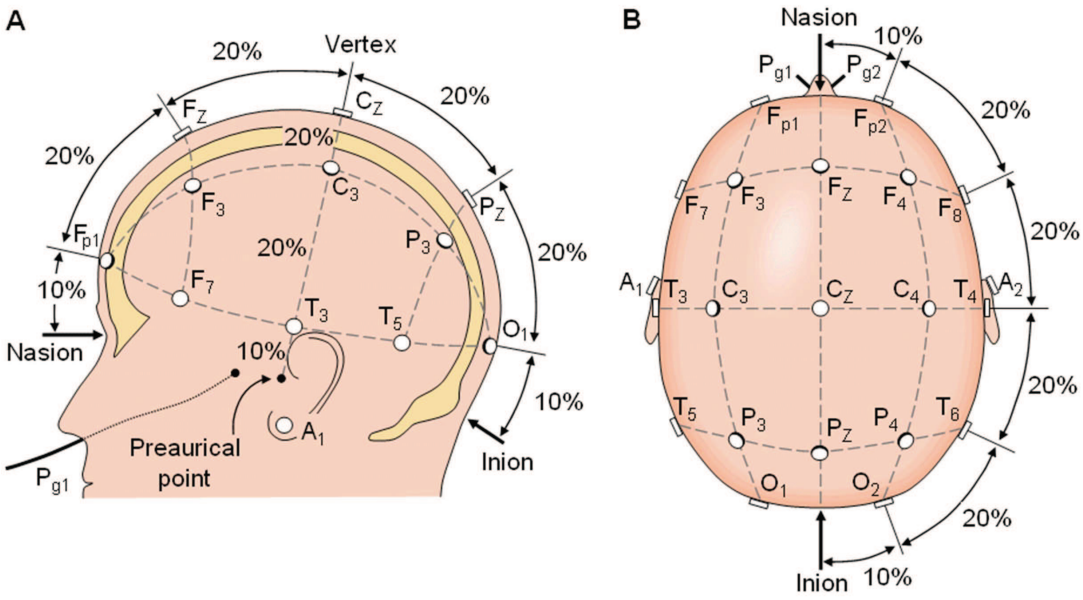


Fig. 2.3: International 10-20 system seen from (A) left and (B) above the head. A = Ear lobe, C = central, Pg = nasopharyngeal, P = parietal, F = frontal, Fp = frontal polar, O = occipital [113].

In the multi-channel setting, the EEG signals recorded at each time point can be considered as a spatial matrix corresponding to the spatial distribution of the electrodes. However, the spatial resolution of EEG is quite poor. Due to the volume conduction through the scalp, skull and other layers of the brain, the EEG signal recorded by a sensor is a "blurred" copy of multi-source activities (see Figure 2.4) [68, 138]. Spatial filters are used to improve the spatial resolution by enhancing the control signal and/or reducing the noise by using the neighborhood or global information [117]. This section reviews four popular spatial filtering methods: bipolar recording, common average reference, Laplacian derivation and common spatial pattern.

2.6.1 Bipolar recording

Bipolar recording is an EEG acquisition technique that measures the potential differences between two "active" electrodes. Bipolar recording can increase the SNR by reducing the common noise of both electrodes [108]. Let D_1 and D_2 be the discriminative patterns (e.g. motor-imagery related ERD/ERS patterns, event-related potential) of EEG signals at the

[11] The 10-20 system is an internationally recognized method to apply the location of scalp electrodes in an EEG experiment. This method was proposed to ensure standardized reproducibility so that comparisons can be made between subjects or over time for the same subject (for details, see [84]).

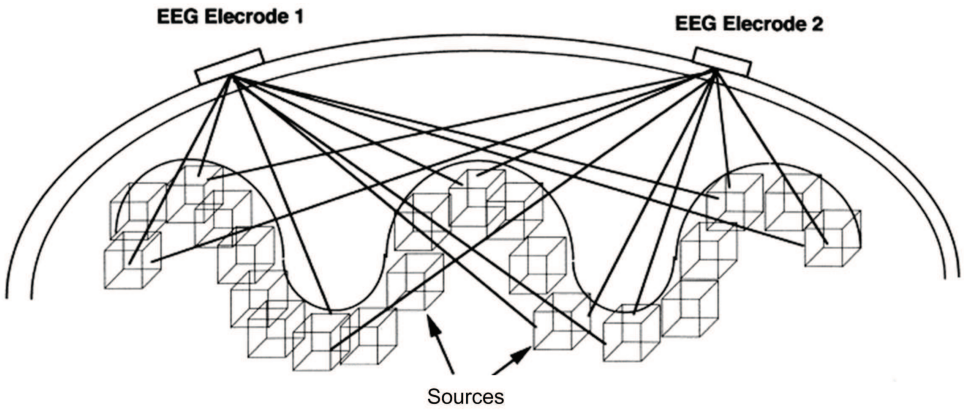


Fig. 2.4: Volume conduction effects in EEG recording (this figure is modified from [197]).

two "active" electrodes. Let C represent the mixture of common additive background noise of the two electrodes. The signals X_1 and X_2 recorded at the two electrodes can be expressed by the following additive model [216]:

$$X_1 = D_1 + C, \;\; X_2 = D_2 + C \tag{2.4}$$

This additive model is based on a basic assumption that a scalp EEG is a linear combination of source components. Although this assumption may not exactly reflect the complex composition of neural responses (which is still under exploration) in scalp EEG [11], it is typically used as an approximate model in practice for artifact reduction and source analysis [205, 212, 215]. Based on this model, the EEG signal in the bipolar recording channel derived from these two electrodes is:

$$X^{BI} = X_1 - X_2 = D1 - D2 \tag{2.5}$$

Thus, the bipolar recording can remove the shared additive background noise between two "active" electrodes.

2.6.2 Common average reference

In the common average reference method (CAR), the EEG signal is centered according to the following formula:

$$X_i^{CAR} = X_i - \frac{1}{N}\sum_{j=1}^{N} X_j \tag{2.6}$$

where X_i is the potential difference between the i-th "active" electrode and the common reference (i.e. EEG in the i-th monopolar channel), N is the number of "active" electrodes, and X_i^{CAR} is the corresponding EEG signal after the CAR transform. The CAR tends to result in a spatial voltage distribution with a zero mean, and therefore the CAR provides approximately reference-free EEG [154]. This setting reduces the components that exist in a large proportion of the electrode population, so as to serve as a high-pass spatial filter in EEG-based communication [117].

2.6.3 Laplacian derivation

Laplacian derivation is also a widely used spatial filtering algorithm in BCI [13, 78, 136]. The Laplacian value of each channel is obtained by subtracting the mean activity at surrounding electrodes from the channel of interest (see Figure 2.5) [75, 189]. Denote by X_i^{LAP} the Laplacian EEG signal in channel i, and S_i an index set of the four electrodes surrounding the i-th electrode. The Laplacian derivation is computed according to the following formula:

$$X_i^{LAP} = X_i - \frac{1}{4}\sum_{j \in S_i} X_j \tag{2.7}$$

Two different sets of surrounding 4 electrodes are often used: nearest-neighbor elec-

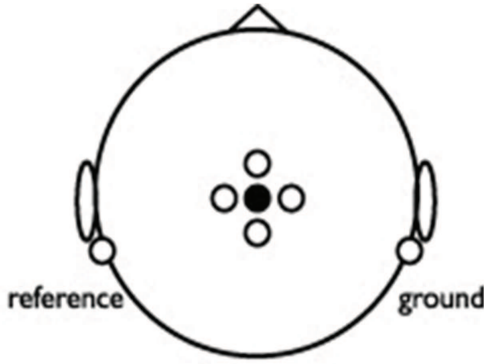


Fig. 2.5: Laplacian derivation: • represents electrode i, and four ∘ around electrode i represent four surrounding electrodes (they have the same distances from electrode i).

trodes (the distance from each surrounding electrode j to the center i is $d_{ij} = 3cm$) and next-nearest-neighbor electrodes ($d_{ij} = 6cm$) [117]. The Laplacian derivation employing nearest-neighbor electrodes is called a small Laplacian, while the one using next-nearest-neighbor electrodes is named a large Laplacian. The characteristics of the Laplacian highly depend on d_{ij}. Experiments show that the Laplacian becomes more sensitive to the components with high spatial frequencies when d_{ij} increases [117]. Thus, the small Laplacian is often used to improve the SNR when the control signal is highly localized and stable over time, while the large Laplacian is more suitable when the signal is not highly localized and non-stationary. Compared to CAR, the Laplacian derivation uses local instead of global information to obtain a reference-free EEG. This method can be used in any BCI system which has limited electrodes to record EEG data.

2.6.4 Common spatial pattern

Unlike spatial filters reviewed above, common spatial pattern (CSP) is a data-driven approach for constructing subject-specific spatial filters. It aims at maximizing the difference between the classes (denoted L and R in the following) through finding directions (i.e. spatial filters) that maximize the variance of signals for one class and at the same time minimize the variance of signals for the other class [37]. The basic strategy of the CSP algorithm is to simultaneously diagonalize the covariance matrices of the two classes. Let X_L and X_R be the $(time \times channel)$ data matrices of the frequency filtered EEG signals

(concatenated trials) under the two conditions (e.g., right-hand or left-hand imagination, respectively). The corresponding covariance matrices C_L and C_R are estimated by:

$$C_L = X_L^T X_L, \quad C_R = X_R^T X_R. \tag{2.8}$$

The spatial filters $W = [w_1, ..., w_N]$ are constructed as follows:

$$W^T C_L W = \Lambda_L, \quad W^T C_R W = \Lambda_R, \quad \Lambda_L + \Lambda_R = I \tag{2.9}$$

Those filters W decompose the N-channel EEG $X = [x_1, x_2, ..., x_N]^T$ into N uncorrelated filtered signals $Z = [z_1, z_2, ..., z_N]^T$ through the transformation $z_j = w_j^T X$, $(j = 1, 2, ..., N)$. A toy example that shows how it works in 2D is given in Figure 2.6. In classification, the features of single-trials are obtained by the log-variance of CSP filtered signals z_j [135]. For more details, please refer to Chapter 4.

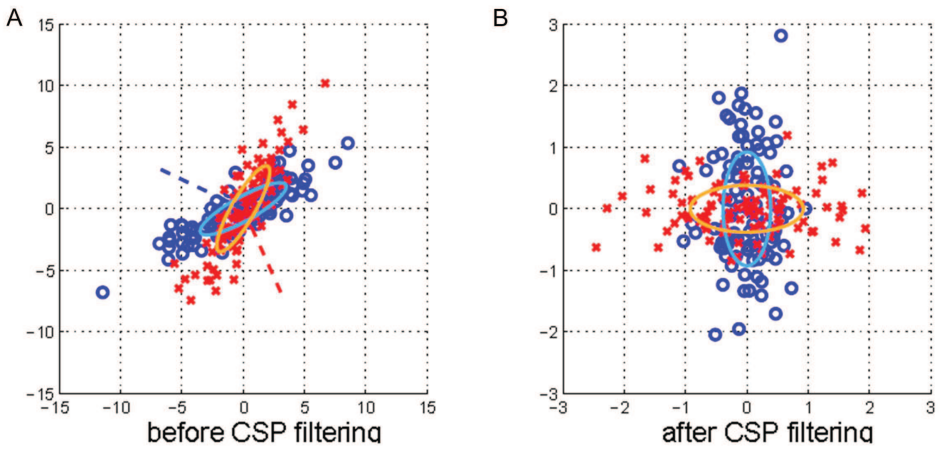


Fig. 2.6: A toy example of CSP filtering in 2D space [37]. Two sets of points marked by red ellipses and blue ellipses are from two Gaussian distributions (two classes). A. Distribution of samples before filtering. Ellipses show the estimated covariances of distributions and the dashed lines show the directions of the pair of CSP projections. B. Distribution of points after the filtering. Both classes are uncorrelated at the same time in this process. The horizontal axis gives the largest variance in the red class and the smallest in the blue class, while the vertical axis in the opposite.

2.6.5 Discussion on spatial filters

As recording EEG activities during a special BCI control may vary with different subjects, subject-specific data-driven spatial filters, such as CSP, are often recommended in BCI researches, in particular for motor imagery BCI [37, 68]. Figure 2.7 shows the spectra of left vs. right hand motor imagery after using different spatial filters on the same data [37]. In this example, the raw channel shows a spectral peak around 9Hz that provides no visible discrimination between two classes. The bipolar recording and CAR can only improve it slightly compared to the raw channel. Both the Laplacian derivation and CSP generate a second discriminative spectral peak around 12Hz. Compared to Laplacian derivation, CSP filters yield more remarkable discriminative peaks around both 9Hz and 12Hz.

However, the CSP algorithm usually requires a multi-channel setting (around 20 "active" electrodes) for a good performance [62]. Such a multi-channel setting inevitably reduces the portability and practicability of BCIs, which represents a main drawback for end-users. Recent researches showed that some feature extraction or classification methods using few electrodes with bipolar recording or Laplacian derivation may also achieve fairly good performances [12, 46, 108, 152]. Meanwhile, some improved CSP algorithms, such as Sparse CSP, are also proposed for reducing the number of electrodes in CSP-based classification [239]. Thus, a balance between better performances and less electrodes has to be achieved in the spatial filter selection. This aspect is discussed in Section 2.8.

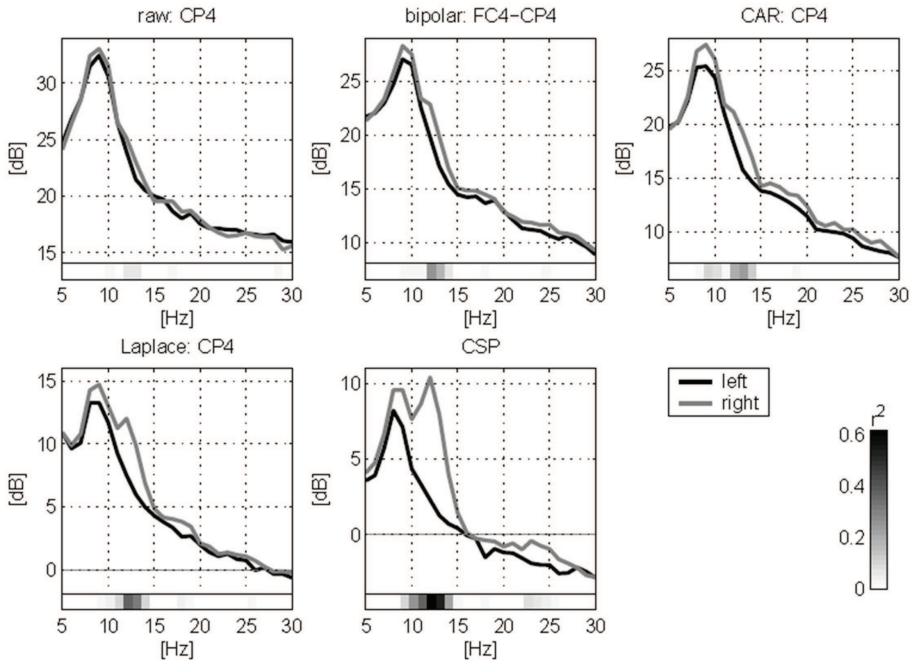


Fig. 2.7: Spectra of left vs. right hand motor imagery [37]. All plots are calculated from the same data but using different spatial filters. The discrimination between the two conditions is evaluated by the r^2-value (Equation 2.25).

2.7 Feature extraction and classification in motor imagery BCI

2.7.1 Pre-processing

In the EEG recording, the electrooculogram (EOG) signals from ocular activities are the main artifacts, which need to be removed before feature extraction. Various methods can be used for EOG reduction (for details, see [64]), but they are rarely fully automatic and online processing methods. Here, we describe a fully automated online method recommended by BCI competition IV [177]. The basic assumption of this method is that the recorded EEG data E are a superposition of the "pure" EEG data X and three spatial

EOG components O (horizontal, vertical and radial). So the "pure" EEG data X can be obtained by:

$$X = E - O \cdot b \tag{2.10}$$

where b is the correction coefficient. Denote by C_{OO} the autocorrelation matrix of the EOG components O, and by C_{EO} the cross-correlation between the recorded EEG data E and the EOG components O. The correction coefficient b is estimated by:

$$b = C_{OO}^{-1} C_{EO} = (O^T O)^{-1} O^T S \tag{2.11}$$

Figure 2.8 shows that most visible EOG artifacts in a EEG recording can be removed by using this online processing method. The independent evaluation of expert scorers demonstrated that this method can successfully reduce 80% EOG artifacts in raw EEG signals [177]. Although this method cannot completely eliminate EOG artifacts, it is the most popular method for online EOG correction, because it is fully automatic and available in a popular open source software library of biomedical signal processing – BioSig [176]. Recently, it has been discussed whether some artifacts, such as eye movement, can be used as useful control signals in BCI systems [21, 120]. This idea will be explored in Chapter 3.

Imaginary movements of a body part typically cause an EEG power modulation (ERD/ERS) at μ (8-12Hz) and/or β (12-30Hz) bands. Thus, a band-pass filter is usually applied to the "pure" EEG signals in the preprocessing after artifact reduction to extract the desirable frequency bands for feature extraction. For general use, a 8-30Hz band-pass filter (typically using a 5-order Butterworth filter) is recommended, since this frequency band covers both μ (8-12Hz) and β (12-30Hz) bands [135].

2.7.2 Popular features in motor imagery BCI

A great variety of features have been used for classification of signal in a motor imagery BCI [107]. Their properties may greatly affect the performances of the classifiers.

Band power

ERD/ERS patterns are the power decrease (ERD) and power increase (ERS) in relation to a reference interval (typically 1s) before the trial begins [154]. As the bandpass filtered EEG data in a short segment approximately follow a zero-mean Gaussian distribution [186], the power of signal is usually estimated by the variance of the signal values for each trial. In a multi-channel recording, denote by x_{jk} the j-th sample in the time interval $[t, t+T-1]$ of the k-th trial of the bandpass filtered EEG data in channel i, and $\bar{x_k}$ the mean value over all samples of filtered EEG in the time interval $[t, t+T-1]$ of the k-th trial. So we have:

$$\sigma_i^2(k) = \frac{1}{T-1} \sum_{j=t}^{t+T-1} (x_{jk} - \bar{x_k})^2 \tag{2.12}$$

However, $\sigma_i^2(k)$ is not normally distributed (it approximately follows a χ^2-distribution) [34]. Thus, the logarithm is applied here to make the distribution of BP features approximately

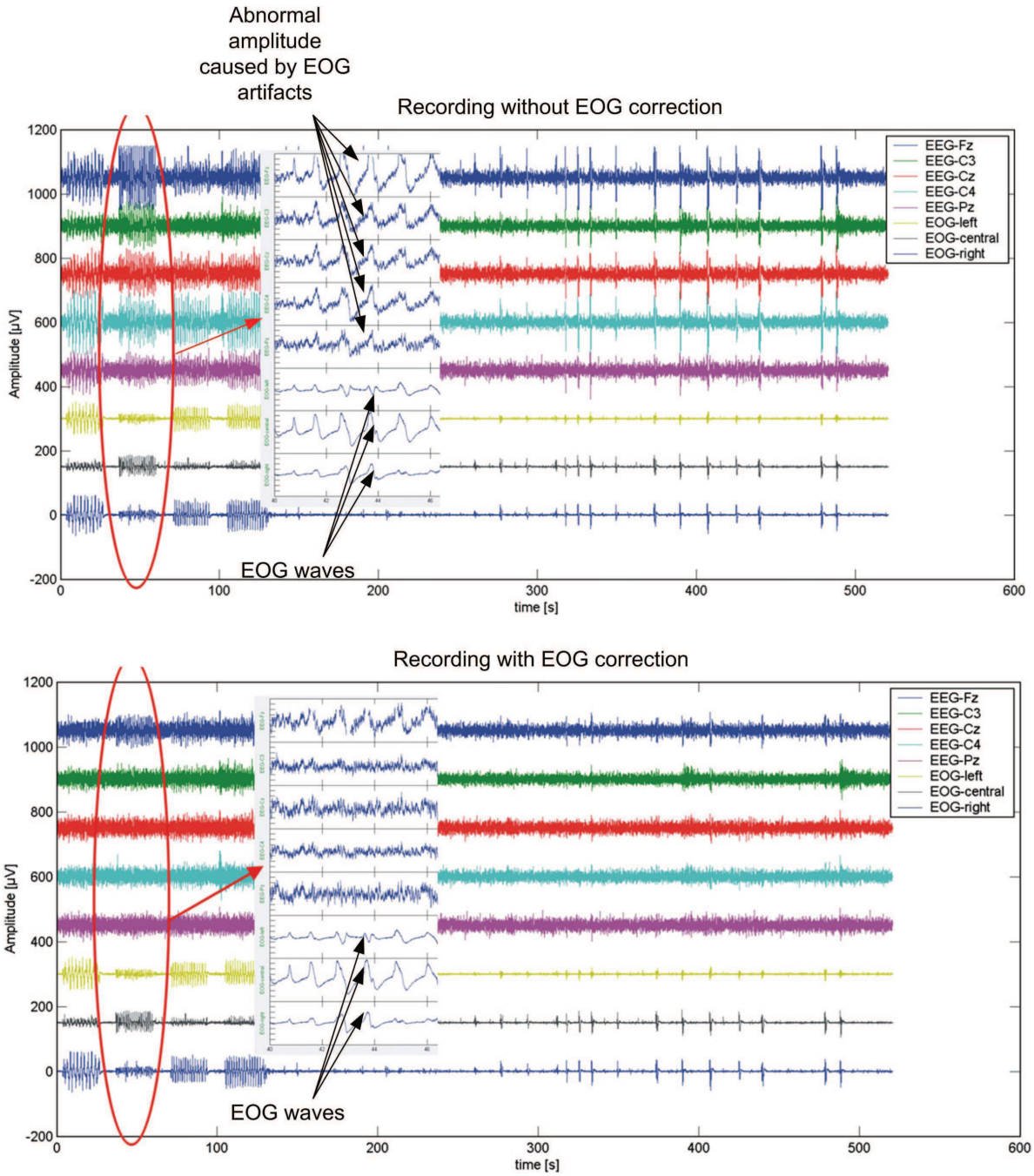


Fig. 2.8: Result of the online EOG correction methods (the figure is reproduced from [177]).

normal, which makes the linear classification more successful [213]:

$$BP_i(k) = \log(\sigma_i^2(k)) \tag{2.13}$$

$BP_i(k)$ is the k-th trial *log-BP* in channel i, which is called log band power (*log-BP*) estimator [213, 232].

Adaptive autoregressive parameters

Adaptive autoregressive (AAR) models are often used to describe non-stationary signals. In AAR models, AAR parameters are allowed to vary in time, so the model is defined as [81, 175]:

$$x_j = a_{1,j}x_{j-1} + a_{2,j}x_{j-2} + ... + a_{p,j}x_{j-p} + w_p \tag{2.14}$$

where x_j is the output sequence (i.e. j-th sample of EEG signal) and p is the model order. The coefficients $a_{1,j}, a_{2,j}, ..., a_{p,j}$ are time changing AAR parameters, which serve as data features in classification, and w_p is a white noise process with zero mean. AAR parameters are very useful for time-varying spectral analysis of EEG, so it is suitable for on-line BCI classification problems [175].

Time domain parameters

Time domain parameters (TDPs) are a set of broad band (i.e. 8-30Hz) EEG features defined in the time domain [213]:

$$TDP^{(p)} = \underset{t}{var}(\frac{d^p x(t)}{dt^p}), p = 0, 1, 2, ... \tag{2.15}$$

The range of t depends on the time segment used in feature extraction (for details, see Chapter 5). Note that the TDPs of order $p = 0$ is the *log-BP* feature we mentioned above. The TDPs of order $p = 0$ to 2 are related to Hjorth parameters [74], which are also often used as time domain features in the classification of EEG data [9, 145, 217]. Hjorth parameters contains three parameters: *Activity*, *Mobility* and *Complexity*. The relationships between Hjorth's parameters and TDPs of order $p = 0$ to 2 are:

$$Activity = \exp(TDP^{(0)}); \tag{2.16}$$

$$Mobility = \exp(TDP^{(1)} - TDP^{(0)})/2; \tag{2.17}$$

$$Complexity = \exp(TDP^{(2)} - TDP(^{(1)})/2; \tag{2.18}$$

Our experiment shows that Hjorth Parameters carry less information than the TDPs of order $p = 0$ to 2 in the discrimination between different motor imagery data except its first parameter, which is the band power (BP) feature (for details, see Chapter 5).

Feature properties and their impacts

Some critical properties of features have to be considered in a BCI system, since they may greatly affect the performances of the classifiers.

1. **Sensitivity to noise and outliers:** All features reviewed above are sensitive to noise and outliers [107]. As raw EEG data typically have very poor SNR, this property will unavoidably deteriorate the performance of the classifier, presenting a challenge to BCI data classification. As a result, artifact removal, spatial filtering and time-frequency selection are often needed before feature extraction.

2. **High dimensionality:** High dimensionality is a common property of all features, in particular for high-order AAR or TDP features, in multi-channel BCI systems. In practice, training datasets should have five to ten times as many trials as the dimensionality of features to guarantee a good performance of a classifier [83, 164]. Thanks to CSP, electrode reduction algorithms (see Section 2.8) and support vector machine (see Section 2.7.3), this challenge can be addressed in most cases now.

3. **Non-stationarity in long-term use:** BP and TDP features are non-stationary in long term use, because both of them are based on the variance of the signal. The variance of EEG signal tends to have large modulations in a long-term use period [230].

4. **Sensitivity to time-frequency region:** For BP and AAR features, only the ones extracted from time-frequency regions that contain signal patterns linked to specific motor imagery tasks are suitable to feed the classifier. Thus, the time-frequency optimization is required for achieving good performances of the classifier. This part will be detailed in the next subsection.

2.7.3 Classifiers

Many classification algorithms can be used in BCIs. They are grouped into five categories in [107]: linear classifiers, neural networks, nonlinear Bayesian classifiers, nearest neighbor classifiers and combination of multiple classifiers. The most popular ones are linear classifiers, since they use simple models but usually achieve fairly good results in practice [132]. As they have less free parameters to tune and are thus less prone to overfitting, linear classifiers are generally more robust than their non-linear counterparts when there is some noise in the data [132]. However, linear classifiers can fail in the presence of strong noise and outliers. Thus, artifact reduction is often required for BCI data. Note that some complex problems, such as separating the data of the relax state ("think of nothing") from active mental state (e.g. counting numbers, reading letters), may not be effectively solved by just using linear classifiers even after artifact reduction. In this case, nonlinear classifiers, such as support vector machine with a kernel, are recommended [66].

Here, we describe two important kinds of classifiers that have been used for BCI design: Fisher's Linear Discrimininant Analysis (LDA) and Support Vector Machine (SVM). Details about other, not very often used, methods can be found in [107].

Fisher's Linear Discriminant Analysis (LDA)

Let us assume that two classes of observations have means $\vec{\mu}^L$ and $\vec{\mu}^R$, and covariances Σ^L and Σ^R, then the separation between two classes in a linear projection on a direction $\vec{v}$ is defined as the ratio of the variance between classes to the variance within the class, and is typically calculated as:

$$FC(\vec{v}) = \frac{(\vec{v} \cdot (\vec{\mu}^L - \vec{\mu}^R))^2}{\vec{v}^T(\Sigma^L + \Sigma^R)\vec{v}} \tag{2.19}$$

The maximum separation occurs when

$$\vec{v} = (\Sigma^L - \Sigma^R)^{-1}(\vec{\mu}^L - \vec{\mu}^R) \tag{2.20}$$

Fisher's LDA separates the different classes of data by a hyperplane perpendicular to $\vec{v}$. This classifier usually works with BP and TDP features [37, 213]. Due to the assumption of normal distribution of data in this Fisher's LDA, it is necessary to make the features at least approximately follow a Gaussian distribution before classification. For motor imagery data, the logarithmic transformation can make the distribution of BP or TDP features approximately normal [213], so it is often applied on those features to feed a Fisher's LDA.

Support Vector Machine (SVM)

The SVM is often used in BCIs, and has shown very good performances in high-dimensional feature spaces, such as high-order AAR or TDPs [32]. The main idea of SVM is to find a discriminant hyperplane $u \cdot \hat{X} + b = 0$ (u is a weight vector and b is a bias term) with the largest possible margin (i.e. the distance of the hyperplane to the nearest training data points) to separate D-dimensional data $\hat{X} \in \mathbb{R}^{K \times D}$ into two classes. It can be realized by solving the following optimization problem:

$$\begin{aligned} &\underset{u \in \mathbb{R}^d}{min} \ \|u\|_2 + C \sum_{k=1}^{K} \xi_k^2 \\ &s.t. \ y_k(u \cdot \hat{X}_k + b) \geq 1 - \xi_k \ (k = 1, ..., K) \end{aligned} \tag{2.21}$$

where $y_k \in Y = \{L, R\}^K$ is the classification label, K is the number of training vectors (i.e. total number of training trials) in the data sets, ξ_k is a slack variable used to ensure that the problem has a solution in case the data are not linear separable. The margin is defined as $\gamma(\hat{X}, Y, C) = 1/\|u\|_2$, C is the regularization parameter that controls the trade-off between a low training error and a high margin. SVM is known to be insensitive to overfitting and high dimensionality problem. It can achieve good generalization performances with a small computational time. The original algorithm proposed by Vapnik in 1963 was a linear classifier [208]. However, nonlinear SVM can be created by applying the kernel trick to large margin hyperplane. Common kernels $k(\hat{X}_i, \hat{X}_j)$ include:

- polynomial:
$$k(\hat{X}_i, \hat{X}_j) = \gamma \cdot \hat{X}_i \cdot \hat{X}_j + c$$
- radial basis function (RBF):
$$k(\hat{X}_i, \hat{X}_j) = \exp(-\gamma \left\| \hat{X}_i - \hat{X}_j \right\|^2)$$
- sigmoid:
$$k(\hat{X}_i, \hat{X}_j) = \tanh(\gamma \cdot \hat{X}_i \cdot \hat{X}_j + c)$$

The sigmoid kernel is not positive semi-definite, which may lead to a problem in the training step (i.e. non convergence of the minimization) at least in theory (for details, see [105]). As a result, kernels generally used in BCI research are the polynomial and RBF kernels [66].

Generally speaking, linear SVM is more often used than SVM with a kernel in motor imagery BCIs, in particular for two-class problems. However, some complicated multi-class problem may need a SVM with a kernel to classify data that are not linearly separable [107].

2.7.4 Time-frequency optimization

As we mentioned above, one of the major problems in feature extraction is to find time-frequency regions that carry discriminative information. Here, we categorize time-frequency optimization methods into two classes: CSP-related and non-CSP-related methods. The reasons for that are: 1) CSP is a popular algorithm in feature extraction, which has many variants and extensions, so that CSP processing is a big family in machine learning approaches for BCI; 2) CSP has some important properties, so as to generate some distinct strategies in the time-frequency optimization, which we will discuss in the following subsection.

CSP-related methods

The CSP algorithm is widely used in decoding the spatial patterns of motor-imagery-related neural responses from EEG signals in motor imagery BCI [5]. It has several critical properties:

- It typically works with BP features, so that it is sensitive to outliers, time-frequency region and non-stationary problem. Its effectiveness depends on the frequency band and time segment of EEG signal used.
- The frequency band for best CSP performance may be different from the frequency band that contains visible discriminative information in raw EEG (see Figure 2.7). Thus, the time-frequency band directly selected by using the information in raw EEG may not work in a CSP-based classification.

Several approaches were proposed to select the optimal frequency band for the CSP algorithm [6, 61, 103, 106, 143, 232]. The typical CSP-related approaches include:

1. **Common spatio-spectral pattern (CSSP)**: The method embeds a time delay in CSP for individually adapted frequency filtering, so as to achieve a more robust machine learning procedure than CSP (for details, see [103]). An improved version of CSSP was also proposed and called Common Sparse Spectral-Spatial Pattern (CSSSP) [61]. This improved version employs simultaneous optimization of an arbitrary Finite Impulse Response (FIR) filter within the CSP. The main problem of those methods is that the obtained filter coefficients greatly depend on the initial points [143].
2. **Sub-band common spatial pattern (SBCSP)**: In this method, the EEG signals are first decomposed into sub-bands using a filter bank. The CSP is performed in each sub-band to obtain CSP features. The SBCSP features are then put into Linear Discriminant Analyzers (LDAs) to obtain scores which reflect the classification capability of each frequency band. Finally, the scores are alternatively fused by Recursive Band Elimination (RBE) or Meta-Classifier (MC) for decision making. Figure 2.9 shows the general scheme of this method (for details, see [143]).
    - The RBE is an iterative algorithm based on SVM. Let us consider the input vector $\hat{X}$ as a concatenation of scores of all sub-bands. SVM will separate $\hat{X}$ into

two classes by finding an optimal hyperplane $u \cdot \hat{X} + b = 0$ (see Section 2.7.3). At each iteration, the algorithm trains the SVM to get weight vector u and then removes one band with the smallest weight in u. This procedure is repeated I times, so as to get the desired number of remaining (selected) bands. The number of selected bands is determined empirically in this method.

- The MC is based on a Bayesian classifier. It transforms band scores into likelihood ratio values by the Bayesian classifier (for details, see [143]) and then puts the values into a SVM to get the classification result (see Figure 2.9).

The SBCSP algorithm yielded superior classification accuracy (mean classification error = 10.0% for RBE-based SBCSP, and mean classification error = 10.3% for MC-based SBCSP) compared to CSSP (mean classification error = 11.8%) and CSSSP (mean classification error = 12.5%) on a standard dataset (BCI competition III dataset IVa) [59].

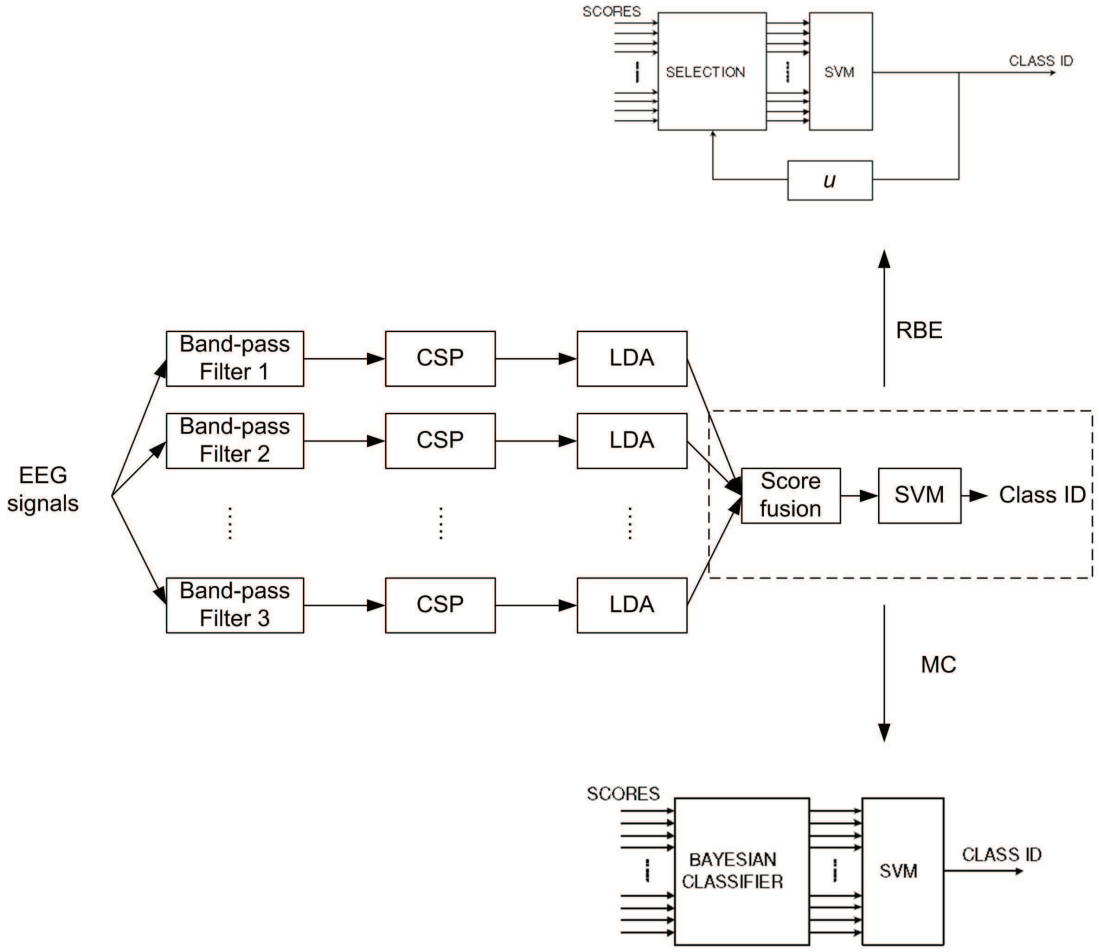


Fig. 2.9: General scheme of SBCSP [143].

3. **Filter band common spatial pattern**: Similar to SBCSP, EEG signals are decomposed into multiple frequency bands through a series of band-pass filters. Then, CSP features are extracted from each of these bands. A feature selection algorithm based on mutual information is employed to select discriminative CSP features for classification. The general scheme of FBCSP is shown in Figure 2.10 (for details,

see [6]). This method can be extended to address multi-class problems in BCI classification [5, 45], and yielded the best mean performance, compared to other submissions on BCI competition IV dataset IIa [39] and IIb [100], respectively. However, the FBCSP algorithm is sensitive to outliers, since it involves a classical estimation of multi-variate covariance matrices in calculating CSP features [7, 238]. A robust FBCSP algorithm was proposed to address this problem [7], which replaces the classical estimator with the robust Minimum Covariance Determinant (MCD) estimator [169].

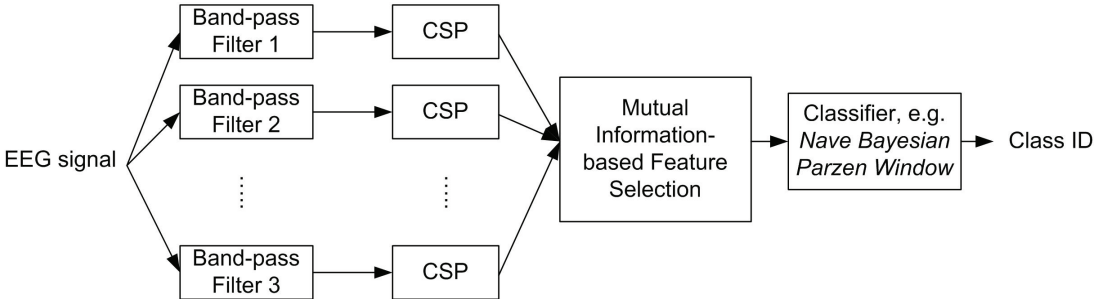


Fig. 2.10: General scheme of FBCSP [6].

Although these methods well address the issue of selecting an optimal frequency band for CSP-based classification, most of them select the time segment manually or heuristically [8]. Recently, an extended version of FBCSP including an optimal time segment selection process based on mutual information was proposed [8]. However, this extended version of FBSCP selects the time segment from only four different options (0.5-2.5s, 1.0-3.0s, 1.5-3.5s and 0.5-3.5s), which did not generate significant improvement on the BCI competition IV dataset IIb [100] compared to previous versions [8].

Non-CSP-related methods

Non-CSP-related time-frequency feature extraction approaches usually employ a simple spatial filter, such as Laplacian derivation, instead of CSP, to improve the SNR of the EEG signals [80, 218, 219, 234]. There are two main strategies in the literature for extracting the time-frequency features after spatial filtering:

1. **Weighting synthesis strategy** [218, 219, 234]: In this strategy, the EEG signals are first decomposed into a series of frequency bands. Then, the envelopes of EEG activities in a series of overlapping time segments are calculated for each frequency band. Those piecewise integrated envelopes form the spatial patterns in a time-frequency grid. Finally, weighted synthesizing of spatial patterns over time and frequency bins is used for decision making. The time-frequency weights are determined by a training process (for details, see [218]). This strategy was tested on 9 subjects for the discrimination between right and left hand motor imagery (the dataset is provided by Dr. Allen Osman from the University of Pennsylvania, for details, see [146]), its performance (ACC=90%) is better than CSP algorithm (ACC = 81%) with the same type of classifier [218].

2. **Divide-and-merge adaption** [80, 81]: The EEG signals are decomposed into a series of basic time-frequency cells which cover different frequency bands and time segments. The merge procedure can be summarized as follows:

   (a) Select a cell and construct a children-mother structure with other nearby cells (see Figure 2.11).

   (b) Define and calculate the discriminant power for each children and mother subspace.

   (c) Merge two neighboring children subspaces if their discriminant power is less than that of the corresponding mother subspace, so as to construct new time-frequency segmentations.

   Features (e.g. AAR parameters) are extracted from the constructed time-frequency segmentations, and then are ordered using a class separability criterion, e.g. Fisher's criterion. The top m features are selected for classification. The parameter m is determined by the user. This strategy was evaluated on BCI competition III dataset IVa [33]. Its mean classification accuracy (96.0%) was better than the one obtained by using the same types of feature and classifier but without time-frequency optimization (87.7%) [81].

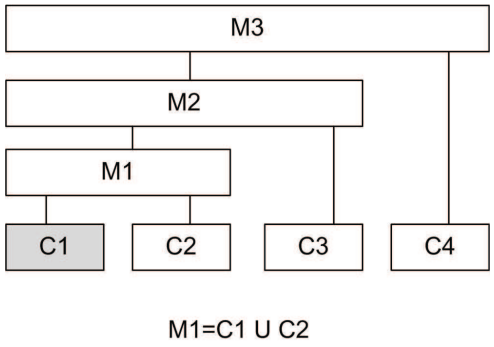


Fig. 2.11: Children-mother structure for the selected cell (marked in gray) [80].

2.8 Electrode reduction

As mentioned before, electrode reduction is an interesting topic in BCI research, particularly for bringing BCIs out of labs and hospitals into general public applications (e.g. entertainments, navigation). The problem of electrode reduction is usually solved in three ways: 1) reducing channels through CSP-based methods [124]; 2) selecting channels based on a specific classifier, such as support vector machine (SVM) [97]; 3) directly applying EEG acquisitions in a few specific areas which are in charge of the target motor intensions to achieve an acceptable accuracy rate of classification [108]. Here, we mainly review the methods of type (1) and (2), so as to provide a general view of how to algorithmically reduce the number of electrodes.

2.8.1 CSP-based methods

The CSP algorithm is widely used for constructing spatial filters for multi-channel EEG as described in Section 2.6.4. However, the performance of CSP-based classification is not proportional to the number of channels, since redundant electrodes may lead to overfitting (i.e. to learn the noise rather than useful discriminative information from the training data) and increase the computational time. Electrode reduction and selection based on CSP can be performed according to three different strategies: 1) constructing sparser spatial filters to reduce the number of channels required in CSP-based classification, such as in [62, 239], 2) selecting optional channels through CSP-based source localization methods, such as in [220], 3) combining CSP and other methods to choose an effective subset of channels in CSP-based classification, such as in [16, 17, 110, 124]. The typical approaches of these three strategies are briefly reviewed below.

Sparse spatial filter optimization

Let us denote by $w = [\omega_1, \omega_2, ..., \omega_N]^T$ a CSP spatial filter derived from N-channel data $X = [X_1, X_2, ...X_N]^T$. A filtered signal z is calculated as:

$$z = w^T X = \omega_1 X_1 + \omega_1 X_2 + ... + \omega_N X_N \tag{2.22}$$

The basic idea of this approach is to construct a sparser common spatial filter w in CSP so that fewer channels are required for the classification.

The classical CSP algorithm can be expressed as the following minimization problem:

$$\begin{aligned} &\min_{w} \; w^T C_L w \\ &subject\ to \;\; w^T (C_L + C_R) w = 1 \end{aligned} \tag{2.23}$$

where C_L and C_R are the covariance matrices of two classes [239]. Sparse CSP adds a $L1$ norm regularization term into Equation 2.23, which favors the sparsity of the solution [201]. The minimization problem is re-defined as follows:

$$\begin{aligned} &\min_{w} \; w^T C_L w + \rho \|w\|_1 \\ &subject\ to \;\; w^T (C_L + C_R) w = 1 \end{aligned} \tag{2.24}$$

This re-defined problem can be solved by sequential quadratic programming, augmented Lagrangian or conjugate gradient methods [30, 201].

In this method, the spatial filter w is obtained using all channels during the training phase. Only a few channels are needed in the testing phase or in real applications, since a lot of elements in w are 0. An essential parameter in sparse spatial filter optimization is the regularization parameter ρ. The vector w becomes sparser with the increase of the ρ value. However, there is a trade-off to be found between the regularization parameter and the classification accuracy. Thus, it is very important to choose an appropriate ρ value to keep a balance between the number of channels and the classification results. Experiments performed on the BCI competition III dataset IVa [33] show that the optimal ρ value varies between subjects [239]. For example, the optimal ρ is 0.02 for a subject, which reduces the number of channels from 118 to 17, with a slight dropping of classification accuracy from

58.6% to 57.5%, while the optimal ρ is 0.44 for another subject, which reduces the number of channels from 118 to 5, with an increase of classification accuracy from 52.5% to 54.4%. The optimal ρ value is manually selected in [239]. A possible solution to automatically determine the ρ value may be to test each possible ρ through a cross-validation procedure during the training phase for each subject.

CSP mapping-based channel selection

CSP algorithm decomposes the original signal $X = [X_1, X_2, ...X_N]^T$ onto a series of common bases $W = [w_1, w_2, ..., w_N]$ through the transformation $Z = W^T X$. The CSP filtered signal $Z = [z_1, z_2, ..., z_N]$ can be considered as uncorrelated EEG source components that include the common and task-specific components. Thus, the original EEG signal can be reconstructed through the inverse transformation $X = \Upsilon Z$, where $\Upsilon = [\upsilon_1, \upsilon_2, ..., \upsilon_N] = (W^T)^{-1}$. As the first and last columns of Υ correspond to the largest variance of one class and the smallest of the other one, these two vectors υ_1, υ_N reflect the spatial distribution of the most important task-specific components. The strategy proposed in [134, 220] selects the optimal channels by searching the maximum of the absolute value of the elements in the vectors υ_1, υ_N, respectively. Experiments on the BCI competition III dataset IVa [33] showed that this method can automatically select the optimal channels (i.e. 2 or 4 channels) from the 118 channel for linear classification. The mean performance is 87.42% when using 2 channels and 88.44% when using 4 channels.

Combination of CSP and other methods

- Criteria of scoring channel [124]

  (1) r^2-*value based criterion*: Denote $X = [X_1, X_2, ..., X_N]$ the N-channel training data, where X_i is a training data in channel i. The r^2-value of each channel is defined as:

  $$r^2(i) = \left(\frac{\sqrt{N_L N_R}}{N_L + N_R} \frac{\langle \|X_i\|_2 \rangle_L - \langle \|X_i\|_2 \rangle_R}{std(\|X_i\|_2)}\right)^2 \tag{2.25}$$

  where $\|\cdot\|_2$ is the $L2$ norm of vectors, $\langle \cdot \rangle_L$ and $\langle \cdot \rangle_R$ denote the mean of training data in class L and class R; $std(\cdot)$ is the standard deviation of all training data (regardless of the class), N_L, N_R denote the number of samples for class L and class R. Then, the CSP algorithm is applied to M ($M < N$) channels with the top r^2-values for CSP-based classification.

  (2) $L1$-*norm based criterion*: As for the r^2-value, the $L1$-norm can be used to give a score to each channel, which is different from the one used in sparse CSP. Let $\tilde{W}$ contain $2 \times m$ spatial filters $\tilde{w}_i$ derived from the entire N-channel EEG data. The $L1$-norm based score for each channel is calculated by:

  $$SC_{L1}(i) = \frac{\|\tilde{w}_i\|_1}{\left\|\tilde{W}\right\|_1} \tag{2.26}$$

  where $\|\cdot\|_1$ is the $L1$-norm. The M channels with the top score values are used in the CSP feature extraction step for classification.

M is pre-determined by the user in the method based on criteria of scoring channels. This contrasts with other methods where M is automatically determined, such as Sparse Spatial Filter Optimization. The authors in [124] set $M = 20$ in the experiment (using BCI competition III dataset IVa [33]) for comparing the performances of the two criteria. The experimental results showed that the $L1$-norm based criterion generated better mean performances (mean classification accuracy 89.68%) than the r^2-value based criterion (mean classification accuracy 87.24%), since the spatial patterns obtained by CSP with the $L1$-norm based channel selection coincide better with the neurophysiological plausibility (e.g. right foot imagery is characterized by enhanced amplitudes over the left hemisphere of the brain) [124].

- Riemannian distance based channel selection [17]

  The Riemannian distance between two covariance matrices C_L and C_R of two classes of EEG data is defined as:

  $$Dist(C_L, C_R) = \left\| Log(C_L^{-1} C_R) \right\|_F \tag{2.27}$$

  where $Log(\cdot)$ is the log-matrix operator and $\|\cdot\|_F$ is the Frobenius norm of the matrix. The channels in the reduced subset are chosen by maximizing the Riemannnian distance between two classes. In this method, the number of channels in the reduced subset should be pre-determined. In other words, this method only finds the optimal electrodes in the reduced subset for the following CSP-based classification, which is similar to the channel selection method based on criteria of scoring channels. Thus, the main drawback is that the optimal number of electrodes cannot be automatically determined by the channel selection method itself. Experiments on BCI competition III dataset IVa [33] showed that reducing the number of electrodes to 10 with the Riemannian distance based channel selection can still achieve 78% mean accuracy in classification.

- CSP-BPSO algorithm [110]

  BPSO is a discrete binary version of Particle Swarm Optimization (for details, see [91]). The CSP-BPSO algorithm first calculates the *Rayleigh coefficient*:

  $$R_c(i) = \frac{w_i^T (C_L - C_R) w_i}{w_i^T (C_L + C_R) w_i} \tag{2.28}$$

  and relies on this coefficient for the paired spatial filters (i.e. w_1, w_N), which are constructed by CSP algorithm from the entire N-channel data. The value $FD = |R_c(1)| + |R_c(N)|$ reflects the feature discrimination in the CSP-based classification. Assuming that M is the number of channels in the selected subset, the optimization of channel reduction is realized according to the following objective function, which should be maximized:

  $$J = FD + \beta(1 - M/N) \tag{2.29}$$

  where β is the trade-off coefficient between FD and the number of selected channels. If FD is large and M is small, then J will be large. However, FD may decrease when M is too small, so setting a right β value is very important in this method.

  The optimization is performed via a fitting process based on BPSO [91] for choosing M channels. In this process, channels are considered as a group of moving points.

Each moving point is named particle, and the entire group is called a swarm in BPSO. Each particle has two possible positions "0" or "1", "1" means the corresponding channel is selected, "0" means the corresponding channel is omitted.

Thus, the selection of channels is realized by the optimization of the position of each particle in BPSO, which can be written as:

$$v_k^{n+1} = \omega \cdot v_k^n + \lambda_1 \cdot r_1 \cdot (p_k^n - \hat{p}_k^n) + \lambda_2 \cdot r_1 \cdot (p_g^n - \hat{p}_k^n) \tag{2.30}$$

$$s(v) = (1 + e^{-v})^{-1} \tag{2.31}$$

$$if\ s(v_k^{n+1}) > \zeta\ \ then\ \hat{p}_k^{n+1} = 1, otherwise\ \hat{p}_k^{n+1} = 0 \tag{2.32}$$

where $\hat{p}_k^n$ denotes the k-th particle's position at the n-th iteration, v_k^n, p_k^n denote the velocity and the local best known position of the k-th particle in the n-th iteration, p_g^n is the best known position of the swarm at the n-th iteration, ω is a weight, λ_1 and λ_2 are the acceleration constants, and r_1, r_2 and ζ are random numbers uniformly distributed between 0 and 1. From Equation (2.30), we can see that the movements of the particles are guided by their own best known position p_k^n in the search-space as well as the entire swarm's best known position p_g^n. This is expected to move the whole swarm toward the best solutions (for details, see [110]).

The selected channels are used for feature extraction through CSP algorithm and then for the classification with a linear classifier. The selection of the β value is still a challenge in this method, which greatly affects the reduction and classification accuracy. Experiments on BCI competition III dataset I, which is a 64-channel electrocorticographic (ECoG) data [96], showed that when β takes a value between 0.4 and 0.8, using only 12 channels can lead to an accuracy above 90%.

2.8.2 Support vector channel selection

Based on SVM (see Section 2.7.3), two channel selection methods, Zero-Norm Optimization [97] and Recursive Channel Elimination (RCE) [180], were applied for motor imagery BCI.

- **Zero-Norm Optimization** The basic idea of this method is to try to obtain a sparser weight vector u in the optimization problem, which is similar to the Sparse CSP method. Based on the suggestion given in [223], $L0$-norm $\|u\|_0$ is used in Equation (2.21) instead of using $L2$-norm:

$$\min_{u \in \mathbb{R}^d} \|u\|_0 + C \sum_{i=1}^{n} \xi_k^2 \tag{2.33}$$
$$s.t.\ y_i(u \cdot x^{(k)} + b) \geq 1 - \xi_k\ (k = 1, ..., n)$$

  Its solution is usually much sparser than that of Equation (2.21), because of the use of the zero-mean. Here, the parameter C can be estimated from the training data (for details, see [97]), so that prior knowledge is not necessary in this method.

- **Recursive Channel Elimination** This method is based on the concept of margin maximization and derived from the recursive feature elimination method [71, 97,

180]. The process is similar to RBF (see Section 2.7.4), and performed through iteration steps.

At each iteration, it used data from the channel set to train a SVM. Let Ω be the inverse of the margin value:

$$\Omega(X, Y, C) = \frac{1}{\gamma(X, Y, C)} = \|u\|_2 \tag{2.34}$$

The importance of a channel (i.e. whether it can be removed or not) is determined by its contribution in maximizing the margin. Let F_i denote the features extracted from channel i. The score of channel i can be defined as:

$$SC_{RCE}(i) = \frac{1}{|F_k|} \sum_{l \in F_k} |u_l| \tag{2.35}$$

The channel with the lowest score is removed at each iteration. The parameter C can be estimated from the training data by finding the smallest classification error in the cross-validation (for details, see [97]). The iteration runs I times so as to get the desired number of channels, which is determined empirically in the literature.

Like most CSP-based methods, the main advantage of SVM-based methods is that electrode reduction and selection can be done robustly without any prior knowledge of neurophysiology (e.g. the spatial distribution of brain activities of a mental task). On the other hand, SVM-based methods provide us an idea to select task-relevant electrodes based on the performance of a classifier instead of only the intrinsic discriminative features of the signal. Experiments and comparisons provided in [97] showed that the optimal number of channels in the reduced subset is 17 with 23% mean error by using the RCE method. The mean error slightly increases when the number reduces to 12 but dramatically increases when the number is less than 12. Meanwhile, more than 17 channels in the subset also increase the mean error for some subjects. The reason is that only the channels that carry discriminative information are helpful for the classification. Adding channels without discriminative information amounts to put noise into useful data, so that it may deteriorate the performance of the classifier. Subsets of small size (less than 20 channels) selected by RCE generate better performance than those selected by Zero-Norm Optimization, showing that RCE is more effective (see Figure 2.12).

2.8.3 Discussion on electrode reduction

Generally speaking, the goal of electrode reduction is to decrease the number of electrodes participating in the classification. This challenge can be solved by 1) using similarity measures to find the electrodes that carry significant information; or 2) employing sparse representations to reduce necessary electrodes in the projections from the signal space to the feature space. Multiple factors, such as spatial filter, classifier, algorithmic parameters, might affect the performance of electrode reduction. But the most important thing is to keep a balance between classification results and the number of electrodes. This rule may finally determine the choice of the method and the corresponding parameters in real applications.

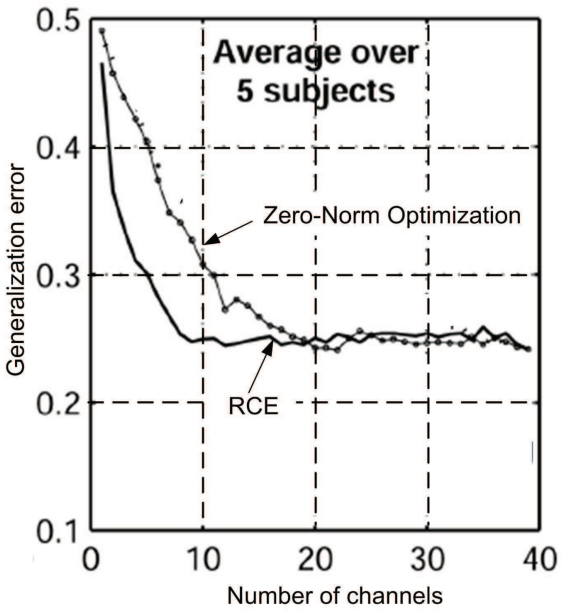


Fig. 2.12: Comparison between RCE and Zero-Norm Optimization (this figure is reproduced from [97])).

Methods based on similarity measures usually cannot directly tell users how many electrodes are required. They need users to decide the optimal number of electrodes based on their experience or on the results of classification. For a user who wants to achieve the best classification through a pre-determined number of electrodes (e.g. 4 electrodes), methods based on similarity measures should be a better solution than those based on sparse representations. Meanwhile, a trade-off between the number of electrodes and the classification results is possible for the methods based on similarity measures, because the user can try different numbers of electrodes to find acceptable classification results.

In contrary, methods based on sparse representations can automatically calculate how many electrodes are required. Thus, they might be a good solution when users do not know how many electrodes are necessary. These methods typically involve a parameter to find a trade-off between the number of electrodes and the classification results.

2.9 Discussion

The traditional BCI systems exclusively employ brain signals to realize the communication and control, aiming to provide assitive technology for disabled people. With several decades of evolution, BCI techniques have shown a relative maturity, which leads to some new understanding on BCI definition and applications. Hybrid and self-paced BCIs provide users more freedom in BCI experiences, which brings BCI to a broad range of BCI applications for the healthy users as well. Several types of EEG-based hybrid BCIs, such as Neurosky's Mindset, Emotiv, etc., have gone out of labs to be daily use products for entertainment.

However, the number of electrodes presents a key challenge for daily use BCIs. A large number of electrodes 1) increases the time cost in skin preparation and helmet installation, 2) reduces the portability of BCIs, 3) increases the price of system, 4) may deteriorate

the classification by a high dimensional feature space. Although dry sensors can address the problem of skin preparation and helmet installation, they are very sensitive to noise and can only be used in a well-controlled environment. CSP algorithms can reduce the dimensionality of features and improve the classification performance, but they cannot address other problems caused by the large number of electrodes.

Recently many algorithms and methods have been proposed to realize the electrode reduction. However, fine characterizations of subject-specific parameters are still needed to find a trade-off between the number of electrodes and the performance of the classifiers. The time-frequency optimization is a key step in the subject-specific parameterization for improving the classification performance under a given number of electrodes. However, most of time-frequency selection methods are based on multi-channel setting. Thus, new methods based on few electrodes are needed to identify subject-specific time-frequency patterns. On the other hand, the existing electrode reduction methods still have some drawbacks. Methods based on similarity measures cannot automatically find the optimal trade-off between the number of electrodes and BCI performance, while the ones based on sparse representations may not reduce the number of electrodes to the exact number demanded by users. Thus, advanced methods need to be proposed to meet automatical reduction and user needs. Moreover, the existing methods usually ignore the intrinsic relationship between subject-specific time-frequency parameters, and the number and positions of electrodes, simply expressing BCI features in one (or two) domain(s) while disregarding the other(s). As a result, the optimal combination of those factors may not be achieved in previous researches.

In this thesis, we will focus on developing an experimental setup and algorithms for BCI systems based on few electrodes for general public applications. A hybrid BCI based on one EEG electrode and a few EOG electrodes is proposed for healthy users. As the motor-imagery BCI is a popular one, which has not only applications in the entertainment but also in the industry usage, several novel algorithms are proposed to address the challenges of this type of BCI. Subject-specific parameterization is performed both in the CSP algorithm for reducing the dimensionality of features and number of channels, and in the time-frequency optimization for systems based on few electrodes, so as to find the optimal combination of subject-specific parameters for easy-to-use BCI systems.

Chapter 3

Turn "Artifacts" into Control Signals: A Hybrid BCI based on Alpha Rhythm and Ocular Activities for Human-computer Interaction

3.1 Introduction

As we have mentioned in Chapter 2 Section 2.2, conventional BCIs are mono-modal and externally-paced, only using one kind of brain signal to realize a computer-driven communication. This control mode raises several challenges for moving BCI systems from the laboratory to home, for example, (1) how to overcome the deficits (e.g. artifacts, mental fatigue) of only one type of brain activity (e.g. motor imagery, cue-driven neural response) and no physiological activities (e.g. eye movements, heart rate) engaged in the communication [64, 225]; (2) could BCI techniques be combined with other existing assistive techniques to utilize the advantages of different approaches and provide users with a more flexible way in the human-computer interaction [126]; (3) how to give users more freedom in choosing the time and speed of communication [173].

A possible solution to the challenge (2) and partly to challenge (1) is the concept of hybrid BCI (hBCI), which we have mentioned in Chapter 2 Section 2.2.2. For challenge (3), the control mode of a self-paced BCI, which we have introduced in Chapter 2 Section 2.2.3, could be a nice answer. Based on these two modern BCI concepts, we propose in this chapter a self-paced hybrid BCI utilizing the multi-modal fusion of user's mental states and ocular activities, and present a control mode, which might provide a potential solution to all the three challenges listed above.

Physiological activities (e.g. ocular activities, heart rate) are the main sources of artifacts in the conventional BCI, since they could trigger communication, which is not based on brain activities [64]. Although the hBCI concept allows muscle activities to be engaged in the human-computer interaction as an on/off switch or a complement to traditional BCI channel [101, 151], there is still no clear answer to how to deal with those traditional "artifacts" (part of challenge (1)). Based on the TOBI improved concept of hBCI, the proposed hBCI no longer considers ocular activities as "artifacts"; on the contrary, it uses them as main control signals instead as an on/off switch in the communication. In this case, it turns the "artifacts" into control signals.

This chapter is organized as follows. The background of utilizing alpha rhythm and ocular activities in BCI is given in Section 3.2. Section 3.3 gives an overview of the

proposed hBCI system. The signal acquisition and experimental setup are described in Section 3.4 and Section 3.5, respectively. Section 3.6 details the signal analysis and processing methodology, whose results are presented in Section 3.7. Finally, we discuss the potential applications of this hBCI and also our future work in Section 3.8.

3.2 Background

3.2.1 Alpha rhythm

Recognition of user's mental states, e.g. tiredness or attention level, could provide useful information to identify the intentional control or non-control states in a self-paced BCI system [125]. Alpha rhythm, also known as alpha wave or Berger's wave, is a kind of neural oscillation in the frequency range of 8-13 Hz arising from synchronous and coherent electrical activity of thalamic pacemaker cells in humans, which can be recorded either by EEG or magnetoencephalography (MEG) [19, 20]. As alpha rhythm is considered to be associated with the subject's mental states and cognitive operations, several BCIs have used alpha rhythm as a control signal in their system [89, 207]. A typically successful example is Neurosky's mindset[1], which analyzes alpha rhythm to estimate the user's "attention" and "mediation" levels for BCI control. Although many types of mental/cognitive-related information can be extracted from the alpha rhythm, the most typical and obvious behavior of alpha rhythm is the strongest oscillations generated in the occipital area during periods where eyes are closed, and attenuated signal by eye opening or increased attentiveness [148]. Thus, we propose to exploit this remarkable behavior of alpha rhythm to realize a self-paced on/off brain switch for the communication in the system.

3.2.2 Ocular activities

Voluntary ocular activities, which can be detected by an eye tracking system or electrooculographic (EOG) approach, have been used as a source of control signal for modern human-computer interactions [77, 82]. However, most BCI studies still consider ocular activities as a main source of physiological artifacts [64, 177, 225]. In this case, EOG electrodes placed around the user's eye are aimed at removing the EOG signals from the raw EEG signals instead of recording EOG for the communication purpose [177]. Although some hBCI studies point out that ocular activities could be used as a source of control signal, the approach is only limited to using eye tracking system but does not exploit EOG techniques [4, 151]. So at least two different recording systems, eye tracking system and brain signal recording system (e.g. EEG recording system), are needed in such hBCI, which, to some extent, increases the hardware cost and complexity of the hBCI system.

In this study, ocular activities are measured with EOG electrodes and transferred into control signal to fulfill a four-direction object movement function. Thus, only one recording system, EEG recording system, is needed in the proposed hBCI-based human-computer interaction. Moreover, as the EOG signal is much stronger than the EEG signal [215], the communication based on EOG signal should be more reliable and less vulnerable to physiological artifacts.

[1] www.neurosky.com/

3.3 System overview

The framework of the whole system is shown in Figure 3.1. In this system, both EEG and EOG data are recorded with one EEG recording system and then input to a PC for processing. Compared to traditional BCI systems, the basic steps of EEG data processing are the same, however, the modules of EOG data processing are added to transfer the intentional ocular activities into commands.

Firstly, the EEG data are processed to identify the mental state of the user. If active alpha rhythm is detected, which means that the state of the user is wakeful relaxation with closed eyes, the EOG data processing will not be triggered, and the whole system is in the sleep state. When the system detects that the active alpha rhythm disappears (i.e. becomes lower than a threshold), the system will be waked up and EOG data processing is triggered. The user can control the direction of the object (e.g. a ball or a wheelchair) movement in the real or virtual environment by saccades (simultaneous movements of both eyes). When the signal of a saccade is detected, the system will classify the direction of the saccade. Then, an audio message will be given to ask the user whether the detected direction is correct and intentional. The user confirms the message by blinking his eyes if it is right and intentional or just ignores it by keeping the eyes open until the message box disappears (3s after a saccade is identified) if it is wrong or undesired. All these steps of the procedure are detailed in the following sections.

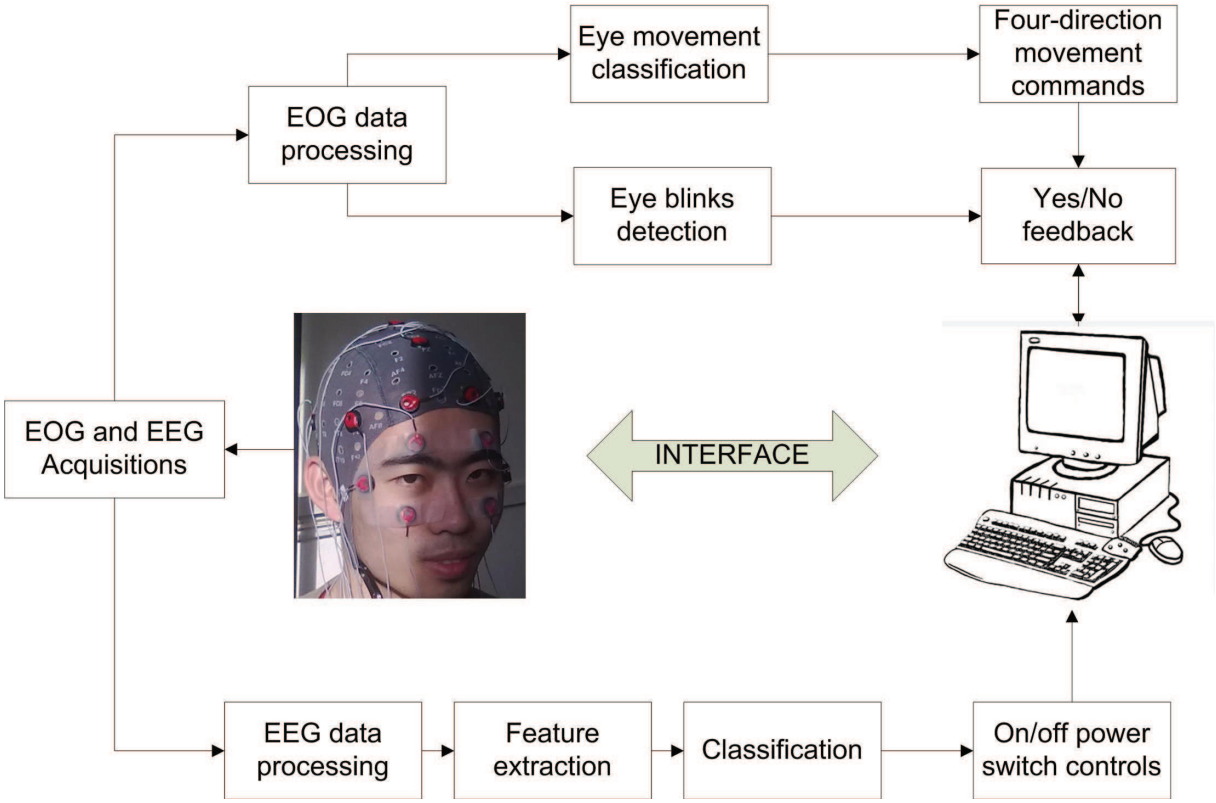


Fig. 3.1: Framework of the proposed hybrid BCI system.

3.4 Signal acquisition

To detect ocular activities, EOG data are recorded above (E1, E2) and below (E3, E4) the left and right eyes, respectively, and from the outer canthi of the left (E5) and right (E6) eyes (see Figure 3.2). This electrode placement scheme is the same as the one used in [49, 159], and the EOG recording is conducted according to the 2006 Clinical EOG standard from the International Society for Clinical Electrophysiology of Vision (ISCEV standard for EOG 2006) [38].

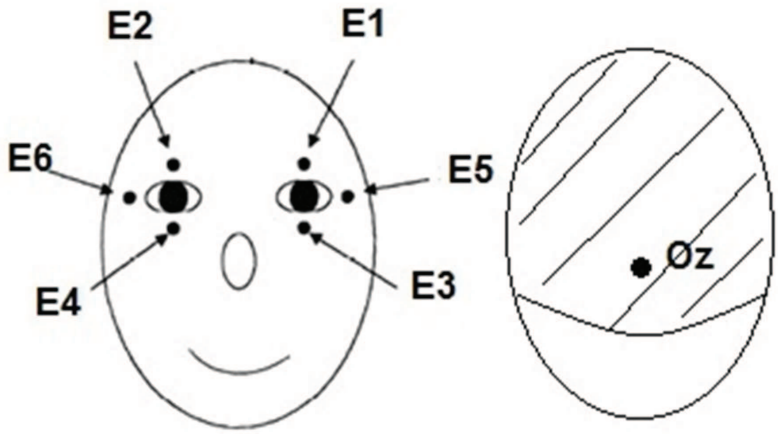
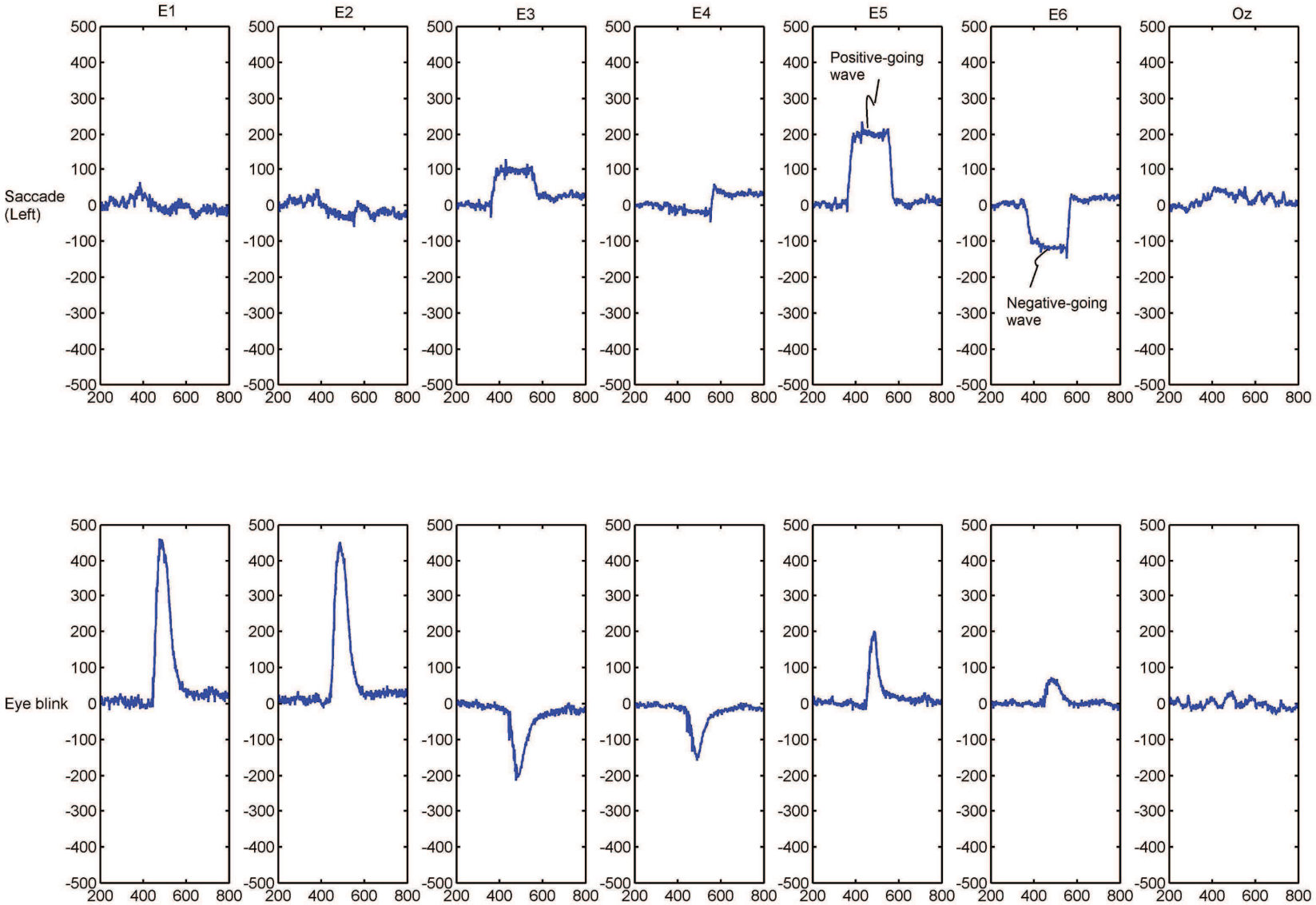


Fig. 3.2: Electrode placement scheme of EOG (left) and EEG (right).

As alpha rhythm predominantly originates in occipital cortex during wakeful relaxation with closed eyes [148], the EEG was recorded in Oz, the center of the occipital area. Both EOG and EEG data are acquired by g.tec EEG recording system[2] (sampling rate: 512Hz) using FCz as the ground and the right ear lobe as a reference.

Examples of original data recorded at each electrode are shown in Figure 3.3. In the first line of Figure 3.3, a left-going saccade evokes remarkable rectangle-like waveforms in E5 and E6, however, the polarity of the waves in these two horizontal electrodes are opposite. In the second line, the example of eye blink signal is presented. From this example, we can see that ocular activities are reflected on some EOG electrodes but have less affect on the EEG electrode Oz, which will be further examined in the next sections.

[2]produced by g.tec medical engineering, Schiedelberg, Austria, http://www.gtec.at/


E1
E2
E3
E4
E5
E6
Oz
Saccade
(Left)
Positive-going
wave
Negative-going
wave
Eye blink

3.5 Experimental setup

Considering the individual variability in ocular and brain activities, the experiments in this study included two sessions: learning session and testing session, for learning the subject-specific features (e.g. the activities of alpha rhythm, saccade amplitudes) and testing the detection methodologies, respectively. Experiments have been carried out on five individuals. None of them had a history of oculomotor pathology. All subjects in the study were entirely voluntary and signed a written informed consent (see Appendix A) before the experiments.

3.5.1 The learning session

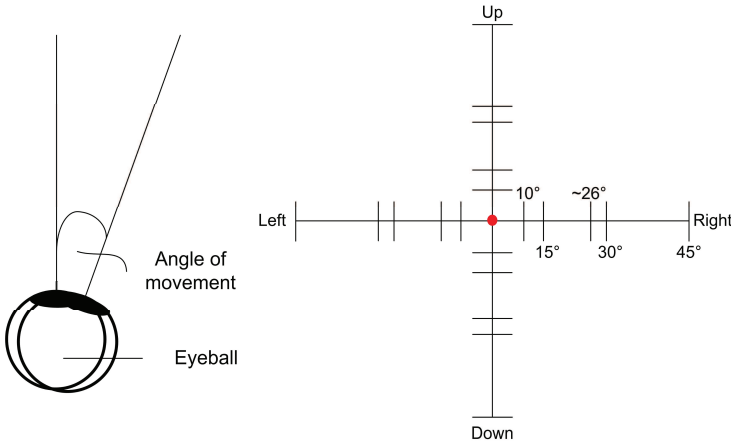


Fig. 3.4: Angle map of intentional saccades used in the experiments.

The data obtained in this section were used to learn for each subject the activities of alpha rhythm under different mental states (wakeful relaxations with closed eyes and open eyes) and the relationship between EOG data and different degree and direction saccades, and then to setup the subject-specific parameters in the methods (described in Section 3.6) for identifying different mental states and saccade controls.

Firstly, the subjects were required to be in a state of wakeful relaxation with closed eyes for 2 minutes and with open eyes for 1 minute. The EEG data recorded here were used to study the difference of alpha rhythm oscillations between closed eyes and open eyes.

Then, EOG data were acquired for learning the relationship between the features (polarity, amplitude) of EOG signal and the characteristics (direction, angle) of an intentional saccade, which was inspired by the work in [41, 142]. In the EOG data recording, four directions of eye movement were considered: left, right, up, down. For each direction, the subject was asked to make a saccade to focus on a target point displayed on a wall at 0.5m from the subject. The target points were marked corresponding to 10°, 15°, $\arctan\frac{1}{2}$ ($\approx 26°$), 30°, 45° from the central position in each direction. Each saccade was performed by quickly moving the eyes to the target points then going back to the center point. The saccade degrees chosen were similar to those in [142]. The difference was that the EOG data of 5° saccade was not recorded due to its small amplitude, which made

precise measurements difficult. The EOG data of $\arctan\frac{1}{2}$ degree saccade was recorded as a supplement to guarantee at least five samples of different degree saccades for estimating the relationship between EOG signal and saccades in each direction [142].

3.5.2 The testing section

In this section, the data were recorded for the same individuals to simulate a human-computer interaction procedure and to test the proposed detection methods and the parameters tuning based on the learning data.

As the half view angle (from the center to right/left edge) for a normal computer screen (17 inch) is close to 15°, the subject was required to perform 15° saccades in the testing section to make the result close to real applications. Different direction saccades were performed according to a sequence that was designed as a simulated object movement path shown in Figure 3.5 (direction sequence: right, down, left, down, right, up), which is inspired by the research about saccades during active visual search (for details, see [55]). The 15° saccades in this section are performed in the same way as in the learning section by quickly moving the eyes to the target points then going back to the center point. A voluntary eye blink was performed after each saccade to serve as the confirmation of the intentional saccade direction. The whole procedure is guided by audio instructions.

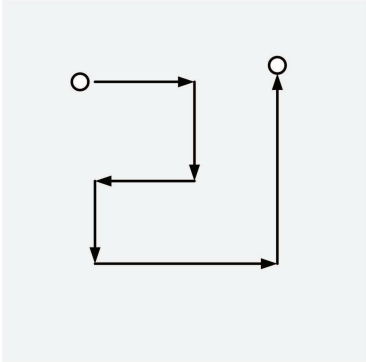


Fig. 3.5: Simulated movement path used in the experiment. The arrows only show the directions of saccades performed.

3.6 Signal analysis and processing

3.6.1 Alpha rhythm

Let x be the acquired EEG signal, which is sampled at frequency f_s. We denote by $x[i]$ its value at $i = t \times f_s$, where t is the time (in seconds). Let $\alpha[i]$ be alpha rhythm, which is extracted from EEG data through the 5-order butterworth band-pass (8-13Hz) filter [44]. The standard deviation σ of signal square $p = \alpha^2$ inside a sliding window (width $\omega = 4$ seconds), which reflects the dynamic changes of signal power, is calculated as a feature of interest to estimate the activity of alpha oscillation:

$$\sigma[i] = \sqrt{\langle p^2[j] \rangle_{N_j} - \langle p[j] \rangle_{N_j}^2}, \tag{3.1}$$

where $\langle \cdot \rangle_{N_j}$ is the average operator on $N_j = \{j \in \mathbb{N} \mid (t - \omega) \times f_s \leq j \leq t \times f_s\}$.

In this study, we analyzed the activities of alpha rhythm in two mental states using EEG data acquired in the learning section: wakeful relaxation with closed eyes and open eyes. Meanwhile, the EEG data acquired in the testing section were also processed using the same method to examine the potential effects of ocular activities on the activities of alpha rhythm.

As the alpha rhythm is more active during wakeful relaxation with closed eyes compared to that with open eyes (see Figure 3.10), a subject-specific threshold was applied to identify the mental state of subject in the proposed system. The user-specific threshold, T_α, was set to:

$$T_\alpha = \frac{1}{2}\left(\min_{i \in CE}(\sigma[i]_{close}) + \max_{i \in OE}(\sigma[i]_{open})\right) \tag{3.2}$$

where $\sigma[i]_{close}$ and $\sigma[i]_{open}$ represent the activity of alpha oscillation $\sigma[i]$ (defined in Equation 3.1) of learning data under the relaxation state with closed eyes (CE) and open eyes (OE), respectively.

In the real application, when the $\sigma[i]$ value is lower than T_α, the system considers that the active alpha rhythm disappeared, so the EOG data processing is triggered for detecting EOG-based control signals.

3.6.2 EOG data

From Figure 3.3, we can see that a specific ocular activity (such as a left-going saccade and an eye blink) can be directly reflected in the original signals of some EOG electrodes. In fact, an eye can be modeled as a dipole with its positive pole at the cornea and its negative pole at the retina, so an electric potential field is built using the eye as its source [23]. When the eyes or eyelids move, the potential field changes according to the modification of the poles. Let us take a left-going saccade as an example. When the eye moves from the center point toward the left periphery, the cornea approaches the left electrode (E5) and the retina approaches the right one (E6). This modification in dipole orientation causes a change in the electric potential field and is reflected in the corresponding EOG signal: positive-going wave in E5 and negative-going wave in E6 (see Figure 3.3). Thus, ocular activities can be monitored by vertical (i.e. E1/E3, E2/E4) and horizontal (i.e. E5/E6) pairs of EOG electrodes. By calculating the signal difference between the opposite electrodes in a pair, the vertical or horizontal eyes or eyelids movement components are easier to detect [41, 49, 51, 158].

Thus, two signal components: the vertical EOG (VEOG), defined in Equation 3.3, and horizontal EOG (HEOG), defined in Equation 3.4, are computed from original signals to monitor the vertical and horizontal movement components, respectively, for the saccade and eye-blink detections:

$$VEOG[i] = x_{E1}[i] - x_{E3}[i] + x_{E2}[i] - x_{E4}[i] \tag{3.3}$$

$$HEOG[i] = x_{E5}[i] - x_{E6}[i] \tag{3.4}$$

where $x_{Ek}, k = 1, 2, \cdots, 6$ is the original signal recorded in the EOG electrode placed at location Ek (see Figure 3.2).

The HEOG and VEOG signals of two typical ocular activities (the original signal in each EOG electrode, see Figure 3.3) are shown in Figure 3.6. A left-going saccade can easily be detected in the HEOG signal due to the horizontal movement of eyeballs, while an eye blink is more visible in the VEOG signal due to the vertical movement of eyelids.

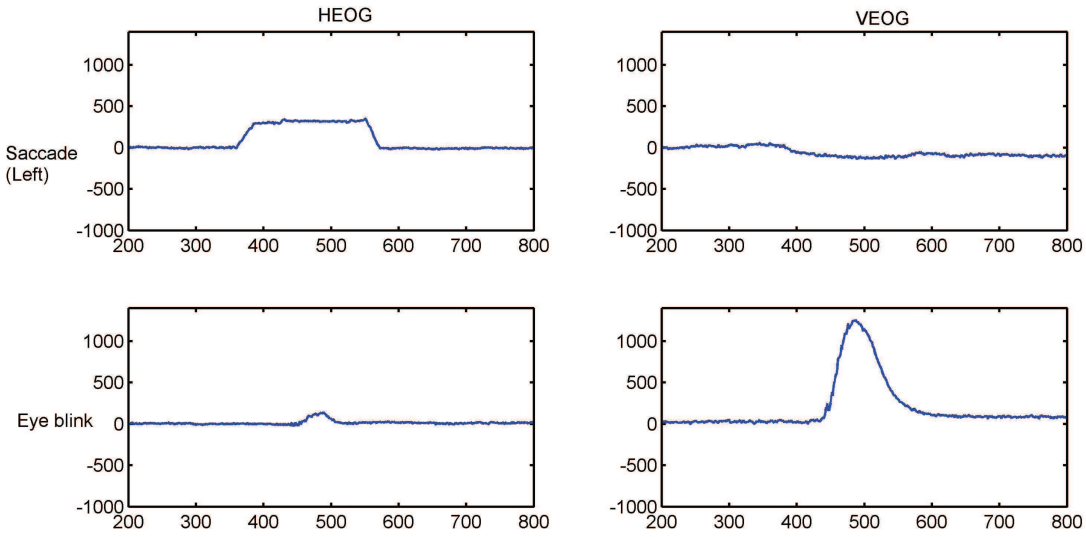


Fig. 3.6: HEOG and VEOG signals of two typical ocular activities: left-going saccade(mostly reflected in the HEOG singal) and eye blink (mostly reflected in the VEOG singal).

Saccades

To analyze the EOG signal of saccades in the learning section, the saccade amplitudes were measured according to ISCEV standard for EOG 2006 [38]. Paired t-tests were performed to check whether a statistically significant difference exists between the two same degree saccades in each direction for each subject.

As the directions and angles of saccades can be reflected in the amplitude and polarity of HEOG and VEOG signals (see Section 3.7), subject-direction-specific confidence intervals $(CI_d, d = left, right, up, down)$ of amplitudes were defined as $1 \pm 15\%$ of the saccade amplitude learned for detecting different direction saccades in our study. An algorithm is applied to process the HEOG and VEOG data sample by sample. When the value of a sample, $x[i]$, is within a confidence interval CI_d, one records the time point of the sample (marked as the beginning point, bp), the next samples are read until a sample out of the confidential interval is detected, and its time point is recorded (marked as the ending point, ep). We then check whether the duration between the beginning and end points, $dur = ep - bp$, is in the range of saccade duration, T, which is predetermined according to the user experience (150-600ms in this study). If yes, a d-direction saccade is detected. The flowchart of the algorithm is shown in Figure 3.7.

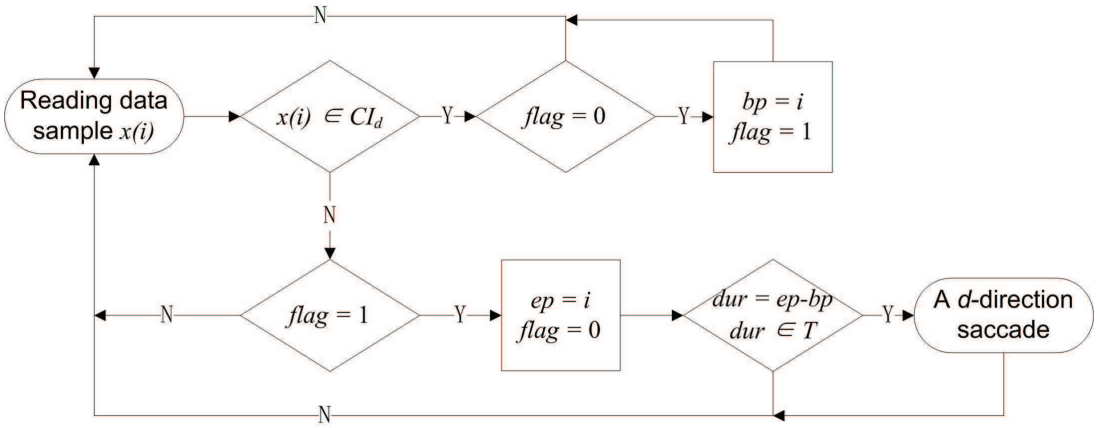


Fig. 3.7: Proposed algorithm for saccade detection.

Eye blink

As a voluntary eye blink serves as the confirmation response of an intentional saccade, it should be detected. A voluntary eye blink can generate a distinguishable triangular waveform in the VEOG signal (see Figure 3.9 A). Thus, the algorithm processes the VEOG data using a 600ms sliding window to detect a triangular wave in the given time range (3s after a saccade is identified) and checks whether its peak value exceeds a predetermined threshold (550μV in this study). The main steps of the algorithm are shown in Figure 3.8. The data scanned in the window are firstly smoothed by a Gaussian low-pass filter (σ = 3) (see Figure 3.9 B). The smoothing process eliminates the glitches in the triangular waveform. Figure 3.9 shows that the time derivative of the smoothed signal has three key points (the positive extremum, the negative extremum and zero crossing point between them). As a result, a triangular-like wave can be detected by detecting these three points: 1) the positive extremum and the negative extremum symmetrically exist in the scanned window (the difference between their absolute values less than 10% of the average of their absolute values), 2) only one zero crossing point should exist between the positive extremum and the negative extremum. Then, the algorithm also checks whether the maximum of the smoothed signal exceeds the predetermined threshold (see Figure 3.9 B). Thus, a voluntary eye blink is detected by the algorithm through both waveform detection and peak checking of the VEOG signal.

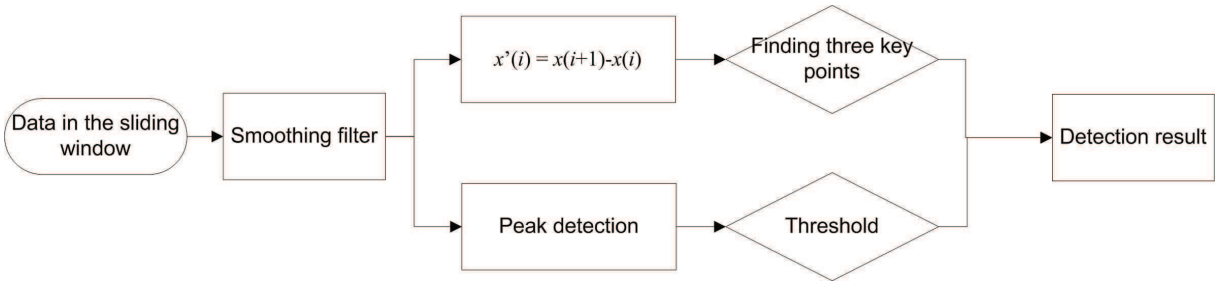


Fig. 3.8: Basic steps of eye blink detection.

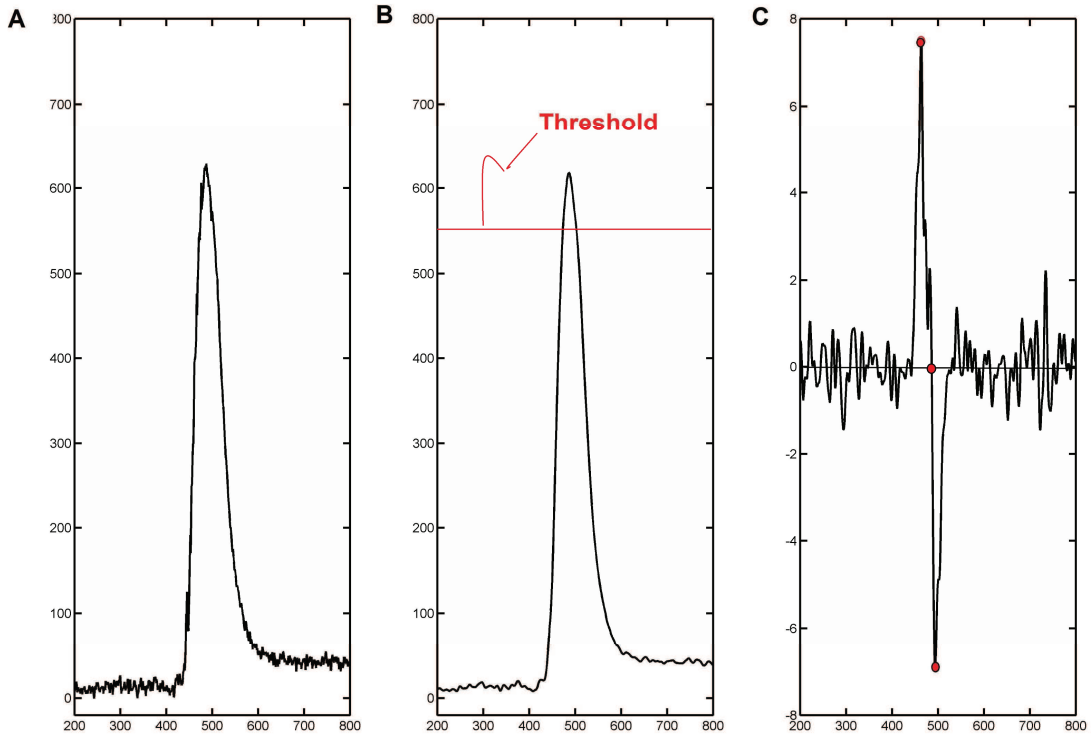


Fig. 3.9: Eye blink detection: A. Original VEOG signal, B. Smoothed signal and predetermined threshold, C. Derivative of the smoothed signal, with three key points marked in red.

3.7 Results

3.7.1 Alpha rhythm

Considering that the ocular activities may affect the alpha rhythm when performing EOG-based communication [209], the activities of alpha rhythm were examined among relaxation with closed eyes and open eyes in the learning section and the period of performing active EOG-based human-computer interaction in the testing section. Figure 3.10 shows alpha rhythms of relaxation with closed eyes (A) and open eyes (B), and alpha rhythm in the state of performing active EOG-based human-computer interaction (C). Relaxation with closed eyes generates stronger alpha oscillations than in other situations with open eyes, which resulted in higher value of $\sigma[i]$ (see Figure 3.10 D). Thus, the mental states with open eyes and closed eyes can be identified by applying a user-specific threshold to the real-time change value of $\sigma[i]$, which serves as a switch in the proposed system. In the experiments, the subject-specific threshold T_α estimated by Equation 3.2 (see Table 3.1) can successfully switch the system according to the subject's mental states with a few seconds delay (D).

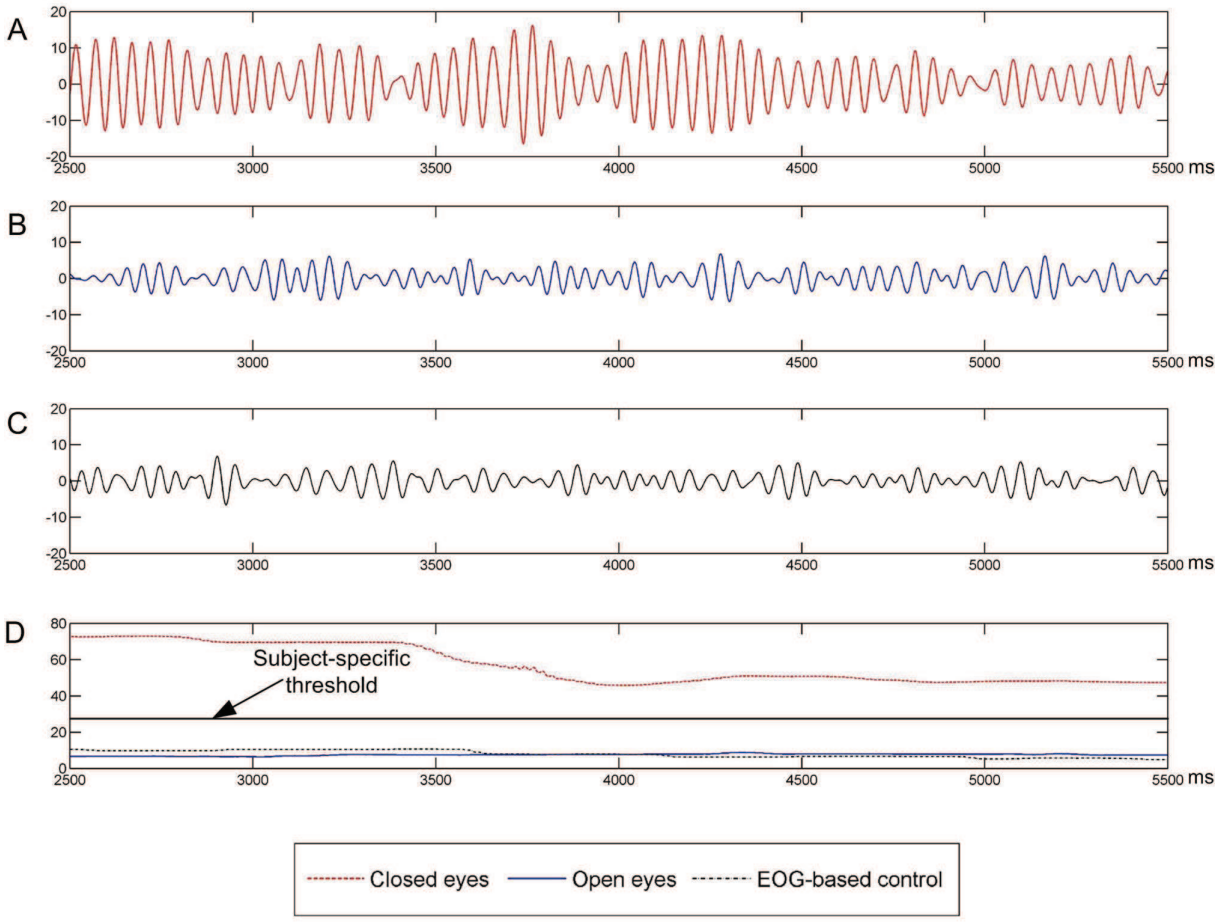


Fig. 3.10: Alpha rhythms at Oz: A. wakeful relaxation with closed eyes, B. wakeful relaxation with open eyes, C. EOG-based control in the testing section, D. corresponding $\sigma[i]$ values.

Table 3.1: User-specific threshold value T_α and delay D of each subject for separating the mental states with open eyes and closed eyes.

Subject	T_α	$D(s)$
1	31.1	2.11
2	27.0	2.07
3	58.8	1.81
4	16	2.58
5	9.9	3.13

3.7.2 EOG data

The examples of VEOG and HEOG data of the four direction saccades are shown in Figure 3.11. No significant difference of saccade amplitude was detected between the two same degree saccades in each direction for each subject (Paired t-tests: all $P \geq 0.2$). As the saccade amplitude is almost linearly increasing with the saccade angles and has

a considerable variability between different individuals/directions (see Figure 3.12), their relationship is modeled as a linear regression and described as:

$$\hat{y} = a\theta + b \tag{3.5}$$

where $\hat{y}$ is the estimated saccade amplitude, θ is saccade angle, a, b are subject-direction-specific parameters.

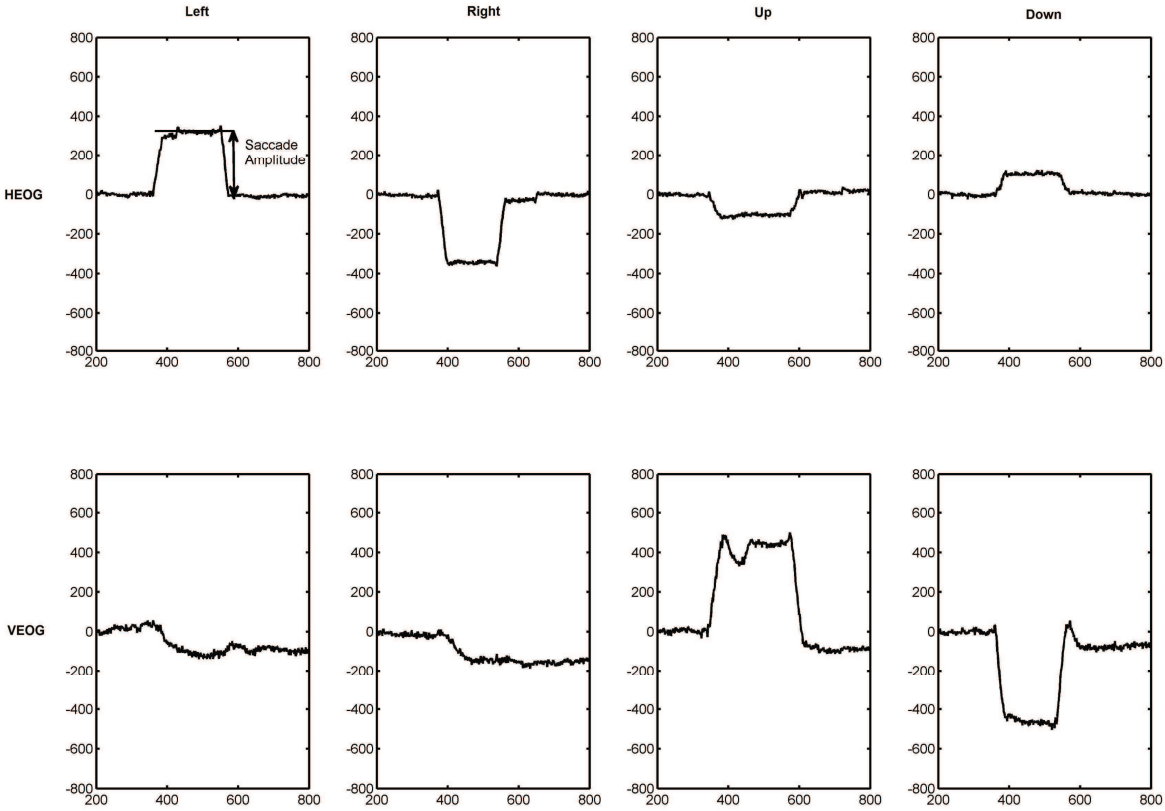


Fig. 3.11: Examples of HEOG and VEOG of saccades in the four directions (left, right, up and down).

The variability of the real data is measured through squared residuals with respect to the linear regression, SS_{err} and squared residuals with respect to the average value, SS_{tot} [139], which are defined as:

$$SS_{err} = \sum_{k=1}^{5} (y_k - \hat{y}_k)^2 \tag{3.6}$$

$$SS_{tot} = \sum_{k=1}^{5} (y_k - \bar{y})^2 \tag{3.7}$$

where the sum is taken over the set of subjects $k \in \{1, 2, 3, 4, 5\}$, y_k are the observed data in the study, $\hat{y}_k$ are the data estimated by Equation 3.5, and $\bar{y}$ is the mean value of the observed data.

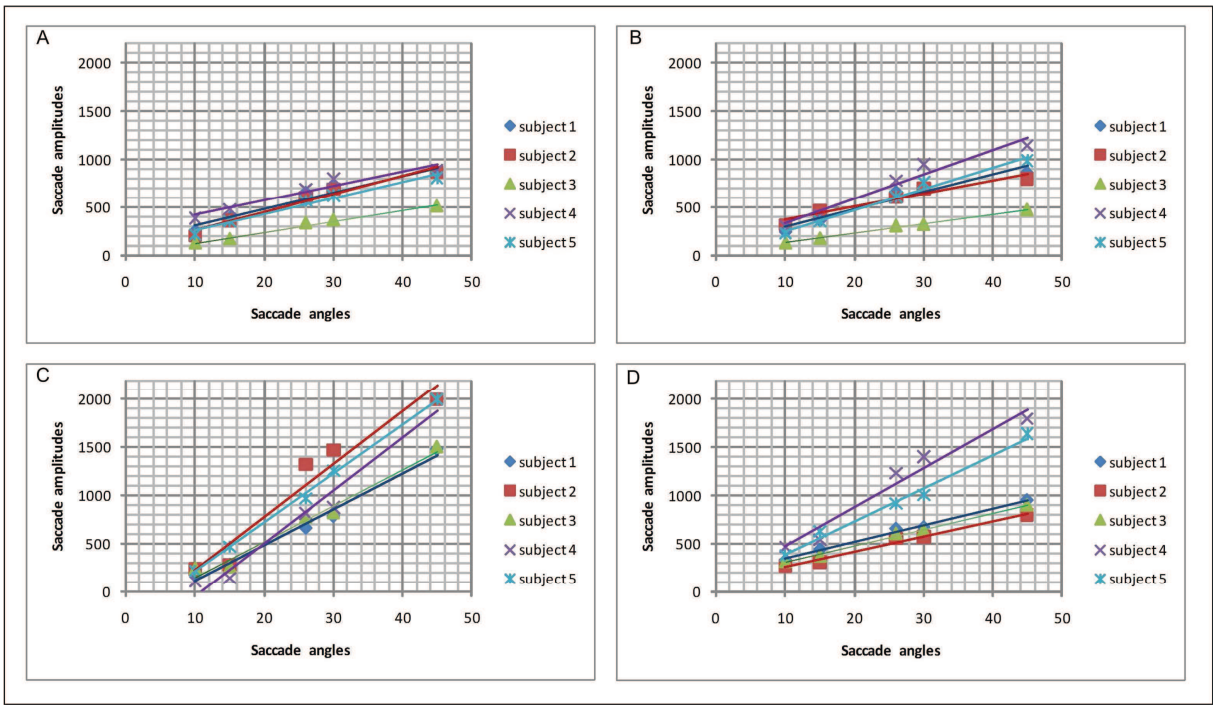


Fig. 3.12: The relationship between saccade amplitudes and angles in the four directions: left (A), right (B), up (C) and down (D). A linear relation is observed.

The coefficient of determination [139], R^2 is defined as:

$$R^2 = 1 - \frac{SS_{err}}{SS_{tot}} \tag{3.8}$$

It evaluates how well the regression line approximates the real data points. The value of R^2 ranges from 0 to 1.0. The closer to 1.0, the better the regression function fits the real data.The subject-direction-specific parameters and the coefficient of determination for each subject in each direction are listed in Table 3.2.

As all R^2 values in the study are close to 1.0, the relationship between saccade amplitudes and saccade degrees can be described by a linear model for each subject in each direction. This relationship can be used for setting the confidence interval of amplitude for detecting a specific degree saccade in the proposed system. On the other hand, it can also be used to deduce the saccade degrees from measured saccade amplitudes in EOG signals for human activity recognition [41] according to the following equation:

$$\hat{\theta} = my + n \tag{3.9}$$

where $\hat{\theta}$ is the estimated saccade degree, y is the saccade amplitude, m, n are subject-direction-specific parameters ($m = \frac{1}{a}, n = -\frac{b}{a}$).

In this study, the boundaries of the confidence interval for detecting a saccade were set to

Table 3.2: Subject-direction-specific parameters, a and b, and coefficient of determination, R^2, for each subject in each direction.

Direction	Subject	a	b	R^2
Left	1	17.32	137.9	0.9292
	2	18.84	73.11	0.9596
	3	11.37	17.19	0.9869
	4	15.22	268.3	0.9397
	5	16.65	98.42	0.9729
Right	1	18.13	117.2	0.9316
	2	13.53	235.7	0.9388
	3	9.708	39.94	0.9949
	4	25.31	83.39	0.9573
	5	22.12	32.41	0.9827
Up	1	37.16	-225.9	0.9844
	2	55.00	-304.5	0.9421
	3	37.13	-213.57	0.9782
	4	55.14	-602.7	0.9668
	5	51.06	-304.48	0.9982
Down	1	17.44	169.88	0.9920
	2	15.6	110.9	0.9858
	3	8.463	72.34	0.9946
	4	20.37	32.01	0.9602
	5	34.40	47.83	0.9861

$(1 \pm 15\%)$ of the saccade amplitude estimated by a linear function, which can successfully identify the intentional saccade-based controls in the testing data. The detection results are illustrated in Table 3.3. The failures of saccade detection are focused on the vertical directions due to the fact that HEOG signals are also sensitive to the movements of eyelids. Figure 3.13 shows HEOG and VEOG signal of an example of a non-detected saccade in this study. It is a down-going saccade. The subject performed non-voluntary eyelid movement during the saccade resulting in considerable noise in the HEOG signal, which finally leads to a failure of saccade detection.

Table 3.3: Saccade direction results for all subjects during the whole testing path. "+" means correct detection, "-" means wrong detection. TP represents true positive rate, FP represents false positive rate.

Subject	right	down	left	down	right	up	TP	FP
1	+	+	+	+	+	+	1.00	0.00
2	+	+	+	-	+	-	0.67	0.00
3	+	+	+	+	+	+	1.00	0.00
4	+	-	+	+	+	-	0.67	0.00
5	+	+	+	+	+	+	1.00	0.00

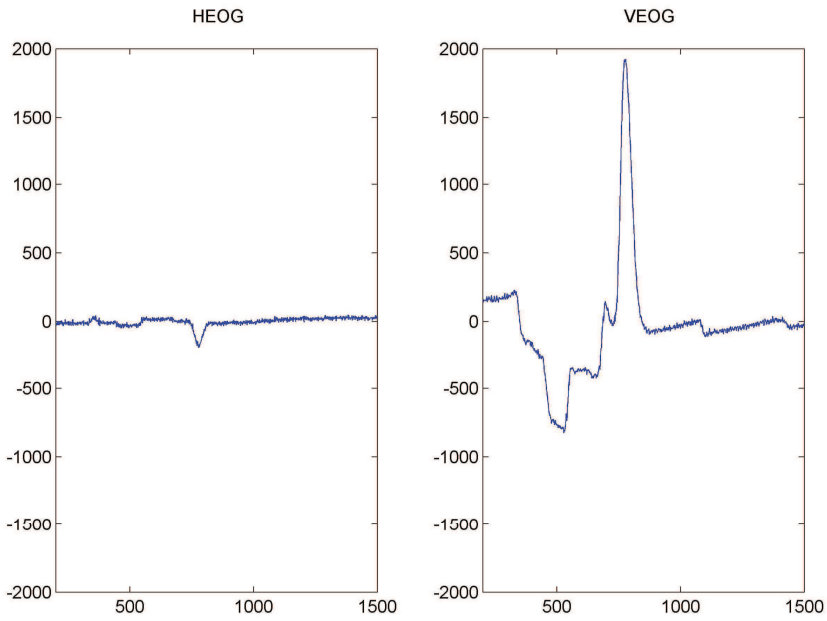


Fig. 3.13: HEOG and VEOG signals of an example of a non-detected saccade.

3.8 Discussion and conclusion

In the proposed hBCI, the user can use brain activity and eyes movements/blinks to realize human-computer interactions, where EOG serves as a control signal instead of being considered as artifacts; and the user can control the system whenever he wants just by opening his eyes. Thanks to the strong signal recorded by EOG electrodes, the features of different ocular activities are easy to detect. So the communication based on EOG is more reliable and less vulnerable to physiological artifacts than the traditional BCI channel. All of these provide possible solutions to the three challenges existing in traditional mono-modal and externally paced BCIs.

All the parameters in the study only need to be estimated once for each subject. All the algorithms applied in the system are on-line methods so that this hBCI technique can be applied to real-world applications. Only one EEG recording system with several electrodes was required in the system, so it is a single-system multi-function BCI, which saves the cost of hardware and also reduces the complexity of the whole system. Moreover, EEG/EOG electrodes are light weight and can be installed with a portable system, which provides a possibility of long-term use in a mobile real-world environment.

The linear relationship between saccade amplitudes and angles modeled in this chapter provides not only a useful reference for setting the parameters (e.g. confidential interval for detecting a saccade) in the proposed hBCI but also a possibility of using this hBCI for more advanced applications, such as a saccade-based speller (see Figure 3.14). In the possible saccade-based speller, the user can type different letters through performing saccades with different angles and directions, which may realize a similar function as P300 speller [183].

Moreover, this hBCI technique could be combined with traditional motor-imagery or P300 BCI technique to build a more powerful hBCI system. In this case, the user can

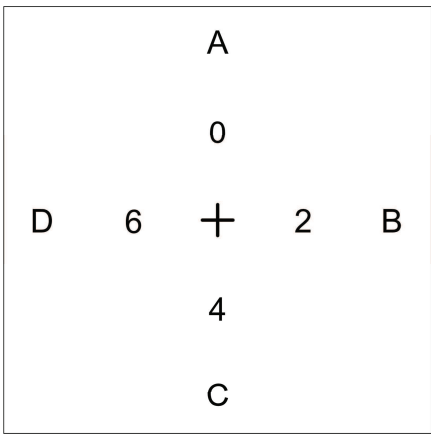


Fig. 3.14: A possible word matrix of a saccade-based speller.

choose to use the ocular activities to control the system when she or he is tired with using traditional BCI channel, where signals recorded by the EOG electrodes are served as control signals. Or the user can choose to use the motor-imagery or P300 as the sources of control when she or he feels tired with the communication based on ocular activities, where EOG electrodes are placed for automatic detection and reduction of EOG artifacts in the BCI channel [50].

In conclusion, this hBCI turns the main source of "artifacts", ocular activities, into a source of control in a hBCI system, giving a possible solution to the three challenges mentioned at the beginning of this chapter. Moreover, it also brings us a concept of single-system multi-function hBCIs, which can be applied to real-world applications, providing users with more freedom and less cost in BCI experiences.

Let us mention that vertical saccades are difficult to detect when the eyelids movement happen during the saccade. How to attenuate the effects from non-voluntary ocular activity is a challenge for this ocular-related hybrid BCI. Thus, one of the potential improvements is to develop a more reliable saccade detection algorithm to increase the robustness of the system in real applications. In the future, it may be interesting to use an eye-tracker in monitoring the ocular activity, so as to improve the quality and efficiency of ocular-based control.

On the other hand, combining this technique with more advanced mind-driven communication (e.g. motor imagery control) is also an important research direction, which will make the system more powerful and meet the diverse needs of a user. Thus, the part of EEG-rhythm based BCI technique, in particular of the popular motor imagery BCI and its challenges, will be further studied in the next chapters.

Chapter 4

Selection of CSP channels and filters for motor imagery BCIs

4.1 Introduction

Neuro-electrophysiologic studies reveal that both real movement and motor imagery of a specific body part typically induce an electro-encephalogram (EEG) rhythmic attenuation termed event-related desynchronization (ERD) in the μ (8-13Hz) and β (13-30Hz) bands over corresponding functional regions in the sensorimotor cortex ("active" brain areas), and may cause an EEG rhythmic enhancement termed event-related synchronization (ERS) over other regions in the sensorimotor cortex ("idling" brain areas) [154]. This phenomenon provides us the possibility to classify motor imagery EEG data, so as to give birth to motor imagery BCI systems. However, raw scalp EEG signals are known to have a poor spatial resolution due to the well-known volume conduction effect in EEG measuring (see Chapter 2 Section 2.6 and [37, 150]). In Chapter 2, we have briefly reviewed and compared several popular spatial filtering methods. They utilize either the neighborhood (like Laplacian derivation, bipolar recording) or global information (like common average reference, common spatial patterns). Among all those spatial filtering methods, common spatial patterns (CSP) algorithm [135] is believed to be the most effective one for extracting discriminative activity (i.e. ERD/ERS), because it is a data-driven approach, which can construct subject-specific optimal spatial filters to overcome the individual differences in EEG data. This algorithm was firstly proposed for a binary discrimination and then extended to multi-class problems through various approaches (for details, see [60]).

However, the performance of CSP usually depends on preprocessing (EEG time segment and frequency band selection), channel selection and number of paired filters. Until now, most studies focused on channel selection (see Chapter 2 Section 2.8.1) and improving a preprocessing procedure for CSP, in particular for selecting the optimal frequency band (see Chapter 2 Section 2.7.4), but not on selecting the number of paired filters in CSP. Most researchers choose the value of this number based on their experience and often use a constant value for all subjects, which ignores the potential individual differences. Although it was mentioned in [37] that this parameter can be alternatively determined via cross validation, this work neither provided any detail nor experimental validation. Moreover, using exhaustive searching strategy to find the optimal value of this parameter in the whole range of values increases the computational time, particularly when the dimension of data is very large. Thus, the method proposed in this chapter includes two steps: 1) a criterion based on *Rayleigh quotient* is applied for pre-selecting the range of the number of filters, 2) an algorithm based on cross validation is then employed for more precise estimation of the optimal value of this parameter in the pre-selected range. Based

on an existing four classes motor imagery dataset (BCI competition IV dataset IIa), we will verify the effectiveness of the proposed method by comparing the classification results using the estimated optimal values with those obtained using the recommended fixed value in existing work first in a popular two-class problem (left/right hand motor imagery discrimination) and then in the multi-class problem for classifying four different motor imagery tasks (left hand, right hand, foot and tongue).

Note that the CSP algorithm typically needs a multi-channel setting, which may present a key challenge for daily-use BCIs (see Chapter 2 Section 2.9). As we have reviewed in Chapter 2 Section 2.8, several machine learning algorithms are available to select several key channels, so as to reduce the number of electrodes in CSP-based classification. However, in practice, some researchers may simply reduce the number of electrodes based on prior knowledge of functional areas in brain, e.g. placing a few electrodes around the sensorimotor cortex, and then apply CSP algorithm with these reduced electrodes [138]. Compared to using machine learning algorithms for electrode reduction (see Chapter 2 Section 2.8.1), prior knowledge based electrode reduction usually does not require a full EEG cap or any additional computational time to find the optimal subset of channels. To explore whether it is an effective way to reduce electrodes in CSP-based classification, we also apply CSP with the estimated optimal number of paired spatial filters on three subsets of reduced channels, which are found based on the prior knowledge of functional areas in brain and recommended by the Graz BCI Lab [138]. The results are compared with those obtained by using a full EEG cap and a popular machine learning based channel selection algorithm, called recursive channel elimination [97].

4.2 Automatic selection of the number of spatial filters for motor-imagery BCI

4.2.1 Pre-selection of paired spatial filters

CSP is a data-driven approach to construct spatial filters, $W = [w_1, ..., w_N]$, which decomposes the N-channel EEG $X = [x_1, x_2, ..., x_N]^T$ into N uncorrelated filtered signals $Z = [z_1, z_2, ..., z_N]^T$ through the transformation $z_j = w_j^T X$, $(j = 1, 2, ..., N)$, so as to calculate CSP features:

$$f_j^{CSP} = \log(var(z_j)) = \log(var(w_j^T X)) \tag{4.1}$$

to feed a classifier (typically Fisher's LDA, see Chapter 2 Section 2.7.3).

In the transformation, w_j is a generalized eigenvector that satisfies:

$$\begin{aligned} w_j^T C^L w_j &= \lambda_j^L \\ w_j^T C^R w_j &= \lambda_j^R \\ \lambda_j^L + \lambda_j^R &= 1 \end{aligned} \tag{4.2}$$

where C^L, $C^R \in \mathbb{R}^{N \times N}$ are the estimated covariance matrices of two classes (i.e. 'L' and 'R') of N-channel EEG signals, respectively. Technically, this simultaneous diagonalization procedure can simply be achieved by solving the generalized eigenvalue problem:

$$C^L w_j = \lambda_j C^R w_j \tag{4.3}$$

where $\lambda_j = \lambda_j^L / \lambda_j^R$. Note that this decomposition transforms two classes of measurements into a common surrogate space. This method is therefore called Common Spatial Pattern (CSP) algorithm (for more details, see [92]). In this surrogate space, $\lambda_j^L \geq 0$ ($\lambda_j^R \geq 0$) is the variance of signal for class L (R). As $\lambda_j^L + \lambda_j^R = 1$, a large $\lambda_j^L (\lambda_j^R)$ indicates that the corresponding spatial filter w_j tends to yield a large variance of signal for class L (R) and a small variance of signal for the class R (L). These contrary effects of w_j on two classes decorrelate the signals, so as to contribute to the discrimination.

We sort the λ_j^L values in descending order. Thus, w_1 corresponds to the largest λ_j^L, and w_N to the largest λ_j^R (the smallest λ_j^L). They compose the first pair of CSP filters. Similarly, w_i and w_{N-i+1} corresponding to i-th largest λ^L and λ^R, respectively, compose the i-th pair of CSP filters.

We define the discriminative activity between two classes as $S_d = C^L - C^R$, and the common activity as $S_c = C^L + C^R$. Thus, the ratio between discriminative activity and common activity projected on the w_j spatial filter is the *Rayleigh quotient* $R(w_j)$ [37] and is obtained by:

$$R(w_j) = w_j^T S_d w_j / w_j^T S_c w_j = \left|\lambda_j^L - \lambda_j^R\right| / (\lambda_j^L + \lambda_j^R) = \left|2\lambda_j^L - 1\right| \tag{4.4}$$

For the i-th pair of filters, $FD(i) = R(w_i) + R(w_{N-i})$ reflects its effectiveness in extracting the discriminative components from the original signal [110]. Usually the first m pairs of spatial filters according to the m largest $FD(i)$ are used, since they tend to extract the discriminative components. The other pairs of spatial filters are not employed, since they tend to pick out the common components, which will deteriorate the classification results. A too small value of m cannot extract all discriminative components for classification. On the contrary, a too large value of m will introduce some common components. Both may deteriorate classification performances (see Section 4.4.1), so that the optimal value of m should be estimated for each subject. A too small $FD(i)$ (typically $FD(i) < 0.1$) indicates that the i-th pair of filters has a very weak ability of extracting discriminative components, and cannot improve classification results (see Section 4.4.1). As all paired filters are sorted in descending order of $FD(i)$, the $FD(i)$ values are used as a pre-selection criterion to shrink the range for seeking the optimal m value. Here, an empirical threshold $fd = 0.1$ is set to shrink the range of m in $[1, M]$, where M is the number of the paired spatial filters with $FD(i) \geq fd$. The effectiveness of this threshold is validated in Section 4.4.1

4.2.2 Refined estimation of the optimal number of paired filters

The optimality criterion for selecting paired spatial filters must satisfy two properties: (1) the number of paired spatial filters must be minimal; (2) it must yield the classification result that is equal or comparable to the best performance, i.e. such that there is no statistical difference between them or their difference is less than a tolerance δ. The tolerance δ is set by experience to avoid the overfitting problem[1] ($\delta = 0.015$ in this work). Here, the classification performances are evaluated via the *kappa coefficient*, κ (for details about *kappa coefficient*, see Chapter 2 Section 2.5.2).

The optimal number of paired spatial filters is evaluated by checking each possible m value ($m \leq M$) to see whether its corresponding κ value is significantly larger than

[1]Sometimes, using a large m may slightly improve the classification results in cross-validation but may deteriorate the classification results in an independent evaluation due to the overfitting [107].

others obtained for smaller values of m. The paired difference test (paired t-test, see Appendix B) is employed for the significance analysis [231]. If several m values yield equal or comparable classification results, the smallest one will be chosen. This procedure is summarized in **Algorithm A**.

Algorithm A: Selection of the optimal number of paired spatial filters

Let $m \leq M$; $\kappa(m)$ is a set of κ for a given m evaluated with a 100 repetitions of 10-fold cross-validation ($\kappa(m) \in \mathbb{R}^{100}$), $\bar{\kappa}(m)$ is the mean value (over the 100 components), $t(a, b)$ represents the p-value of the paired t-test between vectors a and b

1: $m_i \leftarrow 1$; $m_j \leftarrow 2$
2: **while** $m_j \leq M$ **do**
3: **if** $\bar{\kappa}(m_j) > \bar{\kappa}(m_i) + \delta$ **and** $t(\kappa(m_i), \kappa(m_j)) < 0.05$ **then**
4: $m_i \leftarrow m_j$
5: **endif**
6: $m_j \leftarrow m_j + 1$
7: **endwhile**
8: $m_{opt} \leftarrow m_i$
9: **return** the optimal parameter, m_{opt}

The optimal parameter m_{opt} is estimated off-line from the training data for each subject, and then applied to the testing data or on-line applications for the same subject. This strategy can be extended to multi-class problems using One Versus the Rest CSP [60, 233].

4.2.3 Extension to multi-class problems

One Versus the Rest (OVR) CSP is a multi-class CSP approach that computes W for each class against all the others, projects the EEG signals on all the chosen spatial filters to extract the features, and then performs a multi-class LDA classification [60]. For a P-class problem, the OVR CSP algorithm can be described as follows:

1. Estimate the covariance matrices $C^{(p)}$ for all classes ($p = 1, 2, ...P$).

2. Let us denote by $\mathscr{C}$ the set of all classes, and by $\mathscr{C} \setminus \{p\}$ the set of all classes except class p. The spatial filters, $W = [w_1, ..., w_N]$, for class p are obtained by the simultaneous diagonalization of $C^{(p)}$ and $C^{(\mathscr{C}\setminus\{p\})} = C^{(1)} + ... + C^{(p-1)} + C^{(p+1)} + ... + C^{(P)}$:

$$\begin{aligned} w_j^T C^{(p)} w_j &= \lambda_j^{(p)} \\ w_j^T C^{(\mathscr{C}\setminus\{p\})} w_j &= \lambda_j^{(\mathscr{C}\setminus\{p\})} \\ \lambda_j^{(p)} + \lambda_j^{(\mathscr{C}\setminus\{p\})} &= 1 \end{aligned} \tag{4.5}$$

3. Repeat the second step, so as to calculate W for each class.

4. Typically select m paired spatial filters from each W, so as to get $2m \times P$ spatial filters.

5. Project the EEG signals on all the chosen spatial filters to extract $2m \times P$ CSP features (defined in Equation 4.1) for classification.

Based on the pre-selection procedure in Section 4.2.1, each W generates a M value, thus $P \times M$ values are obtained. The largest M value (M_{max}) is chosen as the upper limit of possible m_{opt}. Then m_{opt} is estimated based on the classification results in the range of [1, M_{max}] using **Algorithm A**, so as to choose first m_{opt} paired filters from each W. Without loss of generality, the framework of our method for a four-class problem is illustrated in Figure 4.1.

Note that OVR is a popular strategy in extending two-class discrimination methods to multi-class problems. More discussion about OVR strategy is available in Chapter 7.

4.3 Data description and channel selection

The data used in this work are from BCI competition IV dataset IIa [39], which contains one training session and one testing session of 22-channel EEG data from 9 subjects who performed four classes cue-driven motor imagery (left hand, right hand, both feet and tongue). Each trial began with a fixation cross and an additional short acoustic warning tone. After two seconds, a cue in the form of an arrow pointing either to the left, right, down or up (corresponding to left hand, right hand, foot or tongue) appeared and stayed on the screen for 1.25s. The subjects were asked to carry out the motor imagery task until 4s after cue on-set. No feedback was provided. The experimental paradigm is illustrated in Figure 4.2.

The EEG signals were recorded by 22 Ag/AgCl electrodes (with inter-electrode distances of 3.5cm) using the left mastoid as reference and the right mastoid as ground (sampling rate 250Hz). The electrode montage is shown in Figure 4.3.

We used three subsets for finding the optimal subset of channels for this dataset based on prior knowledge of brain functional areas and their corresponding standard EEG recording positions (see Figure 4.4), and suggestions from the experts in Graz BCI Lab [138]. The first subset consists of channels No. 2 to 21, which use the entire scalp except frontal and occipital regions, since frontal and occipital regions are not involved in any motor imagery task. The second subset includes channels No. 2 to 18, covering mid-central and centro-parietal regions, which are known as sensorimotor areas. The third subset contains electrodes C3, Cz, C4, and electrodes surrounding them (marked in red, see Figure 4.3), 13 electrodes in total, covering the hand (C3, C4) and foot (Cz) representative areas. Although the third subset does not include the tongue representative area, it has been proved that it is possible to identify tongue motor imagery from this subset [138], because tongue motor imagery typically elicits an enhancement of EEG rhythms (ERS) in hand representation areas [152].

4.4 Experimental results

First of all, we use the full channel data to verify the effectiveness of our proposed method in selecting the number of paired filters for both two-class (left vs. right hands) and

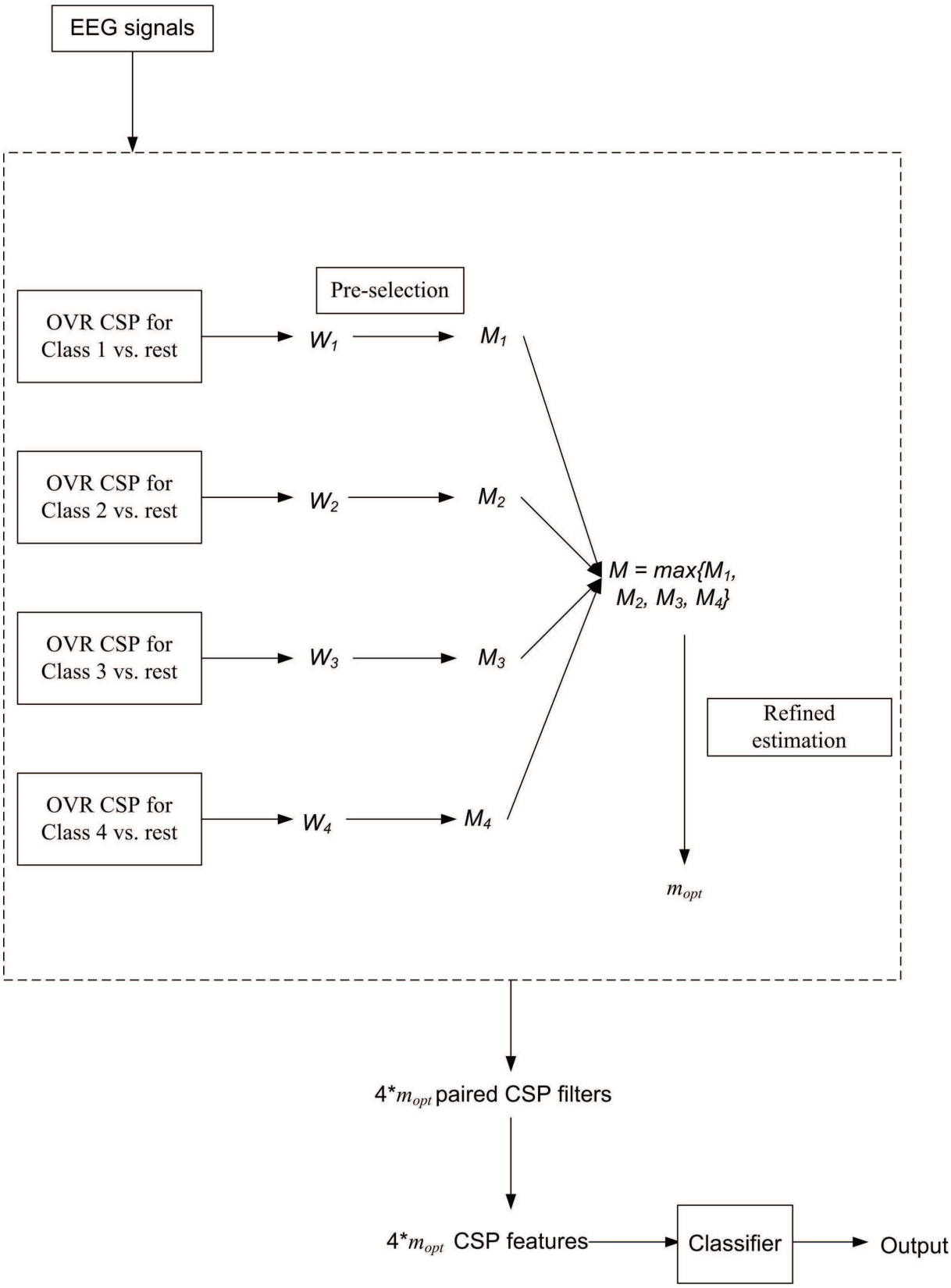


Fig. 4.1: Framework of our method for a four-class problem.

multi-class (left hand, right hand, foot, tongue) problems. Then, we apply CSP with the estimated optimal number of paired spatial filters on three subsets, and compare the

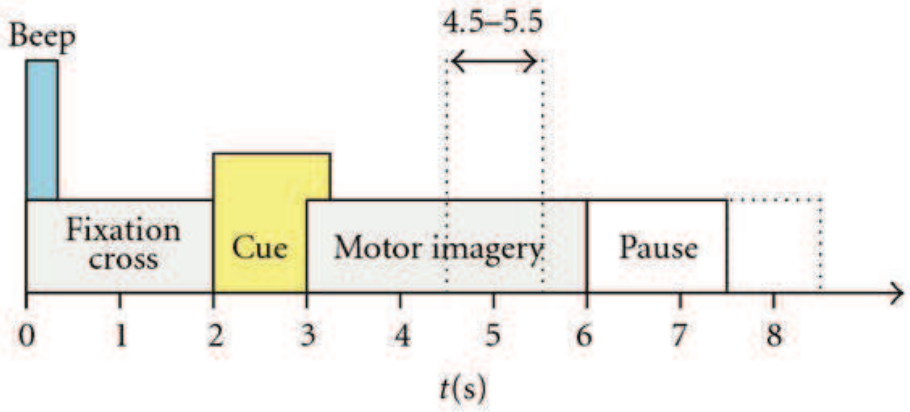


Fig. 4.2: Timing of the experimental paradigm for BCI competition IV dataset IIa.

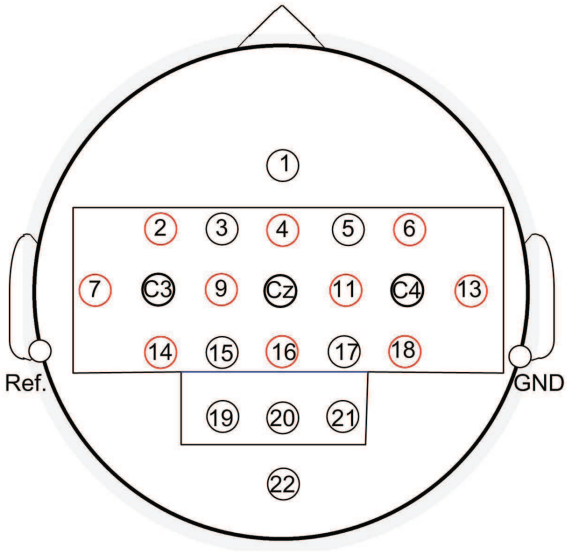


Fig. 4.3: Montage of 22 electrodes in BCI competition IV dataset IIa [39].

classification results with those of using the full channel data and those of using a subset of channels selected by a machine learning method.

4.4.1 Experimental validation of the estimation of the optimal number of paired filters

A broad frequency band of 8-30Hz (μ and β bands) and the segment of 0.5-2.5s of EEG data after the cue on-set were used in this study for calculating the transformation matrix W in CSP and the $FD(i)$ value for each pair of spatial filters, and for training the classifier [135]. The *Fisher's linear discriminant analysis* (LDA), which is classically used with CSP, was employed here for the classification [37]. The effect of the number of spatial filters was studied on the training data using 100 repetitions of 10-fold cross-validations. The classification performances were measured by κ value. Algorithms of CSP, classifier training and evaluation (including calculating κ value) are performed with the BioSig

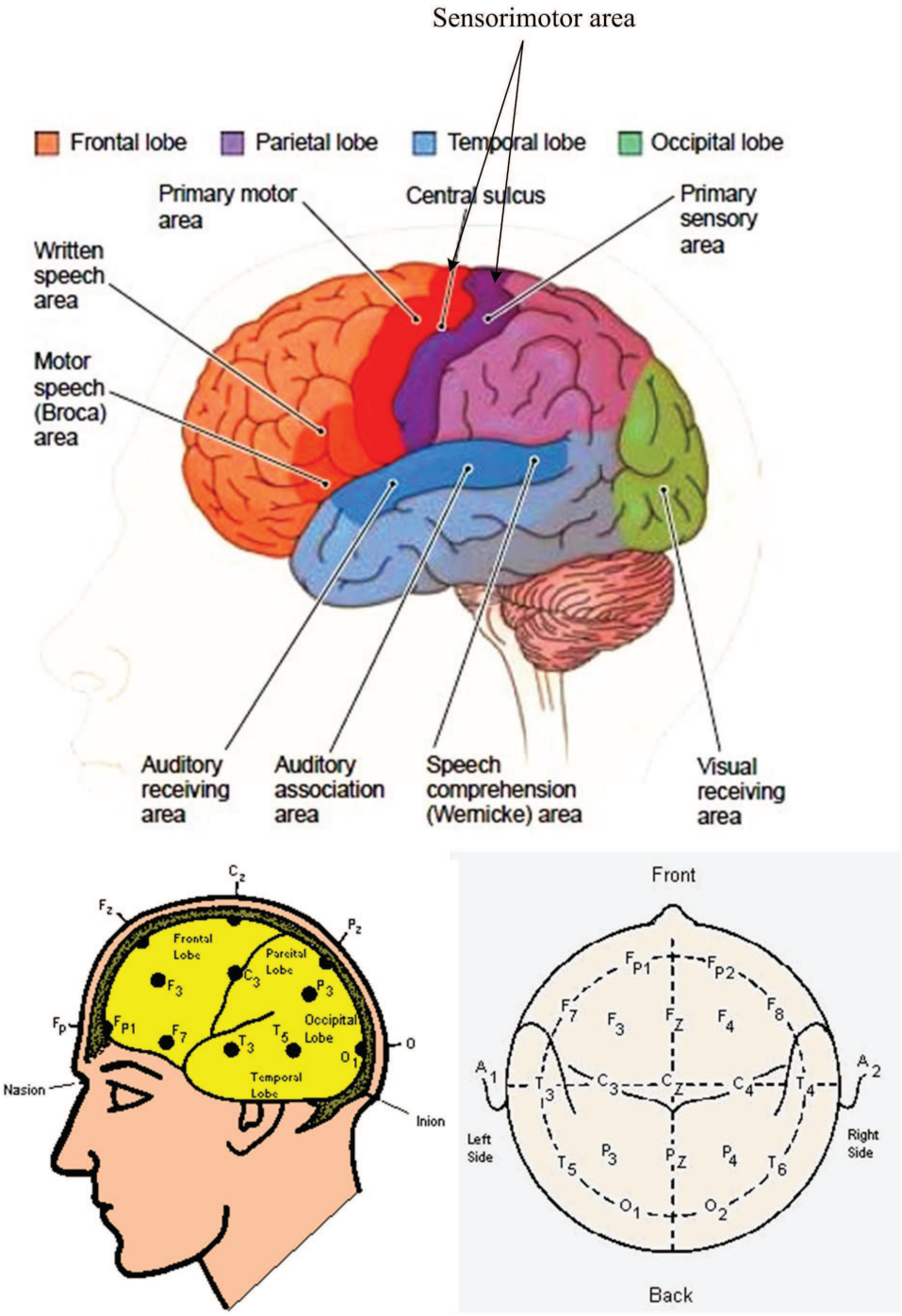


Fig. 4.4: Brain functional areas and their corresponding standard EEG recording positions.

toolbox [176].

The effects of the parameter m on the classification results and $FD(i)$ value of each pair of spatial filters for all subjects are illustrated in Figure 4.5. From Figure 4.5, we can see that the performance of CSP-based classification is not proportional to m but has significant variations depending on m for all subjects. Table 4.1 lists the standard variance of κ values over different m ($m \in [1, 11]$), $\sigma_m(\kappa)$, for each subject, which can be used to evaluate the sensitivities of classification results to m values. From Table 4.1,

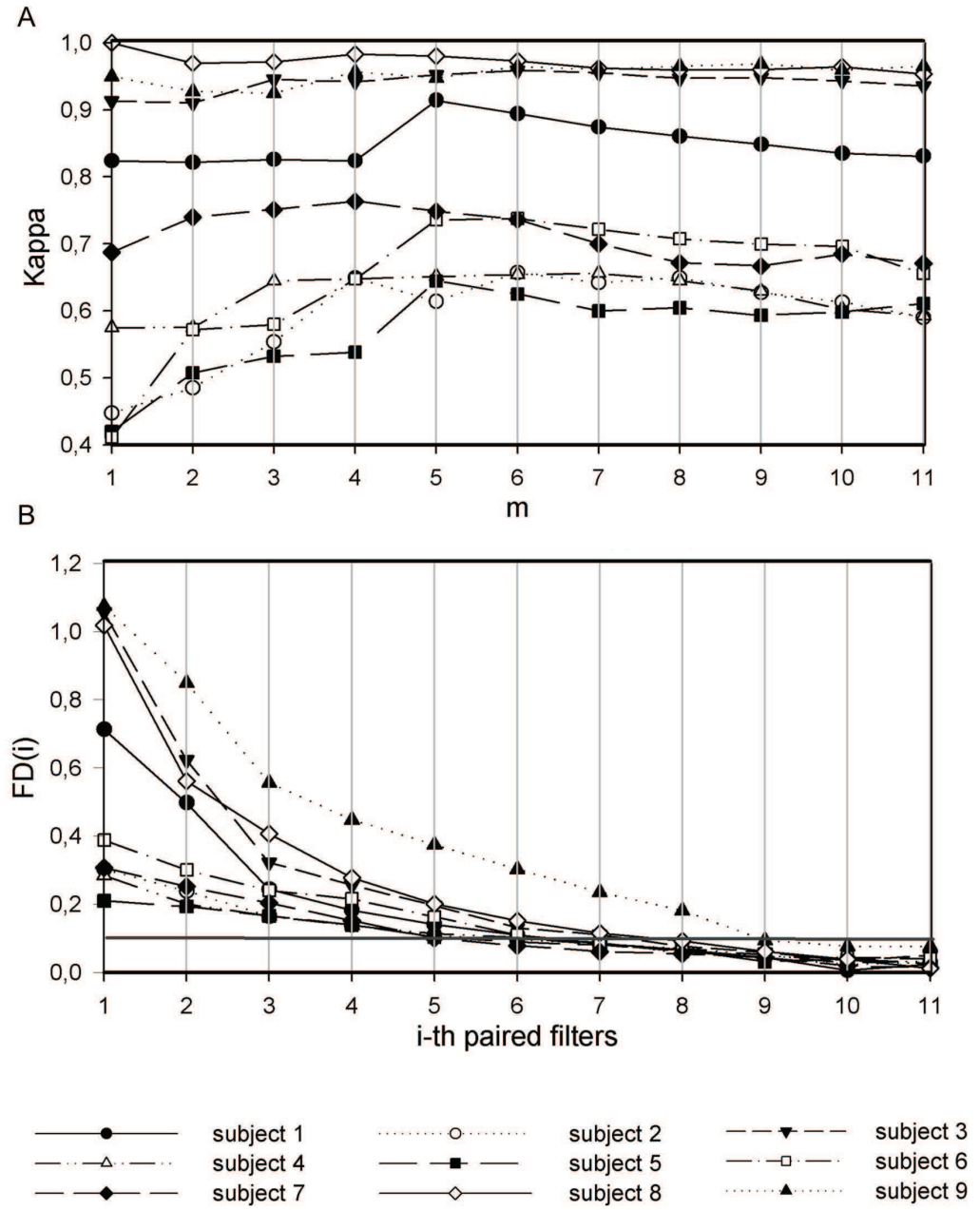


Fig. 4.5: Effect of parameter m on the left vs right hand classification and $FD(i)$ of each pair of spatial filters for all subjects in the BCI competition IV dataset IIa. The horizontal line on the second plot indicates $FD = 0.1$.

we can see that the sensitivity of classification results to m varies among subjects. For Subjects 8 and 9, $\sigma_m(\kappa)$ is relatively low ($\sigma_m(\kappa) = 0.01$), indicating a low sensitivity to m values, while for Subject 2, 5, and 6, $\sigma_m(\kappa)$ is relatively high ($\sigma_m(\kappa) > 0.05$), indicating a high sensitivity to m values. These results prove that it is critical to choose a right m value for each individual in CSP-based classification.

To examine the effectiveness of setting the threshold $fd = 0.1$ to shrink the range of m in $[1, M]$, we list the values of M for different subjects in Table 4.2. From Figure 4.5 and Table 4.2, we can see that adding the paired spatial filters with $FD(i) < 0.1$ does not improve the classification results: e.g. for Subject 1, $M = 6$ and the κ value decreases if

$m > 6$ is used; for Subject 8, $M = 7$ and the κ value remains stable when $m > 7$. Thus, it is reasonable to estimate the optimal m in the range of $[1, M]$ using the threshold $fd = 0.1$.

Table 4.1: Standard deviation of κ values over different m ($m \in [1, 11]$), $\sigma_m(\kappa)$, for each subject.

	Subjects								
	1	2	3	4	5	6	7	8	9
$\sigma_m(\kappa)$	0.03	0.07	0.02	0.03	0.07	0.10	0.04	0.01	0.01

Table 4.2: The M values for different subjects.

	Subjects								
	1	2	3	4	5	6	7	8	9
M	6	4	7	6	5	6	5	7	8

Table 4.3 lists the estimated m_{opt} values learned from 100 repetitions of 10-fold cross-validations in the training data of two classes (i.e. left hand, right hand) and provides a comparison of the evaluation results on the independent testing data using m_{opt} and the classical value ($m = 3$) recommended in [37, 135]. The m_{opt} value varies for different subjects, further indicating the necessity of a subject-specific estimation. Note that for Subjects 2 and 5, the m_{opt} value is equal to the M value, indicating that setting a threshold higher than 0.1 may not lead to the optimal m to achieve the best classification result. From Table 4.3, we can see that classification performances obtained with the estimated m_{opt} value are signification better than those obtained by using the recommended value (paired t-test: $P = 0.01$). For Subjects 8 and 9, whose sensitivities to m are relatively low, one pair of filters can already yield fine performance, while others may need more pairs of filters.

Table 4.3: Estimated m_{opt} values and comparison of independent evaluation results between using m_{opt} and the recommended value ($m = 3$). Estimated m_{opt} values were obtained by 100 repetitions of 10-fold cross-validations on the training data. The evaluation results were computed on the independent testing data using m_{opt} and the recommended value ($m = 3$). The best performances are shaded in gray.

	Subjects									Mean
	1	2	3	4	5	6	7	8	9	
m_{opt}	5	4	3	3	5	5	4	1	1	
κ (m_{opt})	0.75	0.22	0.96	0.40	0.11	0.35	0.70	0.94	0.86	0.59
κ ($m = 3$)	0.67	0.13	0.96	0.40	0.09	0.25	0.69	0.93	0.82	0.55

The comparison of the results obtained with m_{opt} and with fixed recommended m is shown in Table 4.4 for the four-class problem of BCI competition IV dataset IIa. Using m_{opt} leads to better mean performance (mean $\kappa = 0.53$ over all subjects) than using the fixed recommended m (mean $\kappa = 0.51$ over all subjects). Although this result is not statically significant ($P > 0.05$), using m_{opt} generates either better or equal performances (when $m_{opt} = 3$) compared to fixed recommended $m = 3$ for all individuals, indicating the effectiveness of the proposed estimation in multi-class problems.

Table 4.4: Estimated m_{opt} and independent evaluation in a four-class problem using m_{opt} and fixed m, and comparison with the 2^{nd} placed winner in BCI competition IV who also used 8-30Hz data but pair-wise approach (PW CSP) with fixed m ($m = 4$). The best performances are shaded in gray.

	Subjects									Mean
	1	2	3	4	5	6	7	8	9	
m_{opt}	4	3	1	2	3	3	2	4	2	
κ (m_{opt})	0.72	0.30	0.74	0.48	0.20	0.25	0.75	0.71	0.68	0.53
κ ($m = 3$)	0.69	0.30	0.71	0.47	0.20	0.25	0.74	0.71	0.50	0.51
κ (2^{nd}, m=4)	0.69	0.34	0.71	0.44	0.16	0.21	0.66	0.73	0.69	0.52

Table 4.5 lists the classification results of five BCI competition winners on the same dataset. Among them, only the 1^{st}, the 2^{nd} and the 4^{th} place winners used full channels with CSP; and both the 1^{st} and the 4^{th} place winners used extended CSP algorithms based on multiple narrow bands [194]. More detailed information and discussion about these winners and their results are available in [194] and Chapter 7, which focuses on multi-class problems. For this chapter, it makes more sense to only compare with the 2^{nd} place winner [194] (the results of the 2^{nd} place winner are reminded in Table 4.4), since it used the same frequency band and the same number of electrodes as we did. In this case, the effect of frequency band and the number of channels can be eliminated, so as to validate the interest of using subject-specific m_{opt} with OVR CSP. The framework of the method used by 2^{nd} place winner is illustrated in Figure 4.6.

Table 4.5: The classification results (evaluated with κ) of the BCI competition winners on BCI competition IV dataset IIa

	Subjects									
	1	2	3	4	5	6	7	8	9	Mean
1^{st}	0.68	0.42	0.75	0.48	0.40	0.27	0.77	0.75	0.61	0.57
2^{nd}	0.69	0.34	0.71	0.44	0.16	0.21	0.66	0.73	0.69	0.52
3^{nd}	0.38	0.18	0.48	0.33	0.07	0.14	0.29	0.49	0.44	0.31
4^{th}	0.46	0.25	0.65	0.31	0.12	0.07	0.00	0.46	0.42	0.30
5^{th}	0.41	0.17	0.39	0.25	0.06	0.16	0.34	0.45	0.37	0.29

The comparison showed that using m_{opt} with OVR approach in CSP-classification generates better mean performance (mean $\kappa = 0.53$ over all subjects) than the 2^{nd} place winner (mean $\kappa = 0.52$ over all subjects) on this dataset. Although this result is not statically significant ($P > 0.05$), our results outperform those of the 2^{nd} place winner for most subjects (except Subjects 2, 8, and 9), indicating the effectiveness of using subject-specific m_{opt} with OVR CSP)

Note that for Subjects 2, 8 and 9, our method does not yield better results than the 2^{nd} place winner but does generate results either equal to or better than using the fixed value of m. Let us take Subject 9 as an example for further discussion. As we reviewed in Chapter 2 Section 2.8.1, the CSP filtered signals $Z = W^T X = [z_1, ...z_N]$ can be considered as uncorrelated EEG source components. In this case, the original EEG signals can be reconstructed by the inverse transformation $X = UZ$, where $U = [u_1, u_2, ..., u_N] = (W^T)^{-1}$. We call u_j ($j = 1, ..., N$) a spatial pattern, since it reflects the spatial distribution of

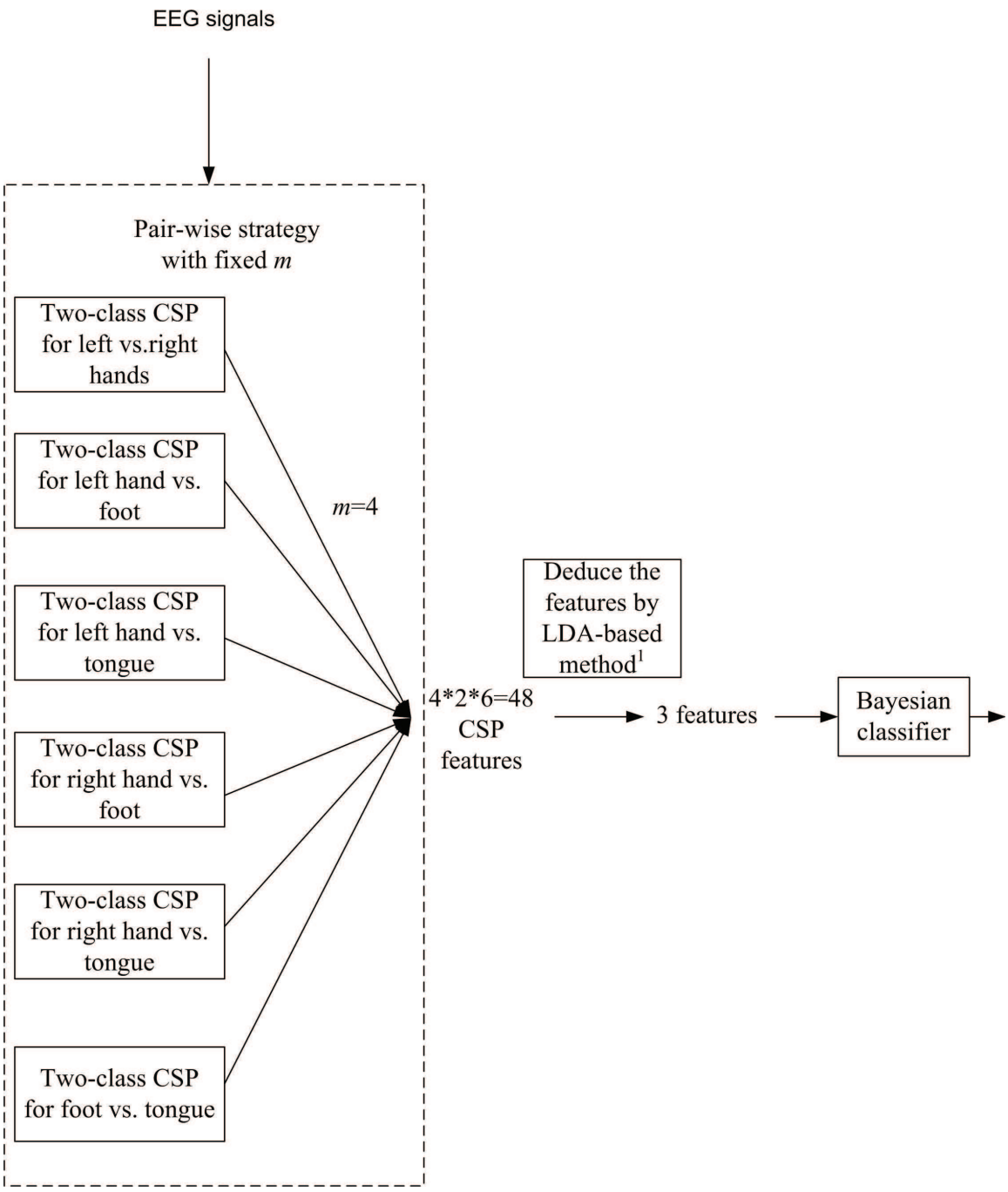


Fig. 4.6: Framework of the method used by the 2^{nd} place winner [194]

uncorrelated source component extracted by spatial filter w_j ($w_j^T u_j = 1$) [220]. Figure 4.7 shows the normalized spatial patterns according to first three pairs of spatial filters obtained by OVR CSP for Subject 9. For all classes, the third pair of spatial filter can only bring in an undesirable source component whose spatial distribution is not consistent with brain functional area (see Figure 4.7). For example, the third pair of spatial filters for left hand movement generates a component mainly distributed around C3 (dark area), which is the representation area of right hand rather than left hand. Thus, using $m_{opt} = 2$ ($\kappa = 0.68$) yields far better results than using fixed $m = 3$ ($\kappa = 0.50$) for this

[1]The details of this method are not available.

subject. However, for foot movement, the second pair of spatial filter extract a component that mainly distributed around P1 (dark area), which is not the foot representation area (around Cz or CPz). It might be a reason why the result of our method on this subject is slightly lower than the 2^{nd} place winner.

The proposed estimation method is not very time consuming. The computational time is 31s on average for two-class problems, in our experimental environment (Matlab 7.10.0, Window 7 Professional 64bits, CPU 2.66GHz, RAM 2.0G). For the four-class problem, the computational time is 88s on average . The optimal number of paired spatial filters can be learned in an off-line analysis with our algorithm and then applied to the on-line applications for the same subject to meet the requirement of rapid processing in real BCI applications.

4.4.2 Channel selection for CSP

The proposed algorithm was applied on the three subsets chosen in Section 4.3 to find the optimal number of paired spatial filters for the four-class CSP-based classification. The classification results are provided in Table 4.6. For comparison purpose, we also provide the classification results obtained when using the full cap (22 channels), and using a subset of channels selected by a machine learning method, recursive channel elimination (RCE) [97, 194]. Recursive channel elimination (RCE) proposed in [97] is a channel selection method based on SVM (margin maximization) and derived from the popular recursive feature elimination method [71, 97, 180]. We have reviewed this method in Chapter 2 Section 2.8.1. In the BCI competition IV, the third place winner applied this method to select channels and then extract CSP features from those channels to perform a multi-class classification using three SVM classifiers with two hierarchies and a voting strategy [194].

Table 4.6: Comparison of the results obtained using subsets of channels with m_{opt}, using the full cap with m_{opt}, and using the subset of channels selected by recursive channel elimination [97] for the multi-class problem of BCI competition IV dataset IIa. The best performances are shaded in gray.

	Subjects									
	1	2	3	4	5	6	7	8	9	Mean
Full cap (22 channels)	0.72	0.30	0.74	0.48	0.20	0.25	0.75	0.71	0.68	0.53
Subset 1 (20 channels)	0.49	0.20	0.47	0.34	0.17	0.20	0.45	0.61	0.33	0.36
Subset 2 (17 channels)	0.50	0.07	0.51	0.33	0.12	0.18	0.34	0.46	0.29	0.31
Subset 3 (13 channels)	0.47	0.14	0.51	0.28	0.22	0.22	0.41	0.53	0.42	0.36
RCE [194] (3^{rd} winner)	0.38	0.18	0.48	0.33	0.07	0.14	0.29	0.49	0.44	0.31

From Table 4.6, we can see that the classification performances significantly drop when using subsets of channels ($P < 0.01$), no matter using the three subsets selected by prior information on the functional areas of the brain or using the subset selected by RCE. This result indicates that a full cap is helpful for improving the classification results when using CSP algorithm. Among three subsets selected by prior information, the third subset that contains electrodes C3, Cz, C4 and their surrounding electrodes is the best choice, since it employs the least number of channels (13 electrodes) and yields the best mean

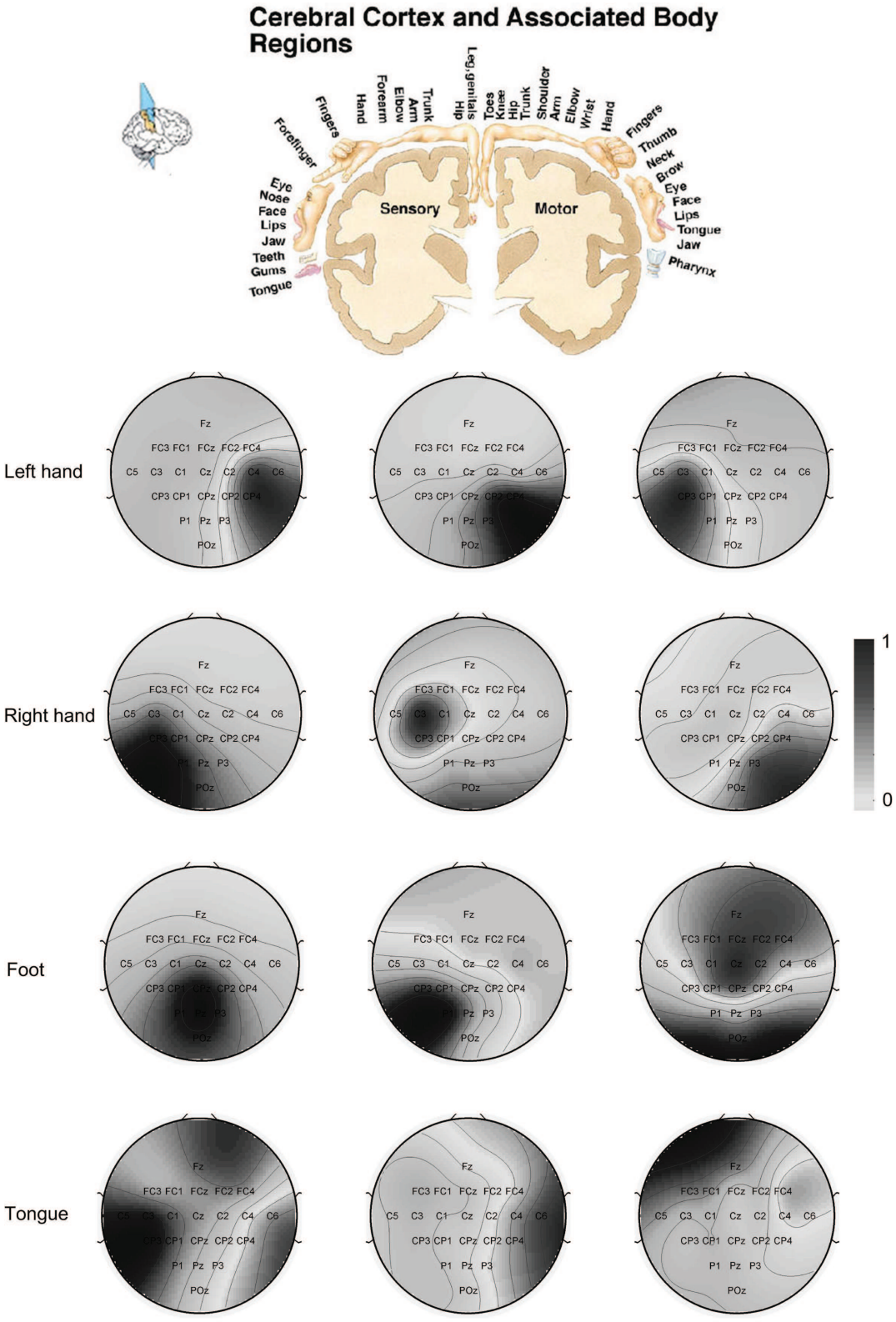


Fig. 4.7: Brain functional areas in sensorimotor cortex and normalized spatial patterns according to first three pairs of spatial filters obtained by OVR CSP for Subject 9 on BCI competition IV dataset IIa.

classification result (mean $\kappa = 0.36$). For Subject 5, using the third subset even generates a slightly better result ($\kappa = 0.22$) than using the full cap ($\kappa = 0.20$). Both the first subset

and the third subset of channels selected based on prior information outperform RCE in the classification results for most subjects (except Subjects 3 and 9 for the first subset, except Subjects 2, 4 and 9 for the third subset), demonstrating the interest of using prior information in BCI problems.

As we have shown the CSP patterns obtained by using full cap for Subject 9 before, we further present the CSP patterns obtained by using different subsets for the same subject in Figure 4.8 to illustrate the influence of reducing the number of electrodes on CSP-based classification. The CSP patterns according to foot motor imagery are given, because the difference between subsets is obvious in these patterns. From Figure 4.8, we can see that reducing the number of electrodes makes the distribution of source component distorted, which results in a deterioration of classification results. However, compared to the first subset and the second subset, the distortion is relatively smaller when using the third subset. The source component still mainly distributes around the foot representation areas (Cz) when using the third subset, while when using the first subset and the second subset, the component mainly distributes around the visual cortex. As a result, using the third subset yields a better result ($\kappa = 0.42$) than using the other two subsets ($\kappa = 0.33$ for the first subset, $\kappa = 0.29$ for the second subset) for this subject.

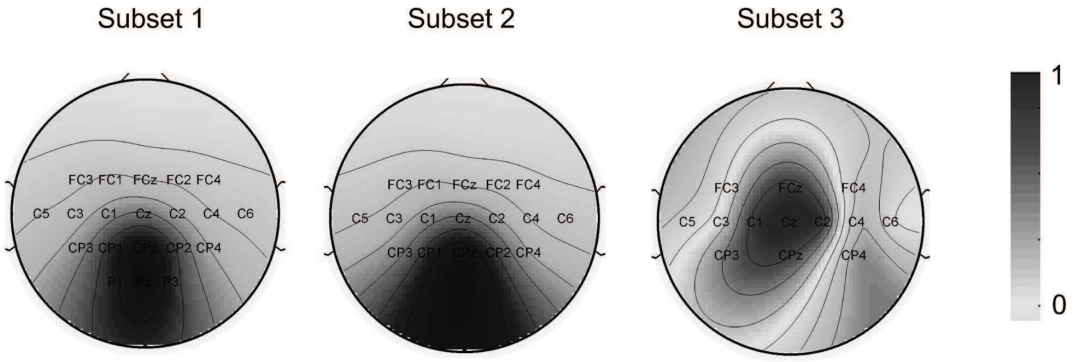


Fig. 4.8: Normalized CSP patterns according to foot motor imagery obtained by using different subsets for Subject 9.

4.5 Conclusion

The number of spatial filters used in feature extraction affects the CSP-based classification results. This chapter proposes an automatic strategy based on *Rayleigh quotient* and cross validation to estimate the subject-specific optimal value of this parameter. Experimental results show that the estimated optimal values of the parameter vary for different subjects and yield better results than those obtained with the fixed recommended value for both binary-class and multi-class problems. The proposed strategy is not very time consuming, and therefore can be applied on the training data to estimate the optimal value of the parameter for each subject and then used for the long term on-line classification of the given classes for the same subject to achieve good performances.

Three subsets of channels selected by prior neurophysiological knowledge were also tested in this chapter to study the channel selection and reduction challenges. The classification results obtained by two subsets selected based on prior information of functional areas

of the brain are generally better than those obtained with a subset of channels selected by a SVM-based machine learning method, called recursive channel elimination (RCE), showing the potential interest of using some prior knowledge in BCI challenges. However, the classification results obtained using reduced channels are not as good as those obtained when using a full cap. Thus, in the next chapters, we will develop some non-CSP based methods to solve the channel selection and reduction challenges for BCIs.

Chapter 5

Subject-specific channel selection using time information for motor imagery BCIs

5.1 Introduction

Although multi-channel recording with a large number of electrodes (e.g. 118) and spatial filtering algorithms (e.g. common spatial patterns (CSP) [37]) can improve the SNR and extract discriminative features from overlapping signals, this setting reduces the portability and practicability of BCI so that it represents a main drawback for final users [236]. Moreover, its effectiveness depends on the choice of the frequency band and the time segment of the EEG data [37].

To develop an easy to use system, several algorithms were proposed to reduce the number of channels in BCI [17, 97, 220]. However, they simply addressed the issue of channel selection by spatial information, disregarding the potential impact of time and frequency information. As a result, the optimal combination of time, frequency and channel position may not be achieved in a BCI design. Although a recent study showed that a broad frequency band (8-30Hz) that covers both μ (8-12Hz) and β (18-25Hz) bands can generally be used when employing features, called time domain parameters (TDPs), the existing channel selection methods mainly work with the popular band power (BP) feature, which is sensitive to frequency band and time segment.

As motor imagery BCIs typically rely on decoding sensorimotor rhythms (see Section 2.3 in Chapter 2), in practice, many researchers simply placed electrodes at three key positions (C3, Cz and C4 of 10-20 recording system [84], see Figure 2.3 in Chapter 2) in the sensorimotor areas to reduce the number of electrodes, which we call 3C setup. The advantages of the 3C setup is that it does not need a full EEG cap, training data and machine learning methods to find the optimal positions for recording. It can be used when only a few electrodes are available. However, due to the limited information and low SNR of signal, it cannot achieve good classification results in most cases. Preprocessing steps (e.g. time-frequency optimization, see next chapters) or advanced classification algorithms (see [12, 46]) are often needed to improve its performance. Moreover, general users may not be skillful enough to place the electrodes at the precise locations of C3, Cz and C4 each time, if a standard EEG cap is not used.

Here, we propose a channel selection method using TDP features. Unlike existing methods, our method considers the effect of the used time segment on channel selection, so as to find the optimal combination of time segment and subset of channels for BCI design.

In this method, a criterion based on Fisher's dicriminant analysis is proposed to measure the discrimination power of TDP features extracted from different channels and different time segments. Comparisons between channel selection using time information (CSTI), channel selection based on the long time segment from the cue on-set to the ending of the cue (CSL), 3C setup and full EEG cap CSP are performed to evaluate the contribution of our method (CSTI). The effect of electrode misplacing on 3C setup is also examined to study its potential influence on classification.

5.2 Method for channel and time segment selection using TDP

5.2.1 Time domain parameters

The EEG signals are bandpass filtered between 8 and 30 Hz using a 5th order Butterworth filter. For one channel (electrode) and one trial, we denote by $x(t)$ the filtered EEG signal in a time segment $[t_0, t_0 + T - 1]$. Time domain parameters (TDPs) are a set of broad band (i.e. 8-30Hz) EEG features defined in the time domain [213]:

$$TDP^{(p)} = \log\left(\underset{t\in[t_0,t_0+T-1]}{var}\left(\frac{d^p x(t)}{dt^p}\right)\right), p = 0, 1, 2, ... \tag{5.1}$$

The logarithm is applied here to make the distribution of TDPs approximately normal (for details, see [213]), since the linear classifier we use here typically assumes that the input features follow Gaussian distributions [132]. Note that the TDP of order 0, $A = TDP^{(0)}$, is the logarithmic band power (BP) of the filtered signal. It characterizes the EEG pattern in terms of amplitude. The EEG signal can be considered as a mixture of sinusoidal waves:

$$x(t) = \sum_m A_m \cdot \sin(w_m t + \Phi_m) \tag{5.2}$$

where $w_m = 2\pi f_m$ is the frequency of each sinusoidal wave. So the derivative provides information on frequency:

$$\frac{dx(t)}{dt} = \sum_m A_m w_m \cdot \cos(w_m t + \Phi_m) \tag{5.3}$$

Thus, the TDP of order 1, $M = TDP^{(1)}$, is a feature that reflects the EEG pattern in terms of frequency, and the TDP of order 2, $C = TDP^{(2)}$, reflects the change in frequency. We use these three TDPs, $[A, M, C]$, in this work, since they carry more information than the only BP feature, and have clearer physical meanings than TDPs of higher orders in BCI research.

5.2.2 A criterion based on *Fisher's* discriminant

*Fisher's*linear discriminant analysis (*Fisher's* LDA) is a very popular classification algorithm in BCI research [132], because it has a very low computational cost and usually

yields good results for motor imagery BCIs [107]. It projects high-dimensional data onto a direction and performs a linear classification in this one-dimensional space. The optimal projection is found by maximizing the separation between two classes. Let us assume that we have two classes of observations, h and f. In a one-dimensional feature space, the separation between two classes is defined using the Fisher criterion [132]:

$$FC = \frac{(\mu^h - \mu^f)^2}{(\sigma^h)^2 + (\sigma^f)^2} \tag{5.4}$$

where μ^h and μ^f are the mean values of the feature over all trials for classes h and f, respectively, and $(\sigma^h)^2$ and $(\sigma^f)^2$ are the variances of the feature.

In feature selection, FC can be used to evaluate the discrimination power of each single feature [132]. However, it is not suitable to evaluate the discrimination power of a group of features. Thus, we propose a novel and simplified criterion based on Fisher's discriminant, called F score, $\hat{F}$, and we use it to estimate the discrimination power of a group of features (here TDP feature vector $[A, M, C]$):

$$\hat{F} = \frac{\left\|\vec{\mu}^h - \vec{\mu}^f\right\|_2^2}{tr(\Sigma^h) + tr(\Sigma^f)} \tag{5.5}$$

where Σ denotes the covariance matrix of the feature vector, $\vec{\mu}$ denotes the mean of the feature vector, $\|\cdot\|_2$ denotes the $L2$-norm (Euclidean norm), and $tr(\cdot)$ the trace of a matrix.

Without loss of generality, let us assume that the feature vector is three-dimensional, $\vec{v}(i) = [v_1(i), v_2(i), v_3(i)]$, $i = 1, ..., K$, where K is the number of samples (trials) for one class. Thus, the mean of the feature vector for the class is $\vec{\mu} = [\mu_1, \mu_2, \mu_3]$, where μ_1, μ_2, μ_3 are the mean values of $v_1(i)$, $v_2(i)$, $v_3(i)$, respectively. We denote by σ_1^2, σ_2^2, σ_3^2 the variances of $v_1(i), v_2(i), v_3(i)$ for the class, respectively. The trace of the covariance matrix for each class can be computed as:

$$\begin{aligned} tr(\Sigma) &= \sigma_1^2 + \sigma_2^2 + \sigma_3^2 = \frac{1}{K-1}\sum_{i=1}^{K}(v_1(i) - \mu_1)^2 \\ &+ \frac{1}{K-1}\sum_{i=1}^{K}(v_2(i) - \mu_2)^2 + \frac{1}{K-1}\sum_{i=1}^{K}(v_3(i) - \mu_3)^2 \\ &= \frac{1}{K-1}\sum_{i=1}^{K}\left[(v_1(i) - \mu_1)^2 + (v_2(i) - \mu_2)^2 + (v_3(i) - \mu_3)^2\right] \\ &= \frac{1}{K-1}\sum_{i=1}^{K}\|\vec{v}(i) - \vec{\mu}\|_2^2 \end{aligned} \tag{5.6}$$

Thus, the trace of the covariance matrix for each class is the mean Euclidean distance between samples to the class center, which reflects intra-class spread.

Compared to FC, $\hat{F}$ is a derived version relying on the Euclidean distance between class centers, $\left\|\vec{\mu}^h - \vec{\mu}^f\right\|_2$, to estimate the difference between classes, and employing the trace of the covariance matrix to evaluate the variance within a class. Note that this simple expression avoids estimating a projection direction as required by the general multi-dimensional expression of *Fisher's* LDA.

5.2.3 F score based channel selection

The TDPs, $[A_e^\chi(i), M_e^\chi(i), C_e^\chi(i)]$, are computed for a time segment $[t_n, t_n+T-1]$ for each single trial i at channel e for class χ ($\chi \in \{h, f\}$). Then, the discrimination power of channel e is estimated by the F score:

$$\hat{F}_e = \frac{(\bar{A}_e^h - \bar{A}_e^f)^2 + (\bar{M}_e^h - \bar{M}_e^f)^2 + (\bar{C}_e^h - \bar{C}_e^f)^2}{\tilde{A}_e^h + \tilde{A}_e^f + \tilde{M}_e^h + \tilde{M}_e^f + \tilde{C}_e^h + \tilde{C}_e^f} \tag{5.7}$$

with

$$\overline{TDP}_e^\chi = \frac{1}{K^\chi}\sum_{i=1}^{K^\chi} TDP_e^\chi(i) \tag{5.8}$$

$$\widetilde{TDP}_e^\chi = \frac{1}{K^\chi - 1}\sum_{i=1}^{K^\chi} (TDP_e^\chi(i) - \overline{TDP}_e^\chi)^2 \tag{5.9}$$

where K^χ is the number of training trials for class χ.

Existing methods typically determine the number of selected channels based on user's experience [220] or exhaustive searching strategy [17, 97], which is either arbitrary or time-consuming. Here, we propose an automatic approach, by considering the properties of both features and classifier to determine the size of the subset of selected channels.

Let $\hat{F}_m$ be the largest F score among all channels:

$$\hat{F}_{max} = \max\left\{\hat{F}_e \mid e \in \{1, ..., 118\}\right\} \tag{5.10}$$

The relative discrimination power of each channel e is defined as:

$$\rho_F(e) = \frac{\hat{F}_e}{\hat{F}_{max}} \tag{5.11}$$

The value of $\rho_F(e)$ is between 0 and 1. A larger $\rho_F(e)$ indicates a larger relative discrimination power. Thus, a threshold $\hat{\rho}$ can be set to extract the channels with $\rho_F(e) > \hat{\rho}$ to be used for classification. A lower value of $\hat{\rho}$ tends to pick out more channels. In practice, the training trials should have several times as many as the dimensionality of features to guarantee a good performance of the classifier [83]. Based on this knowledge, the range of $\hat{\rho}$ can be shrunk to $[P, 1.0]$ to feed the classifier, where P is obtained by:

$$\begin{gathered} \min_P Num(P)\ s.t. \\ P \in [0, 1.0], Num(P) \geq K/3R \end{gathered} \tag{5.12}$$

where $Num(P)$ is the number of selected channels with $\rho_F(e) > P$, K is the number of trials for training, and R is the ratio of the number of trials to the number of features for a specific classifier. Note that each channel yields three TDPs, so here we have $Num(P) \geq K/3R$. As a linear classifier, such as *Fisher's* LDA, typically needs $5-10$ times training trials as many as the dimensionality of features [107], we set $R = 5$ to have a loose range of $\hat{\rho}$ for further optimization. Different subsets of channels according to different $\hat{\rho} \in [P, 1.0]$ are used to train the classifier. The optimal $\hat{\rho}$ is obtained by seeking the subset with the lowest training error (ERR) in the classifier training. The training

error is defined as the observed overall disagreement between classification outputs and true classes. Let $\hat{\rho}^*$ be the optimal $\hat{\rho}$, so it is obtained by:

$$ERR(\hat{\rho}^*) = \min\{ERR(\hat{\rho}) \mid \hat{\rho} \in [P, 1.0]\} \tag{5.13}$$

If there are more than one $\hat{\rho}^*$ obtained by Equation 5.13, we use the largest $\hat{\rho}^*$ as the optimal one.

5.2.4 Channel selection using time information (CSTI)

This method aims to find the optimal combination of time segment and subset of channels for classification. The general scheme of the method is shown in Figure 5.1. First, we compute the TDP features and the F score for each channel in a series of overlapping T-width time segments $[t_n, t_n+T-1]$ $(n = 1, ..., N)$, $t_{n+1} = t_n + T_s$ (T_s is the step), during the motor imagery duration $[T_0, T_e]$, where T_0 is the beginning time of motor imagery and T_e is the ending time. Then, the optimal subsets of channels $S(t_n)$ and their corresponding training error $ERR(\hat{\rho}^*(t_n))$ are obtained by the F score based channel selection proposed above for different time segments $[t_n, t_n+T-1]$ $(n = 1, ..., N)$, where $\hat{\rho}^*(t_n)$ is the optimal $\hat{\rho}$ in the time segment $[t_n, t_n + T - 1]$. The optimal time segment $[t^*, t^* + T - 1]$ is found by seeking the lowest training error $ERR(\hat{\rho}^*(t_n))$ among all time segments:

$$ERR(\hat{\rho}^*(t^*)) = \min_{t_n}\{ERR(\hat{\rho}^*(t_n))\} \tag{5.14}$$

so as to obtain the optimal subset of channels $S(t^*)$ in the optimal time segment $[t^*, t^* + T - 1]$.

5.3 Results

The dataset IVa [33] from BCI competition III is used in this study. As it consists of EEG signals recorded using 118 electrodes, this dataset is very suitable for a fine selection of EEG channels. Five subjects, denoted "aa", "al", "av", "aw" and "ay", have performed 280 trials of cue-driven motor imagery (right hand: 140 trials, right foot: 140 trial) during the recording. The acquisition process was driven by visual cues, presented during 3.5s, and separated by randomly chosen intervals, ranging from 1.75 to 2.25s. Subjects were required to perform the corresponding motor imagery task during the presentation of a cue and to relax in the intermission. Thus, $T_0 = 0$ is the time point of the cue-onset, $T_e = 3.5s$ is the ending of the cue. Ground truth is available for all subjects in this dataset.

The aim of the experiment is to perform classification of the signal, for each subject, into two classes (right hand, right foot), with as few electrodes as possible. The F score based channel selection was performed in five ($N = 5$) overlapping time segments of 0-2.0s, 0.5-2.5s, 1.0-3.0s, 1.5-3.5s and 2.0-4.0s after the cue on-set ($t_n = 0, 0.5, 1.0, 1.5, 2.0s, T = 2s, T_s = 0.5$) to find the optimal combination of time segment and subset of channels by CSTI [**?**]. For comparison purpose, we also performed F score based channel selection, full EEG cap based CSP and 3C setup in a long time segment from the cue on-set to the ending of the cue. *Fisher's* LDA was used as the classifier in this study, since F

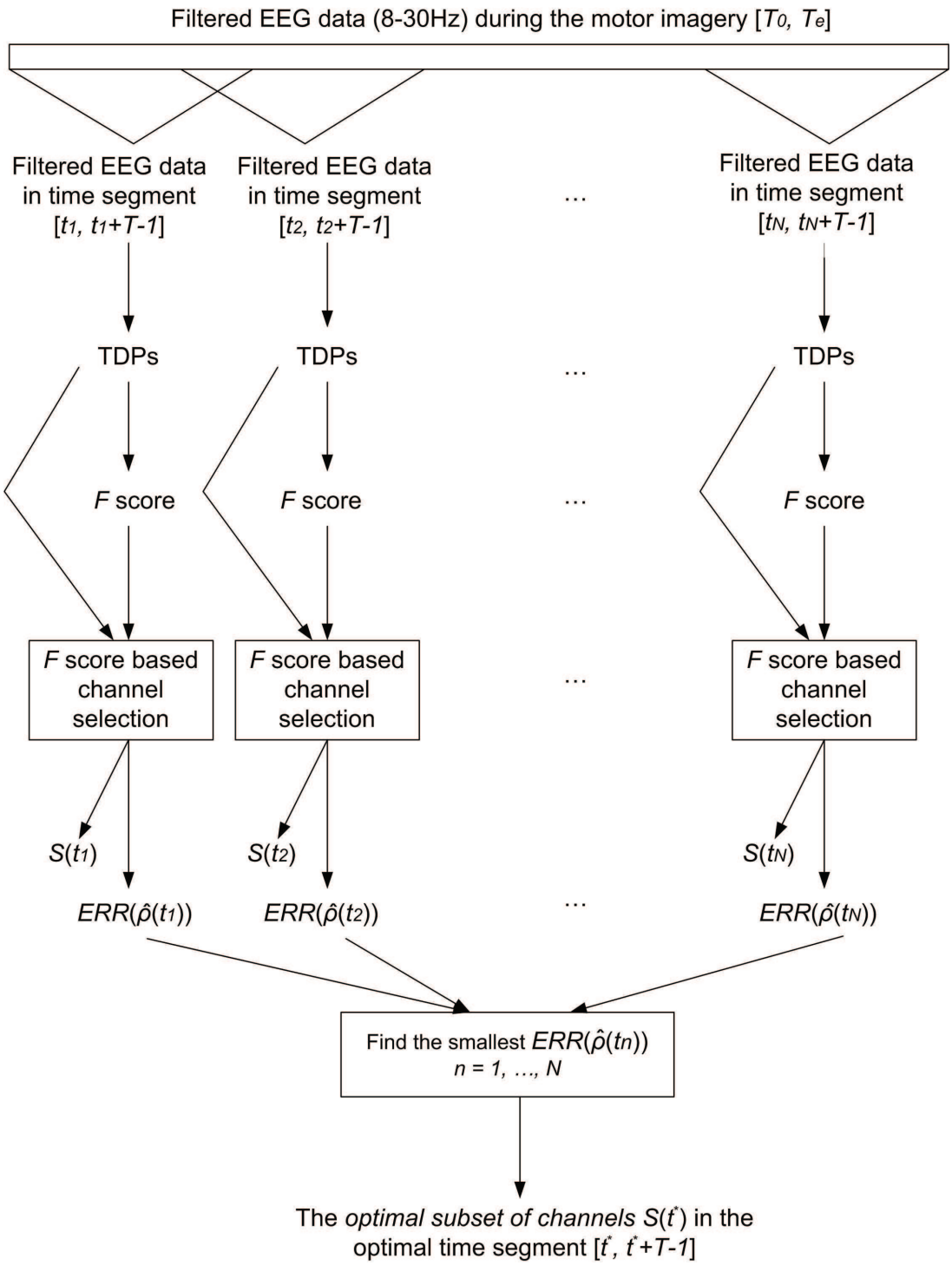


Fig. 5.1: General scheme of CSTI.

score is proposed based on *Fisher's* discriminant, and it works well with TDP and BP features. The paired-sample t-test was employed to reveal the statistical significance of the difference between the results of different methods.

First, we used the first 70 trials for each class for training, and the remaining ones for the independent testing, to evaluate the contributions of our methods. The results are provided in Sections 5.3.1 to 5.3.3. This choice of training/testing data corresponds to a usual situation in real applications, where the training data are recorded before the

testing data. Using 50% trials for training makes the information for training comparable to that for testing. Secondly, considering the data evolution, we also tested our method with randomly selected training and testing data (training : testing = 70 : 70 for each class) to evaluate the robustness of our method. The results are provided in Section 5.3.4.

5.3.1 Effect of time segment on electrode selection

The spatial distribution of the F score and the selected electrodes in different time segments are shown in Figure 5.2, where the selected time segments are marked out by squares. The testing results obtained when using the selected electrodes in different 2s length time segments are provided in Table 5.1, and the results from the selected time segments are shaded in gray. The results are evaluated by classification accuracy (ACC), which is defined as the observed overall agreement between classification outputs and true classes. From Figure 5.2, we can see that the subsets of selected electrodes vary with time segments for each subject, indicating that time segment is an important factor that should be considered in electrode selection. Among all possible combinations of time segment and subset of electrodes, the selected combination yields the highest classification accuracy (ACC) on the testing data. This result shows that CSTI is effective in finding the optimal combination of time segment and subset of electrodes. However, CSTI has a computational cost which is at least N times (N is the number of different time segments, here $N = 5$) the one of methods only performing channel selection in only one single time segment. In our experiments, the computation time for CSTI was 11 seconds, while for channel selection in a single time segment it was 2 seconds (Matlab 7.10.0, Window 7 Professional 64bits, CPU 2.66GHz, RAM 2.0Go). Nevertheless, this additional calibration time remains acceptable for several applications, such as neuro-games.

In this study, we also performed F score based channel selection in a long time segment (CSL) from the cue on-set to the cue ending (that covers the whole period of motor imagery) to see: 1) whether a long time segment will improve the results of channel selection (i.e. selecting less electrodes and/or improving classification accuracy), 2) whether the effect of time segment can be forgotten by just using a long time segment that covers the period of motor imagery, so as to save computation time. Comparisons between CSTI and CSL are provided in Table 5.2. From Figure 5.2 and Table 5.2, we can see that, compared to CSTI, CSL selects less electrodes (except for "ay"). However, CSL only improves ACC for one subject ("av"). For other subjects, CSL yields significantly worse ACC than CSTI does ($p < 0.05$). Considering the deteriorated ACC, using a long time segment in channel selection cannot fully improve the results for most subjects. Thus, the effect of time segment cannot simply be ignored by using the long time segment.

Although CSTI tends to select more electrodes than CSL does, the number of CSTI selected electrodes, which is no more than 11 (see Figure 5.2 and Table 5.2), is less than that of commercial BCI system Emotiv EPOC[1], which has 14 electrodes. Thus, the number of electrodes selected by CSTI is still reasonable and acceptable for general applications (e.g. in a game environment).

[1]https://www.emotiv.com/

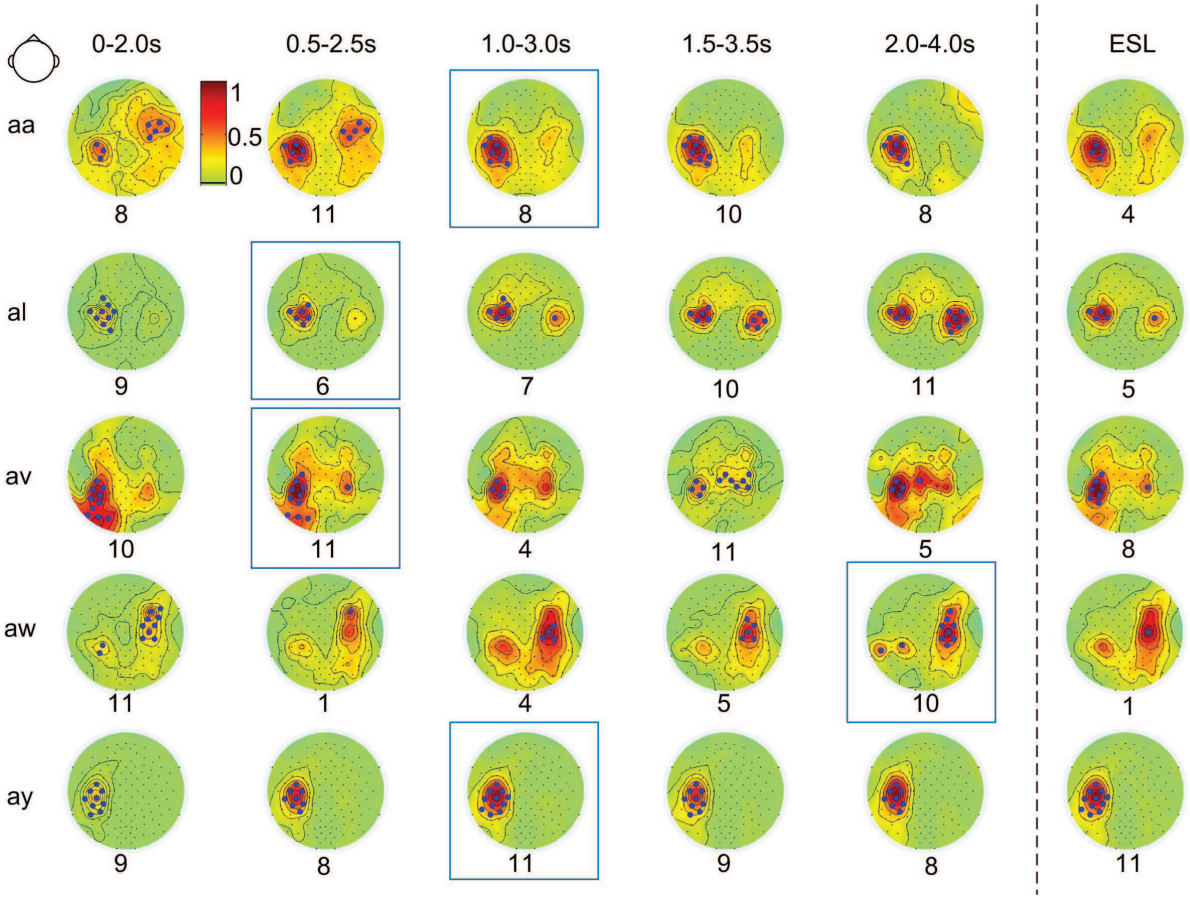


Fig. 5.2: Topographic maps of the F score (color scale) and selected electrodes (marked by bold points) in different time segments. The number of selected electrodes is given below each map. The selected time segments are marked out by rectangles. Results in the long time segment obtained by CSL are provided in the last column.

Table 5.1: ACC results when using the selected subset of electrodes in different time segments (the selected time segment is shaded in gray).

Subject	0-2.0s	0.5-2.5s	1.0-3.0s	1.5-3.5s	2.0-4.0s
aa	0.53	0.67	0.67	0.67	0.64
al	0.78	0.89	0.87	0.87	0.88
av	0.49	0.61	0.61	0.61	0.58
aw	0.66	0.51	0.71	0.75	0.81
ay	0.73	0.90	0.92	0.84	0.80

5.3.2 Comparisons with full-cap CSP and 3C setup

Table 5.3 lists the testing results (evaluated by ACC) of the full-cap CSP and the 3C setup using BP and TDP features. To make the comparison easy, the testing results of CSL and CSTI are also reminded in Table 5.3. For the full-cap CSP as well as for the 3C setup, using TDPs yields better mean ACC ($\overline{ACC} = 0.78$ for full-cap CSP, $\overline{ACC} = 0.72$ for 3C setup) than using BP ($\overline{ACC} = 0.76$ for full-cap CSP, $\overline{ACC} = 0.71$ for 3C setup). The difference is not significant ($p > 0.05$) due to the limited number of subjects in this dataset. For most subjects, using TDPs did improve ACC, which is in line with the results

Table 5.2: Comparisons between CSL and CSTI: number of selected electrodes and ACC.

	Number of selected electrodes		ACC	
Subject	CSL	CSTI	CSL	CSTI
aa	4	8	0.61	0.67
al	5	6	0.82	0.88
av	8	11	0.63	0.61
aw	1	10	0.74	0.81
ay	11	11	0.84	0.92
Mean	6	9	0.73	0.78

in [213]. With the BioSig toolbox [176], TDPs are easy and fast to calculate [213], which only need 2ms to compute (Matlab 7.10.0, Window 7 Professional 64bits, CPU 2.66GHz, RAM 2.0Go). Unlike BP which often requires the selection of frequency bands, TDPs save computation time during the frequency band selection. All of these indicate the interest of using TDPs in motor imagery BCI.

The results obtained using CSTI ($\overline{ACC} = 0.78$) are significantly better ($p < 0.05$) than simply using 3C setup ($\overline{ACC} = 0.71$ when using BPs, $\overline{ACC} = 0.72$ when using TDPs). The mean classification accuracy when using CSTI is better than using full-cap CSP with BP features ($\overline{ACC} = 0.76$, not significant with $p > 0.05$) and equal to using full-cap CSP with TDP features ($\overline{ACC} = 0.78$ when using TDPs). For some subjects ("aa" and "ay"), CSTI even yields higher ACC than full-cap CSP. Thus, CSTI meets the goal of largely reducing the number of electrodes (from 118 channels to no more than 11), without a major loss of classification performance. Moreover, CSTI uses a relatively shorter time segment (2s length) than other methods (3.5s length). For most subjects (except "aw"), the classification outputs were obtained before the ending of cue, which indicates that less time (here, less than 3.5s) is required for recording the training data from those subjects.

Although CSL generates slightly better mean ACC ($\overline{ACC} = 0.73$) than simply using 3C setup, this improvement is not significant ($p > 0.05$) and does not occur for all subjects. Moreover, CSL tends to select more than three channels and needs a full EEG cap to acquire training data for seeking the optimal subset of electrodes. Thus, CSL is not cost-efficient in real applications.

Among all methods, the mean ACC of 3C setup is the worst, but it uses the least number of electrodes (only three channels) and can yield better ACC than the full-cap CSP for one subject in the dataset ("aa"). Moreover, 3C setup has no additional computational cost and does not need full-cap training data for calculating CSP filters or seeking the optimal subset of electrodes. Thus, for electrode reduction, the choice between CSTI and 3C setup may depend on a preference between the number of electrodes, the computation cost, the amount of training data and the classification performance. This choice can be left to the user.

5.3.3 Effect of electrode misplacement

The electrode positions might have undergone slight changes compared to the standard 10-20 recording system [84] in real applications, in particular for general users who may not be proficient in EEG recording. For example, an inexperienced user may put the EEG

Table 5.3: Testing results (ACC) for different methods. The best performance for each subject is shaded in gray.

	CSP (118)		3C setup (3)		CSL (< 11)	CSTI (< 11)
Subject	BP	TDPs	BP	TDPs	TDPs	TDPs
aa	0.46	0.47	0.64	0.59	0.61	0.67
al	0.94	0.94	0.79	0.81	0.82	0.88
av	0.68	0.69	0.58	0.59	0.63	0.61
aw	0.94	0.94	0.73	0.79	0.74	0.81
ay	0.75	0.84	0.81	0.82	0.84	0.92
mean	0.76	0.78	0.71	0.72	0.73	0.78

cap a little bit left; as a result, all electrodes are placed at the left side of the standard positions during the recording.

In practice, the training and testing data may be recorded in two different ways. In the first way, they are recorded in one session without re-placing the electrodes. In this case, if misplacement happens, both the training and testing data are recorded at the same non-standard positions. For machine learning based methods, e.g. CSTI, the effect of electrode misplacement can be neglected, because the optimal subset of electrodes is estimated based on the actual positions, where the data are recorded, instead of standard positions; while for 3C setup, this effect should be examined, because the selected channels (C3-Cz-C4) are defined according to the standard positions. When the cap is put incorrectly, nominal channels (C3, Cz and C4) of 3C setup will not be in their standard positions.

In the second way, the training and testing data are recorded in two sessions (maybe in two different days) with re-placing the electrodes. As a result, the training and testing data may be recorded at different non-standard positions. Usually, not only the shift of electrodes should be consider in this case, but also the change of the mental state of the user [35]. It is a very complicated problem, so called the challenge of "session-to-session transfer" [35]. In fact, all methods face this challenge. As both the change of mental state and the shift of electrodes may exist but are unpredictable, even if a method has achieved a good performance in one "session-to-session transfer" test, it may fail in the next one if the changes are too large. In real applications, commercial BCI systems (Emotiv and Neurosky) require the user to wait a few seconds (or minutes) for calibration after putting the cap (to check the electrode impedance) and to perform a training session with feedback before the real play, to overcome this challenge. As a result, the problem of the shift of electrodes turns to be the same as the one discussed in the previous paragraph, at the price of additional training.

To examine the effect of electrode misplacement on 3C setup, we compared the classification results obtained using the standard 3C setup (C3-Cz-C4) and using the non-standard 3C setup with the electrodes placed a little left (C5-C1-C2), right (C1-C2-C6), forward (FC3-FCz-FC4), backward (CP3-CPz-CP4) with respect to the standard positions (see Figure 5.3). Table 5.4 shows that using the electrodes placed a little backward, the classification results are improved for Subjects "aa", "al" and "av", but deteriorated for Subjects "aw" and "ay". However, for all subjects, the results using the electrodes placed a little forward are significantly worse than using the electrodes placed at the standard positions ($p < 0.01$) and a little backwards ($p < 0.01$). Using electrodes placed a little left or right, the results are deteriorated compared to those obtained with the electrodes placed at the

standard positions. Compared to those obtained with the electrodes placed a little right, the results obtained with using the electrodes placed a little left are better for Subjects "aa", "av" and "ay", but worse for Subjects "al" and "aw".

Figure 5.4 shows that the large values of F score are mainly distributed in the post-central areas of the brain for all subjects, which explains why using the electrodes placed a little backward always generates better results than using the electrodes placed a little forward. Meanwhile, for Subjects "aa", "av" and "ay", the distributions of large values of F score show a left-brain dominance. Thus, the results obtained with using the electrodes placed a little left are better than those obtained with using the electrodes placed a little right for those subjects.

To sum up, the effect of changes of electrode position on classification results depends on the subject and the direction of error placement. As an inexperienced user may unconsciously misplace the electrodes, the effect will be unpredictable when simply using 3C setup and may lead to a deteriorated result. Concerning this effect, CSTI can be recommended to users who are not very professional in EEG recording. However, training data and computation time are needed for finding the optimal subset of electrodes.

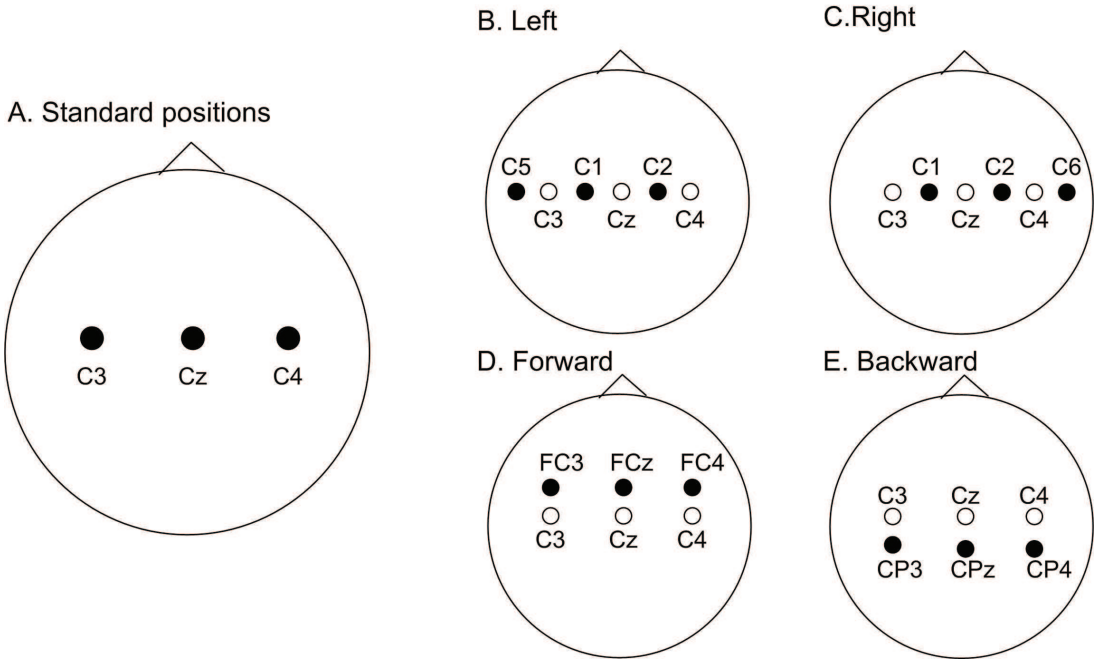


Fig. 5.3: Standard 3C setup (C3-Cz-C4) and the non-standard 3C setups with the electrodes placed a little left (C5-C1-C2), right (C1-C2-C6), forward (FC3-FCz-FC4), backward (CP3-CPz-CP4).

5.3.4 Effect of data evolution

The non-stationarity of EEG is a common problem in BCI. It may result from several causes. For example, changes in electrode impedance may occur when the electrically-conductive gel between skin and electrode dries out or an electrode gets loose. The task involvement and attention level of a subject may change over the course of a BCI experiment. All these factors will lead to some unpredictable modulations in EEG signals,

Table 5.4: ACC results for standard 3C setup (C3-Cz-C4) and the non-standard 3C setups with the electrodes placed a little left (C5-C1-C2), right (C1-C2-C6), forward (FC3-FCz-FC4), backward (CP3-CPz-CP4). The best performance for each subject is shaded in gray.

Subject		Standard	Left	Right	Forward	Backward
aa	BP	0.64	0.57	0.48	0.52	0.69
	TDP	0.59	0.56	0.51	0.49	0.63
al	BP	0.79	0.65	0.74	0.61	0.80
	TDP	0.81	0.70	0.77	0.68	0.86
av	BP	0.58	0.54	0.53	0.59	0.61
	TDP	0.59	0.64	0.59	0.62	0.62
aw	BP	0.73	0.58	0.66	0.65	0.66
	TDP	0.79	0.54	0.66	0.64	0.70
ay	BP	0.81	0.77	0.50	0.69	0.79
	TDP	0.82	0.74	0.59	0.66	0.77

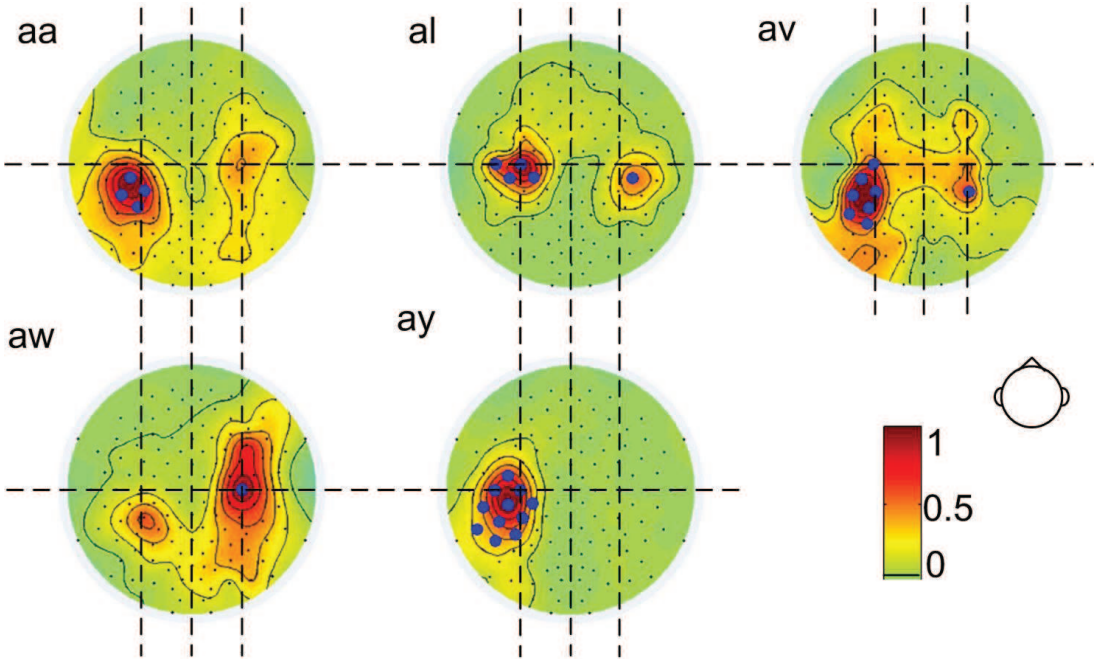


Fig. 5.4: Distribution of F-score for different subjects in the long time segment from the cue on-set to the ending of the cue. Electrodes selected by CSL are marked by bold points.

making the signal in a time segment or at a channel full of artifacts, so as to affect the selection of time segment and channel.

To examine this effect, we randomly selected 140 trials (H/F = 70/70) as the training dataset to find the optimal combination of time segment and subset of electrodes by CSTI for each subject, the remaining data forming the testing dataset. We repeated this procedure 100 times. For comparison, we also calculated the subset of electrodes based on the long time segment by CSL. The experimental results generated by CSTI show that the optimal time segments are not always the same for different training datasets even for the same subject. The possible reason for this result is that the subject may not have the same response time to the cue in different trials due to different mental states and

possible fatigue during the BCI experiment. The distribution of optimal time segments for each subject is given in Figure 5.5. It shows that the optimal time segments mainly appear in the range of 0.5-3.0s (i.e. the second and third time segments) for Subjects "aa", "av" and "ay", while a little bit later (i.e. the fourth time segment 1.0-3.5s) for Subjects "al" and "aw", indicating that some subjects may need relatively longer time for recording the useful data in each trial compared to other subjects. The subsets of selected electrodes also vary with different training datasets for the same subject. Thus, re-selecting the optimal combination of time segment and subset of electrodes might be needed when using a different training dataset due to the effect of data evolution.

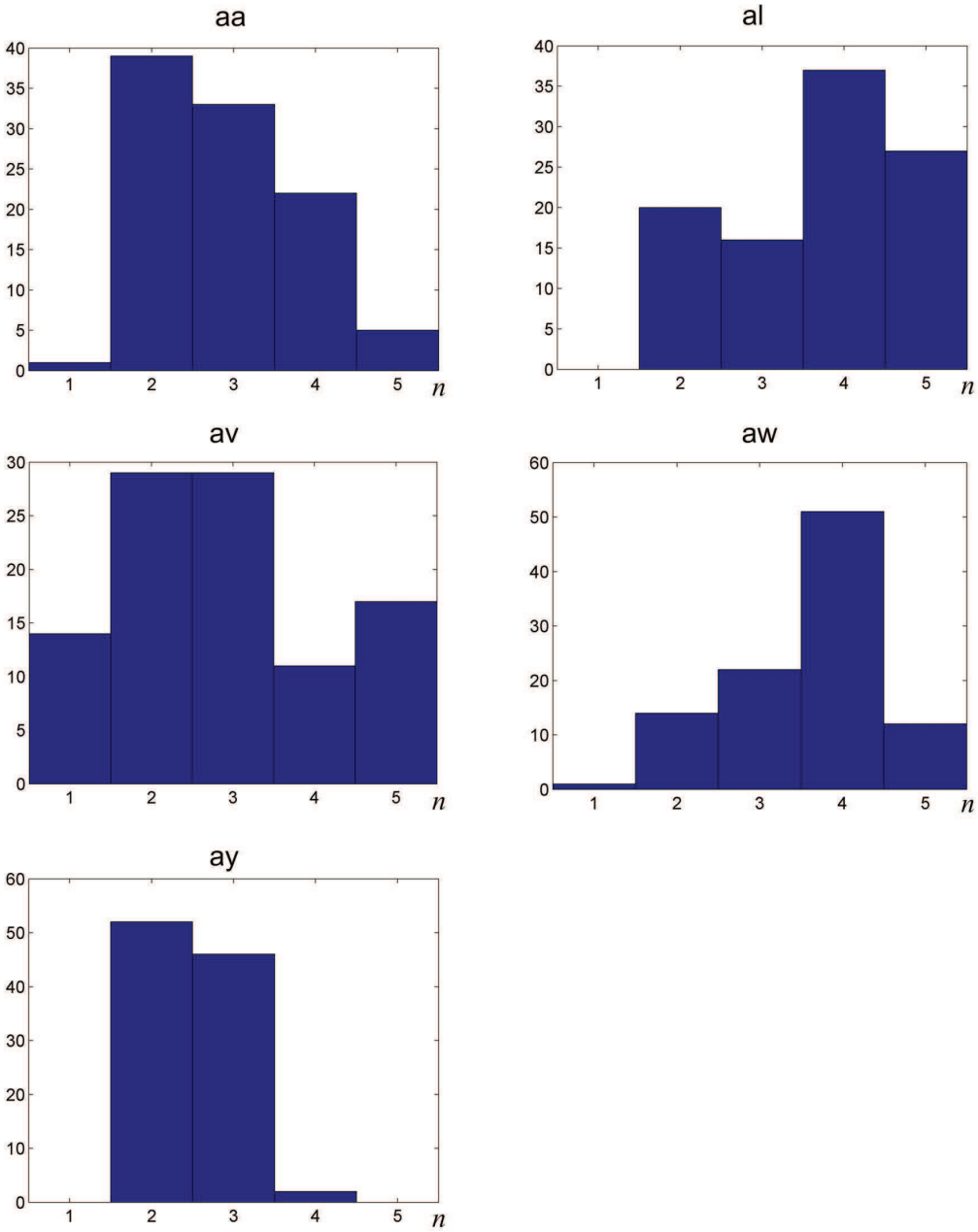


Fig. 5.5: Distribution of optimal time segment for each subject. The horizontal axis n indicates the time segments $[t_n, t_n + T - 1]$ $(n = 1, ..., 5)$. The vertical axis shows the number of times each time segment is selected.

The union and intersection of the subsets of selected electrodes obtained by the 100 repetitions for each subject are shown in Figure 5.6. The union of the subsets contains all selected electrodes in the 100 repetitions, which reflects the range of all possible positions of selected electrodes for each subject. The intersection shows the key electrodes, which were always selected in the 100 repetitions.

From Figure 5.6, we can see that the selected electrodes are in a range instead of covering the whole cap, but the range varies with subjects. This result implies that it is possible to use a part of instead of all electrodes in an EEG cap to find the optimal subset of electrodes.

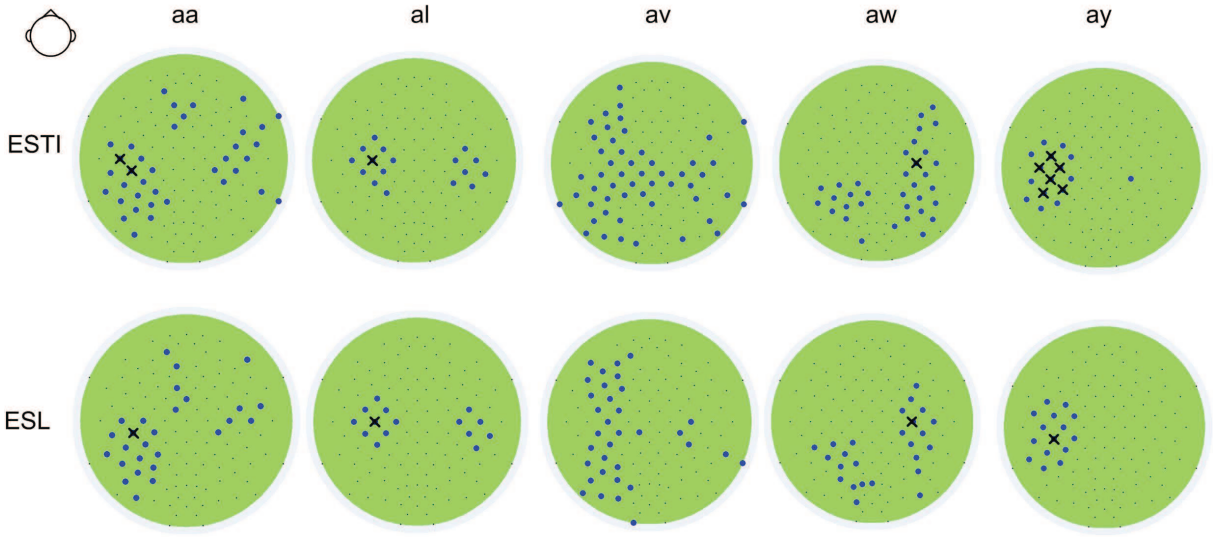


Fig. 5.6: Union (bold points) and intersection (marked by $\times$) of selected electrodes for each subject.

Table 5.5 lists the cardinalities of the unions U_{CSL} and U_{CSTI} of CSL and CSTI selected subsets of electrodes, respectively. CSL generated a smaller range of selected electrodes than CSTI did. Most electrodes selected by CSL have also been selected by CSTI. We use the inclusion ratio, $IR(CSL : CSTI)$ to measure the percentage of electrodes selected by CSL and selected by CSTI:

$$IR(CSL : CSTI) = |U_{CSL} \cap U_{CSTI}|/|U_{CSL}| \times 100\% \tag{5.15}$$

where $|U|$ denotes the cardinality of set U. The results show that more than 88% electrodes selected by CSL have also been selected by CSTI (see Table 5.5). Thus, we can conclude that U_{CSL} approximately lies in U_{CSIT}. For all subjects, both U_{CSL} and U_{CSIT} cover the hand representative area of the sensorimotor cortex of the left brain, because motor imagery of the right hand typically elicits strong ERD in this area (see Figure 5.7). For most subjects, the possibly selected electrodes also exist in the hand representation area of the right brain, because motor imagery of the right foot may elicit ERS in this area (see Figure 5.7), which can also contribute to classification. Electrodes in central, frontal and occipital cortices were seldom selected neither by CSL nor by CSIT, indicating that those areas may be less important in motor imagery of foot and hand movements. Thus, we may not have to place electrodes in those areas for finding the optimal subset of electrodes for classifying motor imagery data of foot and hand movements.

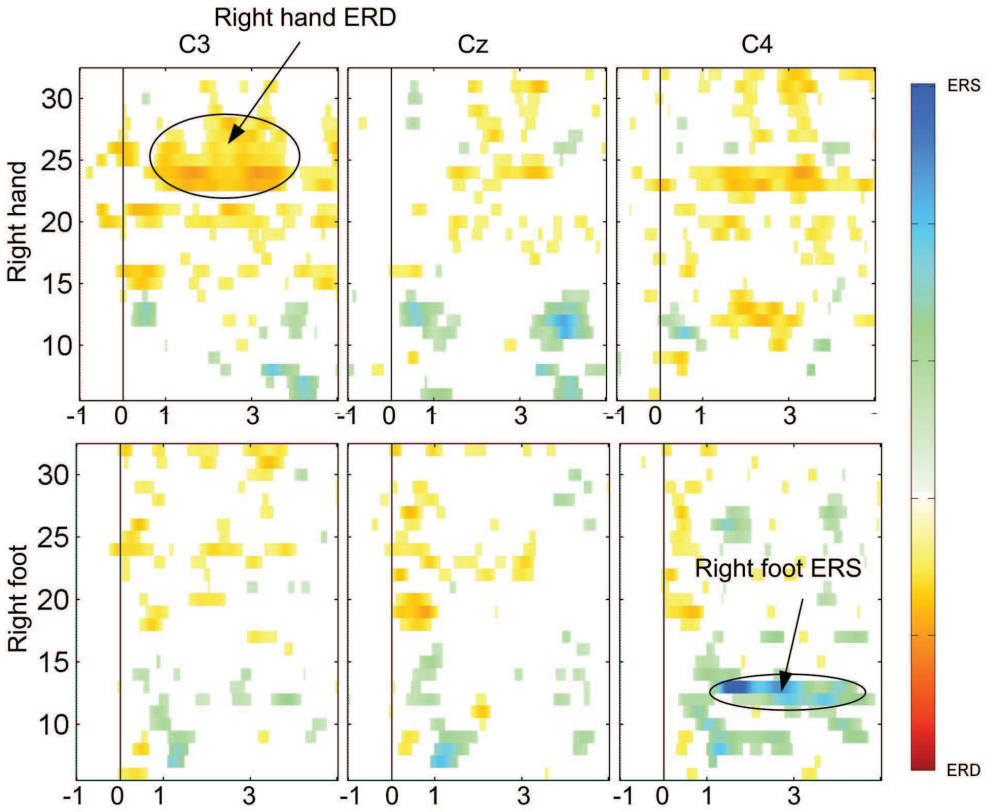


Fig. 5.7: Time-frequency visualization of ERD/ERS for Subject "aw". It was generated by the BioSig Toolbox [176], using overlapping 2Hz bands (step = 1Hz) in the frequency range between 6 and 32Hz, from 1s before cue on-set to 4s after cue on-set (for details, see [152]).

Table 5.5: Cardinalities of the unions of CSL and CSTI selected subset of electrodes and inclusion ratio between CSL and CSTI, $IR(CSL : CSTI)$.

Subject	CSL	CSTI	$IR(CSL : CSTI)$
aa	35	27	92%
al	17	15	100 %
av	52	33	88 %
aw	34	25	88 %
ay	16	14	100 %
Mean	31	21	94 %

Table 5.6 lists the standard positions of the key electrodes (electrodes in the intersection of the optimal subsets) for each subject. Most subjects only have one or two key electrodes, which are located at the right or left hand representation areas. For Subjects "al" and "aw", the key electrode is one electrode (C3 or C4) in 3C setup. For Subject "ay", the key electrodes also contain one electrode (C3) in 3C setup when using CSTI. These results indicate the importance of some positions in 3C setup.

Among all subjects, Subject "ay" has the largest number (6) of key electrodes and the smallest number (16) of electrodes in the union when selecting electrodes by CSTI (see Tables 5.5 and 5.6). These results indicate that this subject needs only a few number of electrodes for seeking the optimal subset. Moreover, for this "good" subject, the six

key electrodes can be always placed for each recording, so that the computation time can be largely reduced by only examining the contributions of the other ten electrodes in the union. On the contrary, Subject "av" is a "bad" subject, who does not has any key electrodes, and has the largest number (52) of electrodes in the union (see Tables 5.5 and 5.6).

Table 5.6: Standard positions of key electrodes (electrodes in the intersection of the optimal subsets). Positions of electrodes contained in 3C setup are marked in red.

Subject	aa	al	av	aw	ay
CSTI	CCP5 CP5	C3	-	C4	C3 CFC5 C5 CCP5 CP5 CP3
CSL	CCP5	C3	-	C4	CCP5

To further compare CSTI with other methods, we also calculated ACC of each method in each repetition. The mean value and standard deviation, σ, of ACC over 100 repetitions for each subject are listed in Table 5.7. CSTI provides significant higher ACC values than those obtained by CSL for most subjects ($p < 0.01$), except Subject "av". Both CSTI and CSL significantly outperform C3, Cz, C4 for most subjects ($p < 0.05$), except Subject "av". The ACC of full-cap CSP are significantly higher than those of CSTI for Subjects "aa" ($p < 0.01$), "av" ($p < 0.01$) and "aw" ($p < 0.01$) but not for Subjects "al" and "ay". These results further prove that CSTI can find an optimal combination of time segment and subset of electrodes to generate fairly good classification results (mean ACC over all subjects: 0.79), which are close to those obtained by using full-cap CSP (mean ACC over all subjects: 0.84) and better than those obtained by using other methods. Moreover, the standard deviations of ACC obtained by using CSTI are relatively small (all $\sigma \leq 0.05$), indicating that CSTI can keep stable classification performances for different training/testing data by finding the optimal combination of time segment and subset of electrodes.

Table 5.7: Mean values and standard deviations σ (given in brackets, 0.01) of ACC for different methods over 100 repetitions for each subject. The best mean ACC values are marked in gray.

	CSP		3C setup		CSL	CSTI
Subject	BP	TDPs	BP	TDPs	TDPs	TDPs
aa	0.76 (12)	0.76 (10)	0.62 (3)	0.62 (4)	0.68 (4)	0.71 (4)
al	0.92 (13)	0.92 (14)	0.85 (2)	0.88 (2)	0.90 (3)	0.92 (2)
av	0.69 (4)	0.68 (3)	0.60 (3)	0.61 (3)	0.62 (3)	0.61 (4)
aw	0.94 (2)	0.96 (2)	0.75 (2)	0.80 (3)	0.81 (5)	0.82 (5)
ay	0.87 (6)	0.88 (5)	0.83 (2)	0.83 (2)	0.84 (4)	0.89 (2)
Mean	0.83	0.84	0.73	0.75	0.77	0.79

5.4 Discussion

Although earlier studies have presented the need for selecting and reducing the electrodes required in a BCI system [17, 97, 220], they addressed this issue based only on spatial information, disregarding the potential impact of temporal information. The contribution in this paper, with the proposition of a novel method, CSTI, emphasizes the potential effects of the chosen time segment on channel selection. A criterion derived from *Fisher's* criterion is proposed to evaluate the discrimination power of a group of features, and applied on time domain parameters (TDP), which overcomes the disadvantage of *Fisher's* criterion [132] on TDP feature selection.

Comparisons between CSTI, CSL, 3C setup and full-cap CSP were performed. The comparisons of their average performances on classification accuracy and reducing the number of channels, their computational costs and training data required for finding the optimal subset of electrodes can be summarized as follows:

- Mean classification accuracy: 3C setup $<$ CSL $<$ $CSTI$ $<$ full-cap CSP;
- Mean number of channels used: 3C setup $<$ CSL $<$ $CSTI$ $<$ full-cap CSP;
- Computational cost for finding the optimal subset of electrodes: 3C setup = full-cap CSP (no computational cost) $<$ CSL (2 seconds in the experiment) $<$ $CSTI$ (11 seconds);
- Training data required for finding the optimal subset of electrodes: 3C setup = full-cap CSP (not needed) $<$ $CSL = CSTI$ (needed).

Although full-cap CSP generates the best average classification performance among all methods, it employs the largest number of electrodes. The tedious placement of EEG electrodes unavoidably reduces its application in non-clinical areas, such as for a home use BCI system. Moreover, the classification performance obtained by full-cap CSP is not always the best and may be even worse than 3C setup in some cases. Thus, the classification performance is not proportional to the number of electrodes, and it is possible to reduce the number of electrodes without deteriorating the classification results.

The 3C setup uses only three channels (C3, Cz, C4) that cover the sensorimotor areas of brain. This setting has the lowest number of electrodes, and does not need a standard EEG cap, training data and computation time to find the optimal subset of electrodes. It is an ideal choice when only a very few electrodes (i.e. less than 10) are available. However, in most cases, its classification accuracies are not as good as for other methods due to the limited information it exploits. Moreover, the 3C setup relies on a precise placement of electrodes, so it may not be easy to use for users who are not professional in EEG data recording.

CSL often chooses more than 3 channels for classification, however, it can only slightly improve classification accuracy compared to 3C setup. Thus, it may not be a good choice in most cases.

CSTI can largely reduce the number of channels (from 118 channels to no more than 11) compared to full-cap CSP, and shorten the training time, without a remarkable decrease of classification performance on a standard dataset (BCI competition III dataset IVa).

The number of electrodes selected by CSTI is less than the commercial BCI system Emotiv EPOC has, so the number (no more than 11) is still reasonable and acceptable for general public applications, such as BCI games. This method can be used in designing BCI systems using few channels (electrodes) for subject-specific applications. This work can also help the user decide on the best compromise between accuracy, easy use and portability, according to his needs.

Chapter 6

Time-frequency optimization for discrimination between imagination of right and left hand movements based on two bipolar EEG channels

6.1 Introduction

In the previous chapter, we have proposed a non-CSP based algorithm to reduce the number of electrodes in a BCI system by selecting a few key channels. However, the algorithm-based channel selection method needs a full EEG cap and additional computational time for finding the optimal subset of channels. Furthermore, the optimal subset of channel should be found for each subject due to individual differences. As a result, this solution will definitely increase the hardware cost of a BCI device, because a subject-specific montage of electrodes should be produced based on the result of channel optimization for each subject.

To overcome these problems, bipolar recordings are recommended in portable BCI systems to reduce the number of electrodes [100, 108]. A bipolar channel of EEG is obtained by subtracting two monopolar EEG signals (for details see Chapter 2 Section 2.6.1) [147]. This acquisition improves the SNR by eliminating shared artifacts between two monopolar channels (for details, see [108]). Therefore they may achieve as good performances as usual multi-channel monopolar settings, using only a few electrodes (i.e. two or three pairs of electrodes) placed around task-relevant sensorimotor areas. Typically, the bipolar electrodes are placed on locations C3 and C4 of the international 10-20 system [185] (see Figure 6.1) for hand-related motor imagery tasks, since these locations correspond to the hand representation areas in the cerebral cortex [156].

However, ERD/ERS patterns are typically short-lasting (half to few seconds) and their frequency range may vary with subjects [234]. Thus, only placing pairs of electrodes around the task-relevant areas may not be sufficient to achieve a good classification. A precise user-specific time-frequency parameterization is required in the feature extraction step for improving classification performances. In the past, a number of approaches were proposed to estimate time-frequency characteristics of motor imagery EEG [234, 8, 80, 232], but only a few were successfully applied to bipolar recording data. Among those methods, the filter bank CSP (FBCSP) method seems to be the most effective one, because it yields the best BCI performances on BCI competition datasets [5]. FBCSP was initially proposed only for frequency band optimization, and then extended to include an optimal temporal selection process [8]. However, FBCSP-based methods involve feature selection

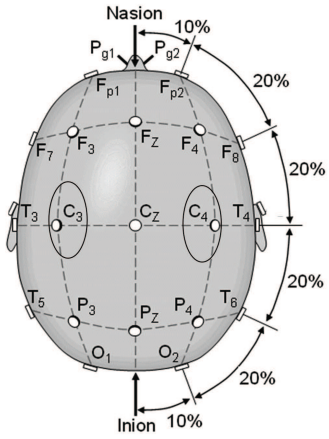


Fig. 6.1: The international 10-20 system standardized by the American Electroencephalographic Society [185]. The positions of C3 and C4 are indicated by ellipses.

procedures based on mutual information, which require tedious iterative steps that greatly increase their complexity.Moreover, the latest version of FBSCP selects the optimal time segment from only four different options (0.5-2.5s, 1.0-3.0s, 1.5-3.5s and 0.5-3.5s), which did not yield better results on bipolar recording data (BCI competition IV dataset IIb) compared to previous versions [8].

In this chapter, we address the issue of time-frequency optimization with only two bipolar channels (C3 and C4) for the discrimination between right and left hand motor imagery tasks. On the contrary to the coarse selection of the time segment in FBSCP, we propose to take into account fine subject-specific time-frequency characteristics for feature extraction. Moreover, our approach is neither based on CSP algorithm nor combined with complex algorithms, such as mutual information based algorithms, and employs less electrodes than CSP-based methods. The strategy of subject-specific time-frequency optimization builds on our preliminary work in [236], and includes three steps: 1) the time-frequency domain of the input bipolar channels is divided into a set of overlapping regions with different time segments and frequency bands, 2) the discrimination power of each time-frequency region is measured, 3) the optimal time-frequency region is selected by finding the region with the largest discrimination power. Once the optimal time-frequency region for each subject is found, the classification is performed using a simple linear classifier, i.e. *Fisher's* linear discriminant analysis (LDA). This classifier has a very low computational cost and usually yields good results for motor imagery BCIs [107].

We adapt the proposed F score (see Chapter 5 section **??**) for estimating the discrimination power of each time-frequency region. Meanwhile, we also propose a novel alternative criterion based on domain specific knowledge (such as location of brain activity during a motor imagery) for the same purpose. The comparison between these two criteria, as well as with state-of-the-art methods, is performed on a standard bipolar dataset (the BCI competition IV IIb) and their contribution to electrode reduction is evaluated on BCI Competition IV IIa dataset.

The chapter is organized as follows: The time-frequency optimization methods are presented in Section 6.3. Section 6.2 introduces the ERD/ERS in motor imagery. Section 6.4 briefly describes the two datasets (BCI competition IV dataset IIa and IIb) used in our experiments. Results are given in Section 6.5.

6.2 ERD/ERS in motor imagery

The ERD/ERS patterns are often expressed as percentage power decrease (ERD) or power increase (ERS) in relation to a reference interval and can be calculated as follows [154, 152]. Let us assume that EEG signals are recorded using K repetitions of the process (K trials) for a motor imagery task. These signals are filtered using a series of bandpass filters in order to obtained ERD/ERS patterns in different frequency bands. We denote by x_i the filtered data from the i-th trial, $i \in \{1, ..., K\}$, x_{ij} the j-th sample in x_i, and $\bar{x_j}$ the mean of the j-th sample averaged over all trials ($\bar{x_j} = \frac{1}{K}\sum_{i=1}^{K} x_{ij}$). Thus, the power value A_j at j-th time point over all trials is defined as:

$$A_j = \frac{1}{K-1}\sum_{i=1}^{K}(x_{ij} - \bar{x_j})^2 \tag{6.1}$$

The percentage power modulation Δ_j at j-th time point is then calculated by:

$$\Delta_j = \frac{A_j - A_{ref}}{A_{ref}} \times 100\% \tag{6.2}$$

where A_{ref} is the mean value of A_j in a reference interval $[t_r, t_r + T_r - 1]$:

$$A_{ref} = \frac{1}{T_r}\sum_{j=t_r}^{t_r+T_r-1} A_j \tag{6.3}$$

A negative value ($\Delta_j < 0$) indicates a power decrease, denoted as ERD_j; and a positive value ($\Delta_j > 0$) indicates a power increase, denoted as ERS_j.

In this chapter, ERD/ERS values of different frequency bands for one EEG channel were subsequently calculated to construct time-frequency maps of ERD/ERS (for details, see [67]). These maps are useful to show the time-frequency distribution of significant band power increase or decrease during the hand imagery tasks, so as to get a better understanding and a visual validation of the time-frequency selection methods. As the standard quantification of ERD/ERS described above employs an averaging procedure over all trials, it is not used to generate features for single-trial classification [152].

6.3 Time-frequency optimization for classification

The EEG signals at C3 and C4 are decomposed into signal components first, in a series of overlapping time-frequency regions ($\omega_m \times \tau_n$), $m \in \{1, , ..., M\}$, $n \in \{1, , ..., N\}$ with different frequency bands $\omega_m = [f_m, f_m + F - 1], f_{m+1} = f_m + F_s$ (F is the bandwidth, F_s is the frequency step) and time intervals $\tau_n = [t_n, t_n + T - 1], t_{m+1} = t_m + T_s$ (T is the interval width, T_s is the time step). The aim of time-frequency optimization is to find a time-frequency region that contains the most discriminative information, so called region of interest (ROI), for a given subject. The selected ROI is then used to extract the band power features [213] that feeds the classifiers.

A measure evaluating the discrimination power of a region should be defined as an increasing function of the discrimination power. This criterion is denoted S. The ROI $(\omega^* \times \tau^*)$ is estimated by exhaustively searching the largest value of $S(\omega_m, \tau_n)$ among all regions:

$$S(\omega^*, \tau^*) = \max \{S(\omega_m, \tau_n) \mid m \in \{1, 2, ..., M\}, n \in \{1, 2, ..., N\}\} \tag{6.4}$$

The exhaustive search is reasonable with the chosen values of M and N (1224 regions in our experiments, as detailed in Section 6.4.5).

Two time-frequency selection criteria are presented here. One is proposed based on neurophysiological background, called Time-frequency Discrimination Factor; the other is the F score (see Chapter 5 section **??**), which is purely based on machine learning.

6.3.1 Time-frequency Discrimination Factor

The proposed criterion for finding the ROI is based on two neurophysiological principles:

1. Imagination of one side of hand movement typically generates ERD in the opposite side of the brain, so that it is possible to discriminate between the imaginations of right and left hand movements by using bipolar electrodes placed over corresponding hand representation areas, i.e. C3 and C4 [156]. To achieve good classification performances, 1) the pattern difference between imaginations of left and right hand movements should exist in the selected time-frequency region (ROI) at each channel; 2) and the difference between C3 and C4 should also exist in the ROI for both motor imageries.

2. Electrophysiological studies have emphasized the role of volume conduction, so that neural activities in one area are distributed on multiple electrode positions [150]. Due to this effect, the signals of some undesirable EEG rhythms (i.e. common components) are also recorded, and mixed with the specific signals of different hand movements, which may deteriorate the classification results [37]. Although bipolar recording can eliminate this effect to some extent, it cannot completely remove all of those common components. Thus, we should consider the influence of those common components in selecting the ROI.

In BCI signal classification, ERD patterns are often estimated by the logarithm of the variance of band-pass filtered EEG in a specific time interval, so called logarithmic band power (BP) estimator [213]. The variance of EEG segment in the time domain for each trial i and each channel e is computed as:

$$v_e(i) = \frac{1}{T-1} \sum_{j=t_n}^{t_n+T-1} (x_{ij} - \bar{x}_i)^2 \tag{6.5}$$

where x_{ij} the j-th sample in the time interval $\tau_n = [t_n, t_n + T - 1]$ of the i-th trial of the ω_m-bandpass filtered EEG data, and $\bar{x}_i$ is the mean value over all samples of filtered EEG in the time interval τ_n of the i-th trial. Then, the band power feature in each channel is defined as:

$$BP_e(i) = \log(v_e(i)) \tag{6.6}$$

The logarithm is applied to make the distribution of BP features approximately normal, so as to feed the linear classifier, *Fisher's* LDA.

According to this definition, the overall BP, $\widetilde{BP}_e^{\chi}$, for each class ($\chi = L,\ R$) and each channel ($e =$ C3, C4) is defined by taking the logarithm of the median or the mean of data variances over trials [236]. Here, we use the median value, because it is more robust to outliers. The overall BP then writes:

$$\widetilde{BP}_e^{\chi} = \log\left(\tilde{v}_e^{\chi}\right) \tag{6.7}$$

where $\tilde{v}_e^{\chi}$ denotes the median of data variances $v_e(i)$ over all trials for class χ.

Thus, the pattern difference (PD_e) between two conditions (left vs. right hand) in a time-frequency region ($\omega_m \times \tau_n$) in each channel is expressed as:

$$PD_{C3}(\omega_m, \tau_n) = \widetilde{BP}_{C3}^{L}(\omega_m, \tau_n) - \widetilde{BP}_{C3}^{R}(\omega_m, \tau_n) \tag{6.8}$$

$$PD_{C4}(\omega_m, \tau_n) = \widetilde{BP}_{C4}^{L}(\omega_m, \tau_n) - \widetilde{BP}_{C4}^{R}(\omega_m, \tau_n) \tag{6.9}$$

The sign of PD_e reflects the tendency (increase or decrease) of the BP modulation from condition L (imagination of left hand movement) to condition R (imagination of right hand movement) in channel e.

Imagination of left and right hand movements usually elicit contrary contralateral dominance of ERD at channels C3 and C4 [156, 152]. These task-related spatial discriminative modulations can be measured by $|PD_{C3}(\omega_m, \tau_n) - PD_{C4}(\omega_m, \tau_n)|$, called *discriminative force* $Fd(\omega_m, \tau_n)$, to estimate this positive contribution in a time-frequency region ($\omega_m \times \tau_n$). A large $Fd(\omega_m, \tau_n)$ indicates that large discriminative modulations occur in the time-frequency region ($\omega_m \times \tau_n$).

On the other hand, it has been proved that other sources (non-target motor imagery sources) will generate signals (e.g. α-rhythm from the visual cortex) in the same frequency as ERD during the motor imagery (for details, see [37, 108]). For example, subjects are looking at the screen during both motor imagery tasks, which can generate visually related common modulations at C3 and C4. Although these sources are not near C3 and C4, they will conduct through scalp and be mixed with discriminative components because of the volume conduction [144]. Meanwhile, neural activities at C3 and C4 will also affect the contralateral channels due to volume conduction. These are what we call common components. They overlap with the discriminative modulations, which present a negative effect on the classification. Thus, we define the *blurring force* $Fb(\omega_m, \tau_n) = |PD_{C3}(\omega_m, \tau_n) + PD_{C4}(\omega_m, \tau_n)|$, to estimate those common modulations in the time-frequency region ($\omega_m \times \tau_n$). A small $Fb(\omega_m, \tau_n)$ indicates that small common modulations happen in the time-frequency region ($\omega_m \times \tau_n$).

Finally, a *Time-frequency Discrimination Factor*, $TFDF(\omega_m, \tau_n)$, is defined as the difference between $Fd(\omega_m, \tau_n)$ and $Fb(\omega_m, \tau_n)$ to evaluate the overall contribution of the data in the time-frequency area (ω_m, τ_n) from electrodes C3 and C4 for two-class discrimination:

$$\begin{aligned} TFDF(\omega_m, \tau_n) &= Fd(\omega_m, \tau_n) - Fb(\omega_m, \tau_n) \\ &= |PD_{C3}(\omega_m, \tau_n) - PD_{C4}(\omega_m, \tau_n)| - |PD_{C3}(\omega_m, \tau_n) + PD_{C4}(\omega_m, \tau_n)| \end{aligned} \tag{6.10}$$

An ideal time-frequency region for classification should have large discriminative modulations (large $Fd(\omega_m, \tau_n)$) and small common modulations (small $Fb(\omega_m, \tau_n)$), so the ROI

$(\omega^* \times \tau^*)$ is estimated by seeking the maximum value of $TFDF(\omega_m, \tau_n)$ among the given $M \times N$ time-frequency regions:

$$TFDF(\omega^*, \tau^*) = \max \{TFDF(\omega_m, \tau_n) \mid m \in \{1, 2, ..., M\}, n \in \{1, 2, ..., N\}\} \quad (6.11)$$

To examine the behavior of $TFDF$, we provide its possible values in Table 6.1 for different cases, and present in Figure 6.2 the ERD/ERS maps and the corresponding $TFDF$ values of 4Hz, 2s-wide time-frequency regions for an example from a standard dataset. From Table 6.1, we can see that 1) the values of $TFDF$ are larger for $PD_{C3} \cdot PD_{C4} < 0$ than for $PD_{C3} \cdot PD_{C4} \geq 0$, 2) the values of $TFDF$ are determined by $\min \{|PD_{C3}|, |PD_{C4}|\}$. Thus, this method tends to seek the time-frequency region where PD_{C3} and PD_{C4} have different signs and large absolute values.

Table 6.1: Values of $TFDF$ for different pairs of PD_{C3} and PD_{C4}.

	$\lvert PD_{C3}\rvert > \lvert PD_{C4}\rvert$	$\lvert PD_{C3}\rvert \leq \lvert PD_{C4}\rvert$
$PD_{C3} \cdot PD_{C4} \geq 0$	$TFDF = -2\lvert PD_{C4}\rvert$	$TFDF = -2\lvert PD_{C3}\rvert$
$PD_{C3} \cdot PD_{C4} < 0$	$TFDF = 2\lvert PD_{C4}\rvert$	$TFDF = 2\lvert PD_{C3}\rvert$

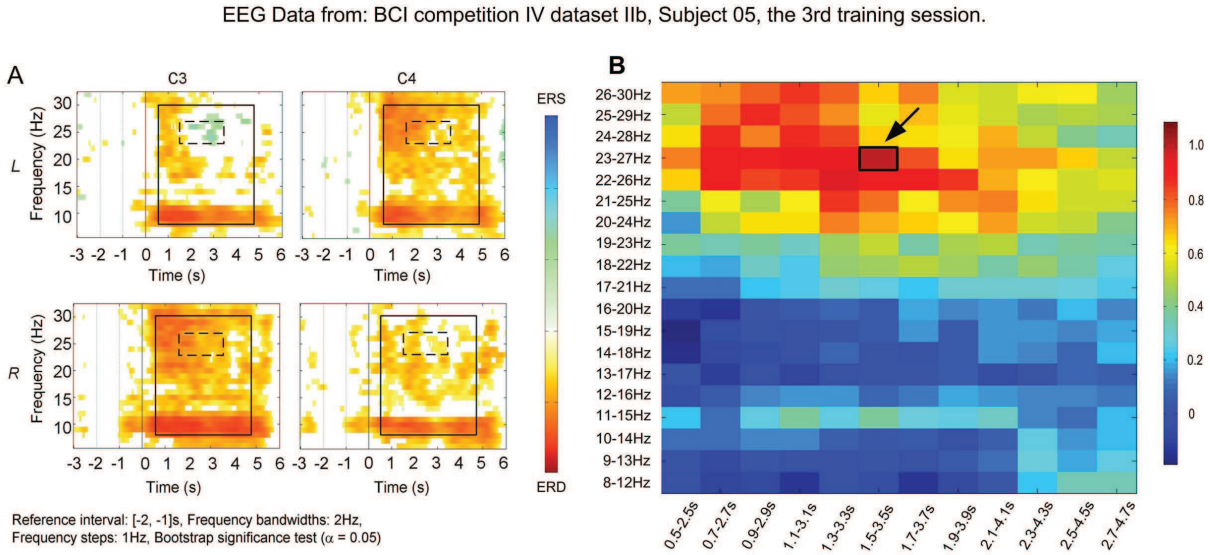


Fig. 6.2: Maps of ERD/ERS and $TFDF$ values for a subject. (A) ERD/ERS maps of one example: time-frequency selection was performed within the large rectangle (solid line). The small rectangle (dashed line) shows the time-frequency region with the largest $TFDF$ value. (B) $TFDF$ values of the time-frequency regions with 4Hz-wide frequency bands and 2s-wide time segments. The largest value is marked out by a small rectangle.

In the ROI, the right hand motor imagery elicits more significant ERD at C3 compared to the left hand motor imagery, which leads to $\tilde{P}^L_{C3}(\omega^*, \tau^*) > \tilde{P}^R_{C3}(\omega^*, \tau^*)$, while left hand motor imagery generates more significant ERD at C4 compared to the right hand motor imagery, so we have $\tilde{P}^L_{C4}(\omega^*, \tau^*) < \tilde{P}^R_{C4}(\omega^*, \tau^*)$. Thus,

$$PD_{C3}(\omega^*, \tau^*) = \tilde{P}^L_{C3}(\omega^*, \tau^*) - \tilde{P}^R_{C3}(\omega^*, \tau^*) > 0 \quad (6.12)$$

$$PD_{C4}(\omega^*, \tau^*) = \tilde{P}^L_{C4}(\omega^*, \tau^*) - \tilde{P}^R_{C4}(\omega^*, \tau^*) < 0 \quad (6.13)$$

These different signs of PD_{C3} and PD_{C4} reflect the spatial difference of significant ERD between right and left hand motor imageries. On the other hand, large absolute values of PD_e represent the large magnitudes of task-related (i.e. right vs. left hand) difference at channel e (i.e. C3, C4), which also contributes to the discrimination between two tasks.

In the literature, a broad frequency band (i.e. 8-30Hz) EEG segments (0.5-2.5s after cue on-set) was typically chosen for feature extraction, because it covers the μ and β bands and usually generates good classification results [135]. For this data example, the frequency band (23-27Hz) of the ROI selected by $TFDF$ is in the range of β band (18-25Hz) but does not completely cover it, and the time segment (1.5-3.5s) is different from the typically used one.

Figure 6.4(A) and (B) show the distributions of the BP features extracted from the time-frequency region with the largest $TFDF$ and of the ones extracted from the recommended broad frequency band EEG segment in a real data example. The linear separation boundary is obtained by *Fisher's* LDA in the figure. From the figure, we can see that the BP features extracted from the time-frequency region with the largest $TFDF$ seem more linearly separable than the ones extracted from the recommended broad frequency band EEG segment. The comparison on classification results will be made in the result section to assess the contribution of $TFDF$ to the discrimination between left and right hand motor imageries.

6.3.2 The F score

The F score proposed in the previous chapter is for estimating the discrimination power of a group of features:

$$\hat{F} = \frac{\left\|\vec{\mu}^L - \vec{\mu}^R\right\|_2^2}{tr(\Sigma^L) + tr(\Sigma^R)} \tag{6.14}$$

where Σ denotes the covariance matrix of the feature vector, $\vec{\mu}$ denotes the mean of the feature vector, $\|\cdot\|_2$ denotes the $L2$-norm (Euclidean norm), and $tr(\cdot)$ the trace of a matrix.

In this chapter, the BP features $[BP_{C3}(i), BP_{C4}(i)]$ (defined in Equation 2.13) extracted from the time-frequency ROI are used for classification, so it is a two-dimensional feature space. We adapt the F score, $\hat{F}$, to estimate the separation between left hand vs. right hand motor imagery in this feature space:

$$\hat{F} = \frac{(\overline{BP}_{C3}^L - \overline{BP}_{C3}^R)^2 + (\overline{BP}_{C4}^L - \overline{BP}_{C4}^R)^2}{\bar{S}_{C3}^L + \bar{S}_{C4}^L + \bar{S}_{C3}^R + \bar{S}_{C4}^R} \tag{6.15}$$

with:

$$\overline{BP}_e^\chi = \frac{1}{K^\chi}\sum_{i=1}^{K^\chi} BP_e^\chi(i) \tag{6.16}$$

$$\bar{S}_e^\chi = \frac{1}{K^\chi - 1}\sum_{i=1}^{K^\chi} (BP_e^\chi(i) - \overline{BP}_e^\chi)^2 \tag{6.17}$$

where K^χ the number of trials for class χ ($\chi \in \{L, R\}$).

We calculate the value of this criterion, $\hat{F}(\omega_m, \tau_n)$, for each time-frequency region $(\omega_m \times \tau_n)$, so as to measure whether $(\omega_m \times \tau_n)$ contains the most discriminative information. The

time-frequency ROI (ω^*, τ^*) is estimated by seeking the maximum value of $\hat{F}(\omega_m, \tau_n)$ among all $M \times N$ time-frequency regions:

$$\hat{F}(\omega^*, \tau^*) = \max\left\{\hat{F}(\omega_m, \tau_n) \mid m \in \{1, 2, ..., M\}, n \in \{1, 2, ..., N\}\right\} \tag{6.18}$$

Note that outliers are taken into account in the $TFDF$ calculation but not in the F score. The reason is that some outliers (which may be caused by muscle movement of one side of face/body during the experiments) may increase the difference between the two sides of brain [37]. In this case, the $TFDF$ value will abnormally increase when using the mean value. As a result, the time-frequency area contaminated by noise may be selected by error, which may deteriorate classification results when using *Fisher's* LDA. On the contrary, for the F score, outliers will increase the intra-class variance, so they will lower the F score. Thus, the time-frequency area contaminated by noise will not be selected (due to a low F score value), though we do not account for the outliers in the calculation of the F score.

As opposed to the $TFDF$, the F score is purely based on statistical characteristics of the features, regardless of neurophysiological phenomena linked to a specific task. So it can be applied in the absence of prior knowledge about task-related neural response.

Figure 6.3(A) and (B) show the F score and the Euclidean distance between two classes $\left\|\vec{\mu}^L - \vec{\mu}^R\right\|_2$ (which reflects only inter-class difference between two classes) in the time-frequency regions with 4Hz-wide frequency bands and 2s-wide time segments for the data example. The large values of F score mainly appear in the frequency band over 20Hz for this example, which is quite similar to the distribution of Euclidean distance. However, the maximum value appears in different regions. Figure 6.4(C) and (D) show the distributions of the BP features extracted from the time-frequency regions with the largest F score, and with the largest Euclidean distance, respectively. The linear separation boundary is also obtained by *Fisher's* LDA in the figure. From Figure 6.4, we can see that the features extracted from the time-frequency region with the largest F score and the largest Euclidean distance are more linearly separable than the ones extracted from the broad frequency band EEG segment when using *Fisher's* LDA as the classifier. Compared to the features from the time-frequency region with the largest Euclidean distance, the intra-class difference is smaller for the features from the time-frequency region with the largest F score. As it considers the intra-class difference, the F score is more reasonable than the Euclidean distance as a two-class separation measurement. The overlap area between two classes is smaller for the F score than for the $TFDF$, indicating that the F score might be more effective than the $TFDF$ in seeking the optimal time-frequency regions for feature extraction. In the result section, the analysis of classification performance with respect to the two criteria is provided.

6.4 Data description and preprocessing

In this chapter, we used data of the BCI competition IV dataset IIa [39] and IIb [100]. Dataset IIa was recorded in a multi-channel monopolar setting (22 monopolar channels). The parameters of bipolar channels can be adapted to the experiments. Dataset IIb was recorded over three bipolar channels C3, Cz and C4. Details of these two datasets are provided below.

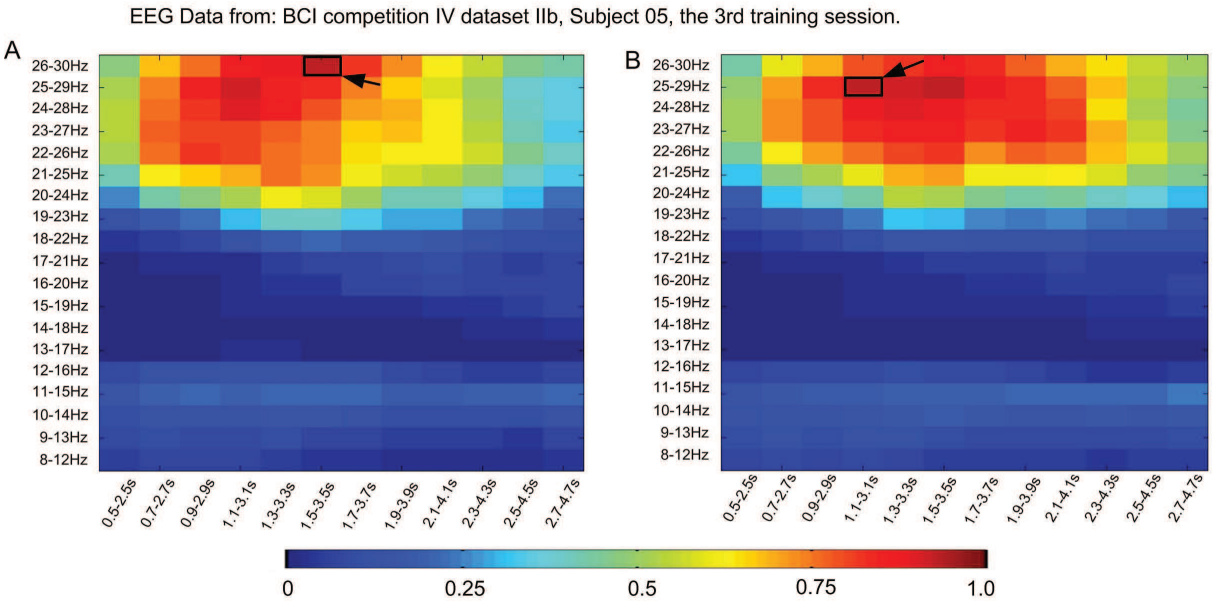


Fig. 6.3: Maps of the F score and Euclidean distance for the same subject in Figure 6.2. (A) F score of the time-frequency regions with 4Hz-wide frequency bands and 2s-wide time segments. The largest value is marked out by a small rectangle. (B) Euclidean distance of the time-frequency regions with 4Hz-wide frequency band and 2s-wide time segments. The largest value is marked out by a small rectangle.

6.4.1 Dataset IIa

As we have introduced this dataset in Chapter 4, we only repeat some key information here. This dataset contains one training session and one evaluation session of EEG data from 9 subjects who performed four classes of cue-driven motor imagery tasks (left hand, right hand, both feet and tongue). The EEG signals were recorded by 22 Ag/AgCl electrodes (with inter-electrode distances of 3.5cm). The electrode montage is shown in Figure 6.5(A). For extracting a bipolar channel at the position of C3 or C4, three different pairs of electrodes can be used, marked as a, p and l in Figure 6.5(A). Thus, nine possible channel combinations for C3 and C4 are generated for time-frequency optimization.

6.4.2 Dataset IIb

This dataset consists of two classes (left vs. right hand) cue-driven motor imagery BCI data from 9 subjects. The EEG data are recorded in 3 bipolar channels, i.e. at positions C3, Cz and C4 (see Figure 6.5(B)). The distances between the two bipolar electrodes forming a channel are pre-determined in this dataset (for details, see [100]). For each subject, 5 sessions are provided, including 3 training sessions and 2 evaluation sessions. The first two training sessions consist of 240 single trials (120 single trials per session) totally without visual feedback. Each trial started with a fixation cross and an additional short acoustic warning tone. Later a visual cue was given to guide the subject to execute the corresponding imagination of hand movement over a period of 4 seconds. The last training session (160 single trials) and both evaluation sessions (160 trials per session) were recorded with visual feedback from 0.5 to 4.5s after the cue on-set (for details, see [100]).

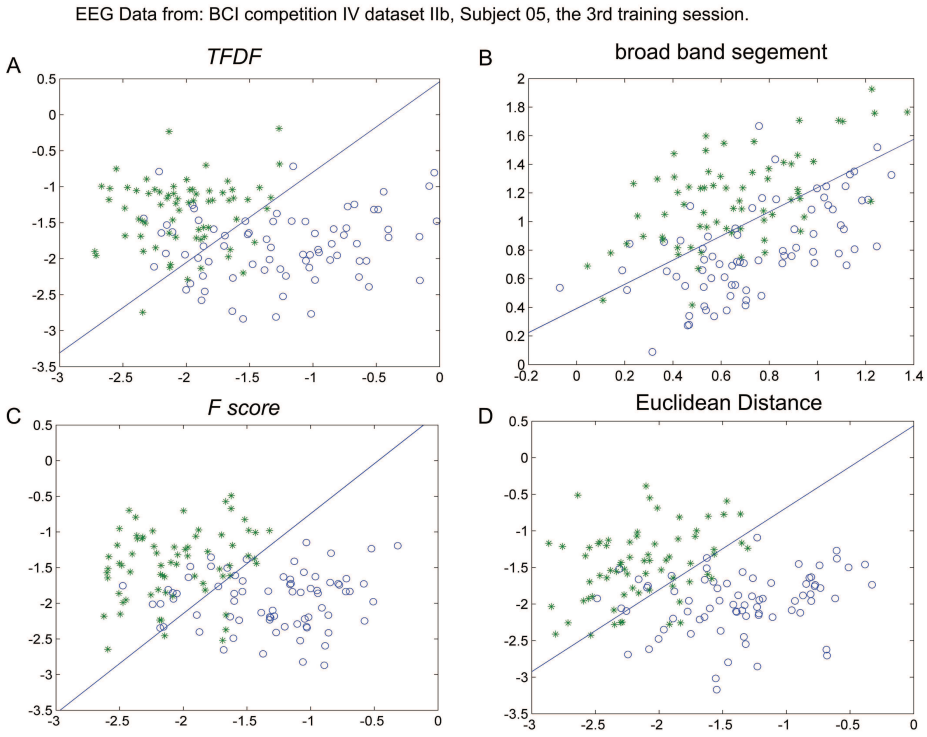


Fig. 6.4: Distributions of BP features extracted from different time-frequency regions for two classes (o represents left hand, * represents right hand). (A) Distribution of BP features extracted from the time-frequency region with the largest $TFDF$. (B) Distribution of BP features extracted from a time-frequency region (8-30Hz, 0.5-2.5s after cue on-set) in the μ band. (C) Distribution of BP features extracted from the time-frequency region with the largest F score. (D) Distribution of BP features extracted from the time-frequency region with the largest Euclidean distance. The line shows the best linear separation boundary in each case

6.4.3 Experimental goals

The proposed time-frequency optimization methods based on different criteria will first be applied on dataset IIb using only two bipolar channels (C3, C4). The goal of this experiment is to evaluate the effectiveness of the methods in improving the performances of BCI based on few channels only. We first train the methods for each subject on the training data, and then evaluate them on the testing data for this subject. This is so called session-to-session transfer. The results of our methods in session-to-session transfers from training sessions to testing sessions will be compared with the winners on this dataset in BCI competition IV. Then, we will apply our methods on two bipolar channels (C3, C4) selected from the 22 monopolar channels of dataset IIa. The classification results obtained on dataset IIa will be compared with those obtained by CSP algorithms using 22 monopolar channels, to evaluate the interest of the method for electrode reduction.

6.4.4 Visualization of ERD/ERS maps

The time-frequency maps of ERD/ERS for both left (L) and right (R) hands in the bipolar channels C3 and C4 were generated by the Biosig Toolbox using overlapping 2Hz bands

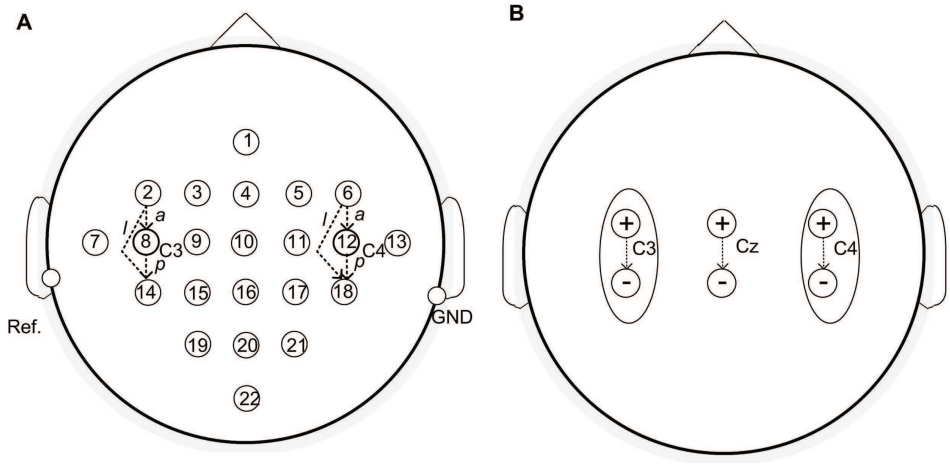


Fig. 6.5: The montage of electrodes in the BCI competition IV dataset IIa [39] (A) and dataset IIb [100] (B). (A): The arrows between the EEG electrodes show the three possible bipolar derivation pairs for bipolar recording of C3 and C4: a, p and l. (B): The arrows between the EEG electrodes show the bipolar derivation, where $\oplus$ is the signal electrode and $\ominus$ is the reference. The distance between two bipolar electrodes forming a channel for each subject is pre-determined (for details, see [100]). Only C3 and C4 (marked by ellipses) are used in this study for discrimination between imaginations of left and right hand movements.

(step = 1Hz) in the frequency range between 6 and 32Hz [176]. The statistical significance of the power increases (ERD) and decreases (ERS) was verified by a t-percentile bootstrap test with the confidence interval of $\alpha = 0.05$ [67]. Only the significant ERD/ERS are shown in the maps.

6.4.5 Data preprocessing for time-frequency optimization

For each bipolar channel, 5th order Butterworth filters are applied to compute 19 successive 4Hz-wide frequency bands of signals (F = 4Hz, F_s = 1Hz): 8-12Hz, 9-13Hz, 10-14Hz, ..., 26-30Hz, and 15 successive 8Hz-wide frequency bands of signals (F = 8Hz, F_s = 1Hz): 8-16Hz, 9-17Hz, 10-18Hz, ..., 22-30Hz. Then 36 overlapping time segments in each frequency band were obtained through 2s, 2.5s and 3s-wide (i.e. T = 2, 2.5 and 3s, respectively) sliding windows (12 segments for each sliding window) with 0.2s-step (i.e. T_s = 0.2) moving from 0.5s after the cue on-set. Those parameters are set based on the experience from competitors reported in BCI competition IV [31] and led to good performances in our previous work [236]. Therefore, there are $(19 + 15) \times 36 = 1224$ time-frequency areas for subject-specific selection. The numbers of different size time-frequency regions are listed in Table 6.2. It has to be mentioned the selection procedure is not very time-consuming (2 mins 21 seconds in average, using Matlab 7.10.0 on Window 7 Professional 64bits, CPU 2.66GHz, RAM 2.0G), and is done offline, so the computational cost is acceptable.

Table 6.2: Numbers of different size time-frequency regions ($F_s = 1$Hz, $T_s = 0.2$s). Total = 1224 regions.

		F(Hz)	
		4	8
T(s)	2	228	180
	2.5	228	180
	3	228	180

6.5 Experimental results

In the experiments, the paired-sample t-test (see Appendix B) was employed to reveal the statistical significance of the difference between the performances of different methods. The test rejects the null hypothesis at the 0.05 significance level.

6.5.1 Improving classification performance for dataset IIb

The time-frequency ROI ($\omega_{ROI} \times \tau_{ROI}$) of each training dataset is obtained by maximizing the $TFDF$ and the F score criteria, respectively. Results are reported in Table 6.3. These results show that 1) the estimated time-frequency ROIs vary among different subjects; 2) even for the same subject, the estimated ROIs vary among different training sessions; 3) the two criteria picked out different ROIs for the same training session. The fact that the estimation results depend on the subjects is also reflected in the individual differences of timing and frequency of ERD/ERS patterns. Even for the same subject, the timing and frequency of ERD/ERS may shift across sessions [156], which leads to the intra-subject difference in the estimation of ROIs between sessions. A typical example of time-frequency maps displaying significant ERD (red) and ERS (blue) in a training session (Session 3) for a subject (Subject 6) in the dataset is shown in Figure 6.6. The ROIs estimated by the $TFDF$ are marked out by solid rectangles (10-14Hz, 0.7-2.7s), while the ROIs selected by the F score are displayed as dashed rectangles (11-15Hz, 1.1-4.1s). Although the ROIs estimated by the two criteria are different, both ROIs contain discriminative ERD patterns between the two classes, indicating that these two criteria could successfully capture the discriminative part of the signal.

To evaluate the contribution of the proposed time-frequency optimization to classification, 10 repetitions of cross-validation are performed on each training session for each subject, using the BP features extracted from the estimated time-frequency ROIs by $TFDF$ and F score respectively. In each run, we randomly separated the data into calibration (90%) and test (10%) sets, and classified the test data using the Fisher's LDA obtained from the calibration set. The classification accuracy (Acc) is defined as the observed agreement between classification outputs and true labels [178]. The cross-validation results are obtained by averaging Acc over 10 runs.

The results are compared to those obtained by using the BP features from the broad band (8-30Hz) EEG segments (0.5-2.5s) with or without CSP filtering. The number of spatial filters used in CSP-based classification for this dataset is 2 (1 pair), because only three bipolar channels (C3, Cz and C4) are provided in this dataset. The comparisons between different methods are shown in Figure 6.7 using scatter plots. We can see that using the

Table 6.3: Time-frequency ROIs selected by the $TFDF$ and F score criteria, respectively, for each training session of each subject on dataset IIb.

Subject	Session	ω^*(Hz)		τ^*(s)	
		$TFDF$	F score	$TFDF$	F score
1	1	10-14	11-15	2.9-4.9	0.9-3.4
	2	10-14	10-14	2.3-4.3	0.9-3.9
	3	10-14	17-25	1.5-3.5	0.9-3.9
2	1	18-22	11-19	1.5-3.5	2.1-4.1
	2	11-15	18-22	2.7-5.2	1.3-4.3
	3	11-15	9-13	2.7-5.2	2.9-5.4
3	1	8-12	12-16	0.5-3.0	2.3-4.3
	2	9-13	21-25	0.5-2.5	2.9-5.4
	3	8-12	11-15	3.5-6.5	2.1-5.1
4	1	8-12	18-26	0.5-2.5	0.7-2.7
	2	9-13	10-18	0.9-2.9	0.7-2.7
	3	8-12	8-12	0.9-2.9	0.9-2.9
5	1	26-30	22-30	0.9-2.9	0.7-2.7
	2	26-30	22-30	0.7-3.2	0.5-3.0
	3	23-27	22-30	1.5-3.5	0.5-3.5
6	1	11-15	11-15	1.3-3.8	1.9-4.4
	2	10-14	11-15	1.7-3.7	0.7-3.7
	3	10-14	11-15	0.7-2.7	1.1-4.1
7	1	12-16	12-16	2.7-4.7	2.1-5.1
	2	12-16	11-15	3.1-5.1	2.9-5.4
	3	11-15	12-16	1.1-3.1	0.9-2.9
8	1	11-15	9-13	1.3-3.3	1.5-4.5
	2	9-13	13-17	1.1-3.1	1.1-3.6
	3	10-14	8-16	0.9-2.9	1.3-3.3
9	1	22-26	21-25	1.9-4.4	1.1-3.1
	2	9-13	8-12	3.1-5.1	3.1-6.1
	3	22-26	23-27	0.9-2.9	0.9-3.4

time-frequency ROI estimated by each criterion can greatly improve the accuracy in most sessions compared to using a broad band EEG segment with or without CSP filtering. Using the F score generates higher accuracy than using the $TFDF$ for most sessions (63.0%), indicating that the F score may be more effective than the $TFDF$ in selecting optimal time-frequency regions for discrimination. In addition, we also observe that using classic CSP filtering generates the worst results in the cross-validations, indicating that classic CSP filtering is not very useful for the data recorded by a very few number of electrodes.

To further examine the contributions of these two criteria, session-to-session transfers are performed using the training session which has the best classification result in the cross-validation for each subject. As the independent evaluation data are recorded on a different day than the training sessions, EEG signals of the subjects may change significantly from the training data to the evaluation data. This test aims at evaluating the robustness of the methods to non-stationary signals.

In this test, the classifier is parameterized from the selected training session using the BP features from the corresponding ($\omega^* \times \tau^*$). The ω^*-bandpass filtered EEG segments with

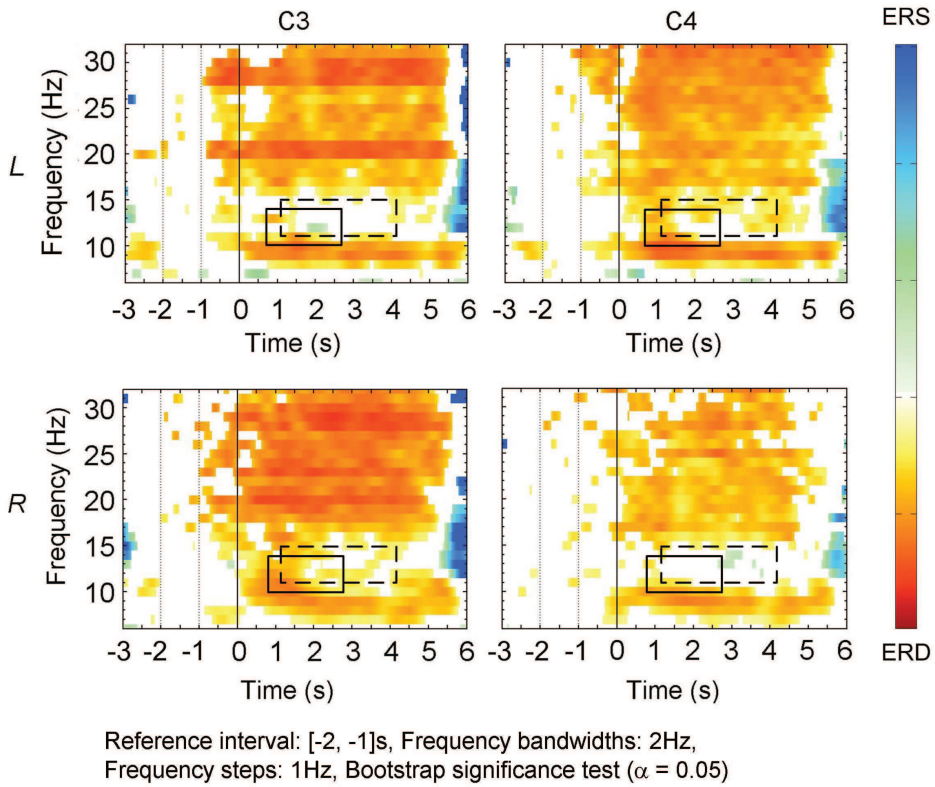


Fig. 6.6: Time−frequency maps displaying significant ERD (red) and ERS (blue) for Subject 6 (a typical example) in BCI competition IV IIb. The areas in the rectangles are the time-frequency ROI selected by the proposed methods based on $TFDF$ (solid line) and F score (dashed line), respectively.

the same time length as τ^* (i.e. T) are obtained from each entire single-trial of testing data via a 0.2s step sliding window to generate continuous classification outputs (see Figure 6.8). According to the BCI competition requirement, the classification performances in the session-to-session transfers are measured by *kappa coefficient* [178]:

$$\kappa = (Acc - P_e)/(1 - P_e) \tag{6.19}$$

where P_e is the chance level for agreement (i.e. $P_e = 0.5$ for two-class problems, so here $\kappa = 2Acc - 1$). Thus, a larger κ value indicates a better classification performance. The mean kappa value over all subjects of the dataset is denoted by $\bar{\kappa}$.

For this dataset, six BCI groups have reported their results of session-to-session transfers in the BCI competition [194]. We provide information on the methods of the first three winners in Table 6.4, since their classification results are better than the other three. The 1^{st} place winner used Filter Bank CSP (FBCSP) [5], which we have briefly introduced in the introduction section. Except our method based on $TFDF$ [236], no one has generated better results than FBCSP on this dataset until now [194]. The 2^{nd} place winner employed Common Spatial Subspace Decomposition (CSSD) with frequency band and time segment selections [194]. The 3^{rd} place winner applied CSP on spectrally filtered neural time series prediction preprocessing (NTSPP) signals [194]. Note that these methods involved frequency and/or time optimization process(es). The results of session-to-session transfers for all methods are shown in Table 6.5. $TFDF$ generates the best mean κ value ($\bar{\kappa} = 0.62$)

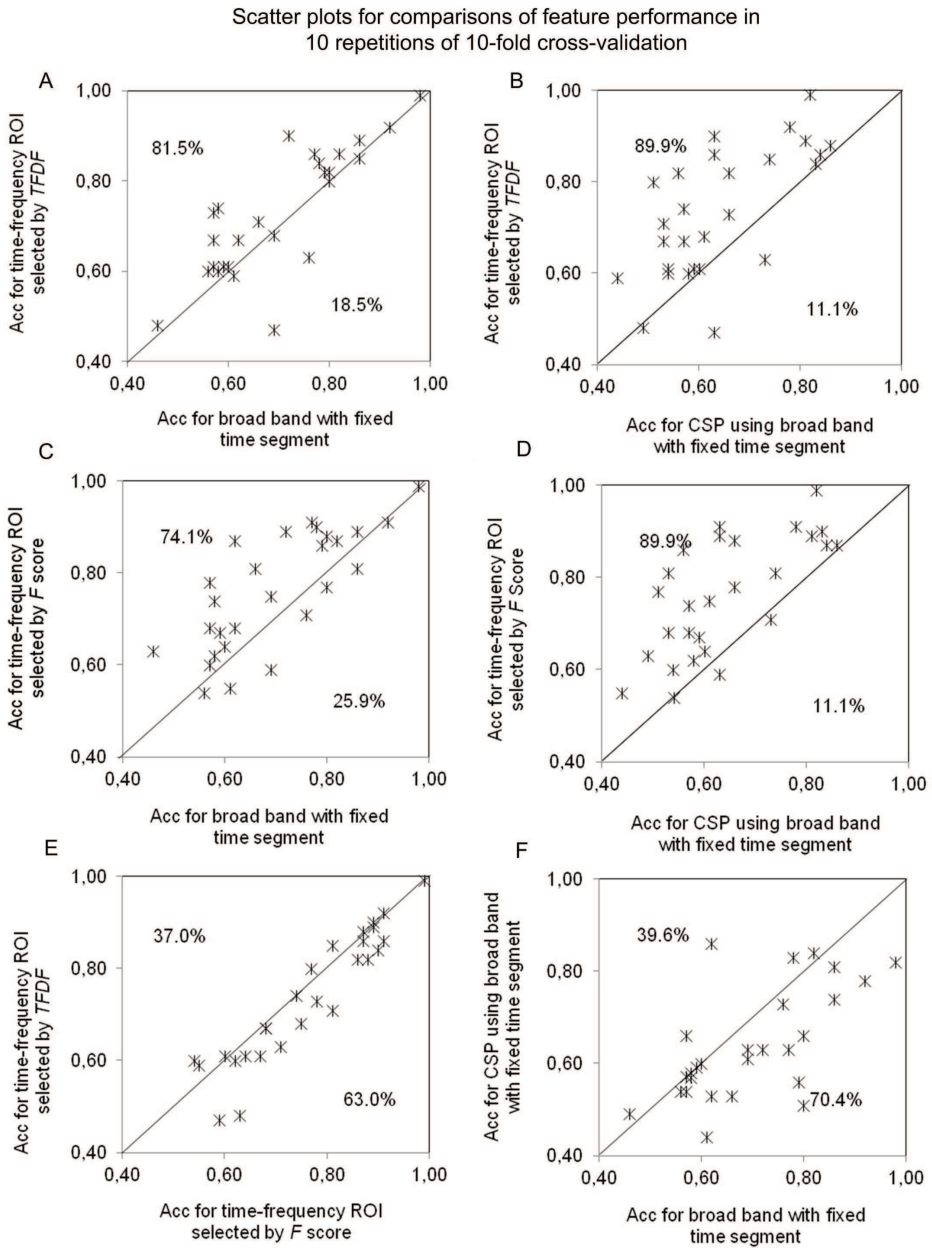


Fig. 6.7: Comparison of method performances in the cross-validation on dataset IIb. (A), (B) Scatter plots of classification accuracies (Acc) obtained by using $TFDF$ in time-frequency selection, vs. those obtained by broad band (8-30Hz) EEG in a fixed time segment (0.5-2.5s) without and with CSP, respectively. (C), (D) Scatter plots of Acc obtained by using F score in time-frequency selection, vs. those obtained by broad band (8-30Hz) EEG in a fixed time segment (0.5-2.5s) without and with CSP, respectively. (E) Scatter plots of Acc obtained by using $TFDF$ vs. those obtained by using F score in time-frequency selection. (F) Scatter plot of Acc obtained by broad band (8-30Hz) EEG in the fixed time segment (0.5-2.5s) with CSP filtering vs. those obtained by the same EEG but without CSP filtering. For the points above the diagonal in each scatter plot, the method in y-axis outperforms the method in x-axis in the cross-validation on the corresponding training session.

among all methods in the independent evaluation. Although the improvements of κ values

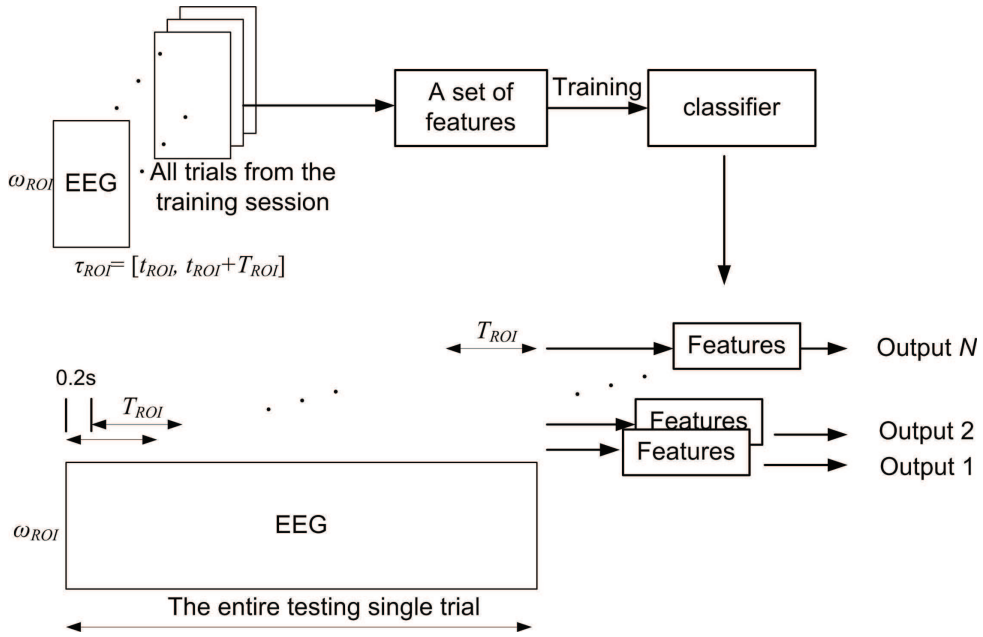


Fig. 6.8: Strategy of session-to-session transfers for BCI competition IV dataset IIa and IIb.

yielded by $TFDF$ compared to the 1^{st} place winner ($\bar{\kappa} = 0.60$, $p = 0.12$) and the 2^{nd} place winner ($\bar{\kappa} = 0.58$, $p = 0.18$) are not statistically significant due to the limited number of subjects, $TFDF$ outperforms the 1^{st} place winner for 6 out of 9 subjects (except Subjects 4, 5 and 8), and the 2^{nd} place winner for 6 out of 9 subjects (except Subjects 4, 7 and 9) too. The mean kappa value obtained by the F score ($\bar{\kappa} = 0.60$) is relatively lower than the one obtained by $TFDF$ (not significantly, $p = 0.29$), but comparable to the 1^{st} place winner, and higher than the 2^{nd} place winner (not significantly, $p = 0.52$). It has to be mentioned that the F score yields the best κ values for most subjects (4 subjects) among all methods. Further examination of the results show that the poor performance for Subject 3 led to a remarkable decrease in mean performance of the F score. In fact, the performances for Subject 3 are much poorer than those for other subjects for all methods, so that results averaged over all subjects might not be representative. Both time-frequency criteria ($TFDF$ and F score) yield better performances than the 3^{rd} place winner ($\bar{\kappa} = 0.46$, both $p < 0.01$), and those obtained by broad band (8-30Hz) EEG segments (0.5-2.5s) with CSP ($\bar{\kappa} = 0.41$, both $p = 0.01$) and without CSP ($\bar{\kappa} = 0.53$, both not significant by $p > 0.05$, even if $TFDF$ and F score outperform it for 7 and 5 out of 9 subjects, respectively). Thus, both criteria are promising for seeking optimal time-frequency patterns to improve classification performance of BCIs based on a few bipolar channels.

As all of the first three BCI competition winners have used all three bipolar channels (C3, Cz and C4) provided by the dataset, our methods not only generate good performances but also use less channels, which indicates that they may also be helpful for channel reduction. This potential contribution is validated in the next subsection.

Table 6.4: Comparison between our method and the first three winners on BCI competition IV dataset IIb.

	Time-frequency selection	Electrodes used	Features	Classifier
Our method	Selected by $TFDF$ or F score	C3, C4	BP	LDA
1^{st} [5]	Mutual information based selection	C3, Cz, C4	FBCSP	Naïve Bayes Parzen Window classifier
2^{nd} [194]	Selected by classification performance in cross-validation	C3, Cz, C4	CSSD	LDA
3^{rd} [194]	Selected by a heuristic search and a selection criterion based on overall classification accuracy in cross-validation.	C3, Cz, C4	CSP	Using the best classifier among 3 variants of LDA and 2 variants of SVM

Table 6.5: Performances of session-to-session transfers on BCI competition IV dataset IIb using the proposed time-frequency optimization methods with $TFDF$ and F score, broad band (8-30Hz) EEG in fixed time segments (0.5-2.5s) without CSP and with CSP, and 1^{st} to 3^{rd} winners in BCI competition. The best performance for each subject is shaded in grey. The last column provides the mean performance over all subjects.

	Subjects									
	1	2	3	4	5	6	7	8	9	Mean
$TFDF$	0.44	0.24	0.25	0.93	0.86	0.70	0.55	0.85	0.75	0.62
F score	0.39	0.25	0.13	0.93	0.88	0.63	0.55	0.88	0.78	0.60
without CSP	0.40	0.24	0.18	0.94	0.39	0.66	0.52	0.81	0.68	0.53
with CSP	0.28	0.13	0.11	0.47	0.56	0.13	0.58	0.76	0.67	0.42
FBCSP(1^{st}) [5]	0.40	0.21	0.22	0.95	0.86	0.61	0.56	0.85	0.74	0.60
CSSD(2^{nd}) [194]	0.43	0.21	0.14	0.94	0.71	0.62	0.61	0.84	0.78	0.58
NTSPP(3^{rd}) [194]	0.19	0.12	0.12	0.77	0.57	0.49	0.38	0.85	0.61	0.46

6.5.2 Electrode reduction for dataset IIa

In this dataset, the time-frequency ROIs are estimated by the two criteria, respectively, for nine possible channel combinations of C3 and C4 (see Figure 6.5, C3-C4: *a*-*a*, *a*-*p*, *a*-*l*, *p*-*a*, *p*-*p*, *p*-*l*, *l*-*a*, *l*-*p*, *l*-*l*). The scatter plot of classification accuracies (Acc) obtained by using the $TFDF$ vs. those obtained by using the F score in time-frequency optimization for all channel combinations and all subjects is shown in Figure 6.9. Using the F score generates higher accuracy than using the $TFDF$ for most cases (74.1%).

The optimal channel combinations are selected by comparing the classification accuracies (choosing the best one) among different combinations in the 10×10-fold cross-validation. Optimal channel combinations of C3-C4 and the corresponding estimated time-frequency ROI for different criteria and different subjects are listed in Table 6.6.

In session-to-session transfers, the optimal channel combinations are used. The classifier

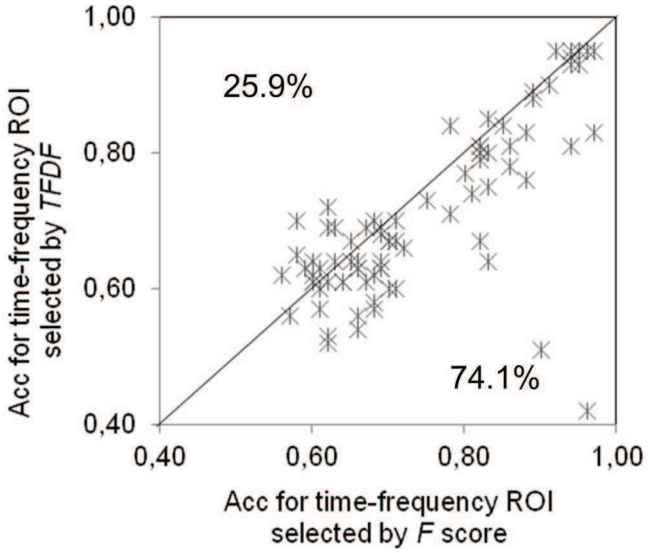


Fig. 6.9: Comparison between performances of $TFDF$ and F score in the cross-validation on dataset IIa.

Table 6.6: Optimal channel combinations of C3-C4 and the corresponding estimated time-frequency ROI for different criteria and different subjects on BCI competition IV dataset IIa.

Subject	C3-C4		ω^*(Hz)		τ^*(s)	
	$TFDF$	F	$TFDF$	F score	$TFDF$	F score
1	*a*-*a*	*a*-*l*	12-16	11-15	1.3-3.8	0.5-3.0
2	*l*-*l*	*a*-*l*	10-14	22-26	1.1-3.1	0.5-3.5
3	*a*-*a*	*a*-*a*	10-14	11-15	0.7-2.7	0.7-3.2
4	*p*-*p*	*a*-*a*	21-25	11-19	0.5-2.5	1.1-3.1
5	*p*-*l*	*p*-*a*	25-29	26-30	1.5-3.5	1.3-4.3
6	*l*-*l*	*l*-*p*	22-26	23-27	0.5-2.5	0.7-2.7
7	*a*-*a*	*a*-*l*	19-23	18-22	0.7-2.7	1.1-4.1
8	*l*-*l*	*l*-*p*	8-12	8-12	0.7-2.7	0.7-3.2
9	*a*-*l*	*p*-*a*	11-15	10-18	0.5-2.5	0.7-2.7

is obtained from the whole training session using the BP features from the corresponding $(\omega^* \times \tau^*)$. The ω^*-bandpass filtered EEG segments with the same time length as τ^* (i.e. T) are obtained from each entire single-trial of testing data via a 0.2s step sliding window to generate continuous classification outputs (see Figure 6.8).

As this study focuses on the two-class (right vs. left hand) problem, it is difficult to compare with BCI competition winners' results (reported based on a four-class problem including tongue and feet motor imagery data) on this dataset. Here, we compared the results obtained by our method with those obtained by FBCSP[1] [5], sparse CSP (SCSP) [10] and classic CSP, respectively. Note that FBCSP is believed to be an effective method that well solves the frequency and/or time optimization [8], which has achieved the best clas-

[1]The results of FBCSP on this dataset for right vs. left hand problem are provided by the BCI lab at Institute for Infocomm Research, Singapore, using all 22 mono-polar channels and the Naïve Bayes Parzen Window classifier.

sification performance on at least two datasets including dataset IIa in BCI competition IV [194]. SCSP is an optimized CSP that selects the least number of channels in CSP-based classification under a constraint of classification accuracy. SCSP has generated better performances than other channel reduction methods (based on the usual Fisher ratio, mutual information, SVM, CSP coefficients) and the regularized CSP (RCSP) on BCI competition IV dataset IIa for the right vs. left hand problem (for details, see [10]). The comparisons of classification results and the number of electrodes ($\#E$) used in classification between different methods are given in Table 6.7. As other researchers provided their classification results as classification accuracy values (Acc, defined in Section 6.5.1) for the right vs. left hand problem on this dataset, we also provide Acc values for the sake of comparison. Table 6.7 shows that all methods generate better mean performances than the classical CSP algorithm with all 22 mono-polar channels (mean classification accuracy, $\overline{Acc} = 77.26$), indicating the interest of time-frequency selection and electrode reduction. Our method based on F score ($\overline{Acc} = 79.67$) yields slightly better results than FBCSP ($\overline{Acc} = 79.17$) and SCSP ($\overline{Acc} = 79.07$) but using far less electrodes on this dataset: our method used only the two bipolar channels C3 and C4 (equivalent to 4 mono-polar channels), FBCSP used all 22 mono-polar channels, and SCSP used 8.55 mono-polar channels in average [10]. Further examination of individual results show that our method based on F score generates the best Acc for most subjects (4 subjects), indicating that it is the most effective on this dataset. Although the mean classification result of our method based on $TFDF$ ($\overline{Acc} = 78.00$) is slightly lower than those of FBCSP and SCSP, the differences are not statistically significant ($p > 0.05$). Comparing individual performances, our method based on $TFDF$ outperforms FBCSP and SCSP for 5 out of 9 subjects. Moreover, our method based on $TFDF$ also employs less electrodes than FBCSP and SCSP in the classification. Thus, the method based on $TFDF$ still meets the goal of electrode reduction without a significant drop of classification accuracy. Generally speaking, our method based on both criteria can effectively select time-frequency ROI for BCI classification based on only a few channels and therefore contributes to electrode reduction.

Table 6.7: Comparisons of classification results (Acc) and the number of electrodes ($\#E$) used in classification between our method, FBCSP [5], SCSP [10] and classical CSP on BCI competition IV dataset IIa. The best performance for each subject is shaded in gray.

		Subjects									
Method	$\#E$	1	2	3	4	5	6	7	8	9	Mean
$TFDF$	4	87.23	66.20	97.81	68.97	64.44	69.44	68.57	96.27	83.08	78.00
F score	4	89.36	69.01	97.81	66.38	66.67	72.22	68.57	97.01	90.00	79.67
FBCSP[1]	22	94.44	52.77	93.05	65.97	88.19	60.41	70.13	94.44	93.05	79.17
SCSP [10]	8.55	91.66	60.41	97.14	70.83	63.19	61.11	78.47	95.13	93.75	79.07
CSP [235]	22	83.51	56.53	97.50	70.00	54.50	62.49	84.50	95.57	90.77	77.26

6.6 Conclusion

Bipolar recording is a typical technique in BCI systems based on very few channels. This work addresses the time-frequency optimization for motor imagery BCIs, and emphasizes the possibility of using only a few bipolar channels to achieve comparable or even better

classification results than using multi-channel settings in motor imagery BCIs. The experimental results indicate the interest of our method for classification improvement and electrode reduction for motor imagery BCI. Thanks to its simplicity and few electrodes requirement, the proposed method is promising for portable BCI systems.

In this method, two alternative criteria, $TFDF$ and F score, are presented for measuring the discrimination power of each possible time-frequency region. Both criteria have their novelties and contributions.

Unlike other criteria, $TFDF$ measures the discrimination power of features based on foundational neurophysiologic phenomena, on which motor imagery BCI relies, considering both discriminative and common modulations instead of only the distribution of features. The classification results obtained by using $TFDF$ are better than those of state-of-the-art methods on BCI competition IV dataset IIb. The $TFDF$ can be adopted in other motor imagery BCI problems by placing the electrodes on the task-relevant areas (e.g. using C3 and Cz for discrimination between foot and right hand motor imagery).

The F score is a simple criterion that is easy to compute, and yields the best individual performances among all compared methods for most cases. As the F score does not require any prior knowledge of neurophysiology and yields the best individual performances among all compared methods for most cases, it may be more interesting for researches who purely focus on machine learning for BCI. Furthermore, it might be possible to apply this criterion on other problems outside BCI field.

In this chapter, the applications of our method focused on the most popular motor imagery BCI task: the discrimination between right and left hand motor imagery data. In the next chapter, this study should be extended to address time-frequency optimization for multi-class motor imagery BCIs.

Chapter 7

Time-frequency optimization for multi-class motor imagery data classification based on F score and Laplacian EEG derivation

7.1 Introduction

It is well known that motor imagery of hand movement desynchronizes EEG rhythms (ERD) in hand representation areas of the sensorimotor cortex (around C3 and C4). In the previous chapter, we have proposed a time-frequency optimization method based on either time-frequency discrimination factor (*TFDF*) or F score for discriminating between imagination of right and left band movements based on two bipolar EEG channels (C3 and C4). However, *TFDF* relies on specific knowledge of neurophysiology, which mainly works for the popular "hand movement discrimination" problem. It is not very easy to adapt to some other motor imagery BCI problems, like motor imagery of tongue, since they do not typically elicit a task-relevant ERD at the recording location. Unlike *TFDF*, F score is based on a machine learning method, i.e. Fisher's discriminant analysis, so that it can be used for seeking the optimal time-frequency areas for classifying other movements than hand movement. In this chapter, we extend the F score based time-frequency optimization to the multi-class case based on one-versus-rest (OVR) strategy [195].

Several studies, such as [152], showed that it is possible to classify four classes of motor imagery (left hand, right hand, foot and tongue) by using three Laplacian EEG channels C3, Cz and C4. The channel Cz is located at the center of the sensorimotor cortex, which is associated with foot movement (see Figure 7.1). Later, another study further proved that channel Cz with Laplacian derivation plays an important role in identifying motor imagery data of foot movement [189]. Although the tongue representation areas are located neither around Cz (foot representation area) nor around C3/C4 (hand representation areas), it typically elicits an enhancement of EEG rhythms (ERS) in hand representation areas, since these areas are completely relaxed when performing the tongue movement [152]. Thus, the three channels are enough for classifying these four classes of motor imagery in real applications.

Laplacian derivation typically needs five active EEG electrodes (one center electrode e and its four nearest-neighbor electrodes $j \in S_e$, where S_e is the set of the four nearest-neighbor electrodes) to derive measures in one "channel" (called Laplacian channel) (see

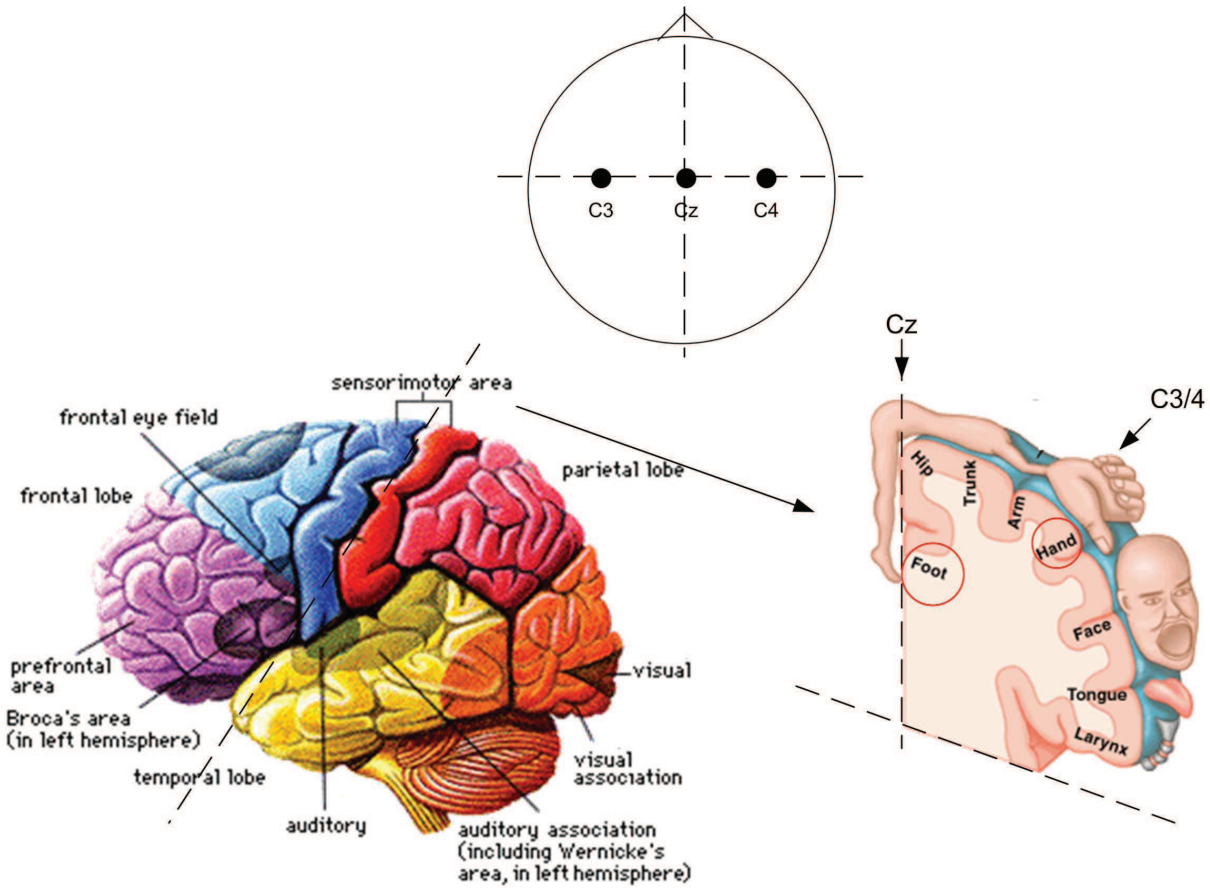


Fig. 7.1: Location of C3, Cz and C4 and the corresponding brain functional areas.

Figure 7.2 and Chapter 2 Section 2.6.3):

$$X_e^{LAP} = X_e - \frac{1}{4}\sum_{j \in S_e} X_j \tag{7.1}$$

where X_e is the EEG signal recorded at electrode e and X_e^{LAP} is the derived signal in Laplacian channel e. Theoretically, for a single channel, Laplacian derivation uses more electrodes than bipolar recording, which only needs two active electrodes $e1$ and $e2$ to build a bipolar derivation pair (see Figure 7.2):

$$X_e^{BI} = X_{e1} - X_{e2} \tag{7.2}$$

where X_{e1} and X_{e2} are the EEG signals at the two active electrodes in a bipolar derivation pair, and X_e^{BI} is the derived signal in the bipolar channel e. However, in practice, bipolar recording usually needs some additional electrodes and computational time for finding the optimal bipolar derivation pair (for details, see the previous chapter and [100, 108]). The electrodes and computational cost typically increase with the number of classes, so that bipolar recording is seldom used in multi-class problems.

Thus, we performed the F score based multi-class time-frequency optimization using three Laplacian EEG channels C3, Cz and C4. The method is applied on two standard BCI competition datasets for classifying four classes of motor imagery data (left hand, right hand, foot and tongue).

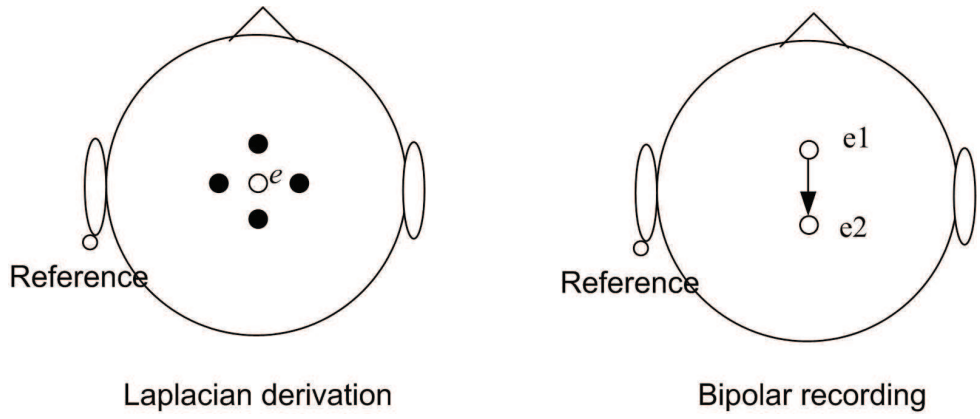


Fig. 7.2: Comparison between Laplacian derivation and bipolar recording. The arrows between the EEG electrodes show a bipolar derivation.

7.2 Multi-class F score based time-frequency selection (MFTFS)

The EEG signals X_e^{LAP} at C3, Cz and C4 are decomposed into EEG segments, in a series of overlapping time-frequency areas $(\omega_m \times \tau_n)$, $m \in \{1, , ..., M\}$, $n \in \{1, , ..., N\}$ with different frequency bands $\omega_m = [f_m, f_m + F - 1], f_{m+1} = f_m + F_s$ (F is the bandwidth, F_s is the frequency step) and time intervals $\tau_n = [t_n, t_n + T - 1], t_{m+1} = t_m + T_s$ (T is the interval width, T_s is the time step).

In two-class problems, the F score is computed to estimate the separation between two classes, so as to find the optimal time-frequency area that yields the largest F score. In the optimal time-frequency area, the populations of features from two different classes have the largest F score distance. This concept can be extended to multi-class problems by the one-versus-rest (OVR) strategy, which is often used in multi-class classification [195, 224]. Let us denote by $\mathscr{C}$ the set of all classes, and then $\mathscr{C} \setminus \{O\}$ for all classes except class O. If we can consider $\mathscr{C} \setminus \{O\}$ as a big class (which we discuss in the following in the case of real BCI problems), the multi-class problem can be transferred to a two-class problem, so as to compute the OVR-based F score:

$$\hat{F} = \frac{\left\|\vec{\mu}^{O} - \vec{\mu}^{\mathscr{C}\setminus\{O\}}\right\|_2^2}{tr(\Sigma^{O}) + tr(\Sigma^{\mathscr{C}\setminus\{O\}})} \tag{7.3}$$

where Σ denotes the covariance matrix of the feature vector, $\vec{\mu}$ denotes the mean of the feature vector, $\|\cdot\|_2$ denotes the $L2$-norm (Euclidean norm), and $tr(\cdot)$ the trace of a matrix.

In this chapter, the feature vector, $[BP_{C3}, BP_{Cz}, BP_{C4}]$, contains the BP features (BP, see Equation 2.13 in Chapter 2) from three Laplacian channels C3, Cz and C4, so the OVR-based F score is calculated by:

$$\hat{F} = \frac{(\overline{BP}^{O}_{C3} - \overline{BP}^{\mathscr{C}\setminus\{O\}}_{C3})^2 + (\overline{BP}^{O}_{Cz} - \overline{BP}^{\mathscr{C}\setminus\{O\}}_{Cz})^2 + (\overline{BP}^{O}_{C4} - \overline{BP}^{\mathscr{C}\setminus\{O\}}_{C4})^2}{(\bar{S}^{O}_{C3} + \bar{S}^{O}_{Cz} + \bar{S}^{O}_{C4}) + (\bar{S}^{\mathscr{C}\setminus\{O\}}_{C3} + \bar{S}^{\mathscr{C}\setminus\{O\}}_{Cz} + \bar{S}^{\mathscr{C}\setminus\{O\}}_{C4})} \tag{7.4}$$

with:

$$\overline{BP} = \frac{1}{K}\sum_{k=1}^{K} BP(k) \tag{7.5}$$

$$\bar{S} = \frac{1}{K-1}\sum_{k=1}^{K}(BP(k) - \overline{BP})^2 \tag{7.6}$$

where K is the number of trials.

As we have mentioned in the previous chapter, the logarithm has been applied on BP features (see Equation 2.13) to make the distribution of BP features approximately normal. In probability theory, a random vector is said to be multivariate normally distributed if every linear combination of its components obeys a univariate (one-dimensional) normal distribution [85]. Thus, the feature vector, $[BP_{C3}, BP_{Cz}, BP_{C4}]$, should be multivariate normally distributed. In practice, Mardia's test is used to check whether a given set of data obeys the multivariate normal distribution with a given significance level of 0.05 [115]. The assumption of multivariate normality is valid when the significance level P is larger than 0.05.

A real example from BCI competition IV dataset IIa [39] is shown in Figure 7.3, which is used to discuss whether it is reasonable to use OVR-based F score in BCI problems. The feature vector, $[BP_{C3}, BP_{Cz}, BP_{C4}]$, was extracted from the data in the time-frequency area with the largest F score for one class (left hand) against all the others (right hand, feet, tongue). Mardia's test shows that the feature vector is multivariate normally distributed for each class ($P = 1.000$). Figure 7.3 (B) shows the distribution of feature vectors from all classes except left hand, which is also multivariate normally distributed ($P = 1.000$). In this case, the set of all the other classes can be consider as a big class, which is normally distributed. Thus, $\vec{\mu}^{\mathscr{C}\setminus\{O\}}$ well represents the center of $\mathscr{C} \setminus \{O\}$. As a result, $\left\|\vec{\mu}^{O} - \vec{\mu}^{\mathscr{C}\setminus\{O\}}\right\|_2$ is the Euclidean distance between the center of one class (blue color) and the center of big class (cyan color) (see Figure 7.3 C), which reflects the inter-class difference between O and $\mathscr{C} \setminus \{O\}$. As we have discussed in Chapter 5 Section **??**, the trace of the covariance matrix of the feature vector reflects intra-class spread for a normal distribution. Thus, using $tr(\Sigma^{\mathscr{C}\setminus\{O\}})$ to estimate the variance of $\mathscr{C} \setminus \{O\}$ becomes reasonable.

For each class, we calculate $\hat{F}(\omega_m, \tau_n)$ using Equation 7.4, for each time-frequency area $(\omega_m \times \tau_n)$, so as to measure its discriminative power for separating the class from all the others. The optimal time-frequency area (ω^*, τ^*) for separating each class is estimated by seeking the maximum value of $\hat{F}(\omega_m, \tau_n)$ among all $M \times N$ time-frequency areas:

$$\hat{F}(\omega^*, \tau^*) = \max\left\{\hat{F}(\omega_m, \tau_n) \mid m \in \{1, 2, ..., M\}, n \in \{1, 2, ..., N\}\right\} \tag{7.7}$$

Without loss of generality, Figure 7.4 presents the scheme of multi-class F score based time-frequency selection (MFTFS) for a four class problem. The optimal time-frequency area for separating one class from all the others can be considered as the characteristic time-frequency area for the class, since it contains information that makes the class different from all the others. Let us assume that the characteristic time-frequency areas for I different classes are $(\omega^*, \tau^*)_i$, $(i = 1, ..., I)$. BP features, BP_{C3}, BP_{Cz}, BP_{C4}, are extracted from the characteristic time-frequency area $(\omega^*, \tau^*)_i$ of each class i in Laplacian channels C3, Cz and C4, so called the class-relevant feature vector. Considering the selected characteristic time-frequency areas for some classes may be the same in practice, we only use the set of feature vector from all different (ω^*, τ^*) for classification. This step can be achieved by:

$$\mathscr{A} = UNIQUE(\{(\omega^*, \tau^*)_i \mid i \in 1, ..., I\}) \tag{7.8}$$

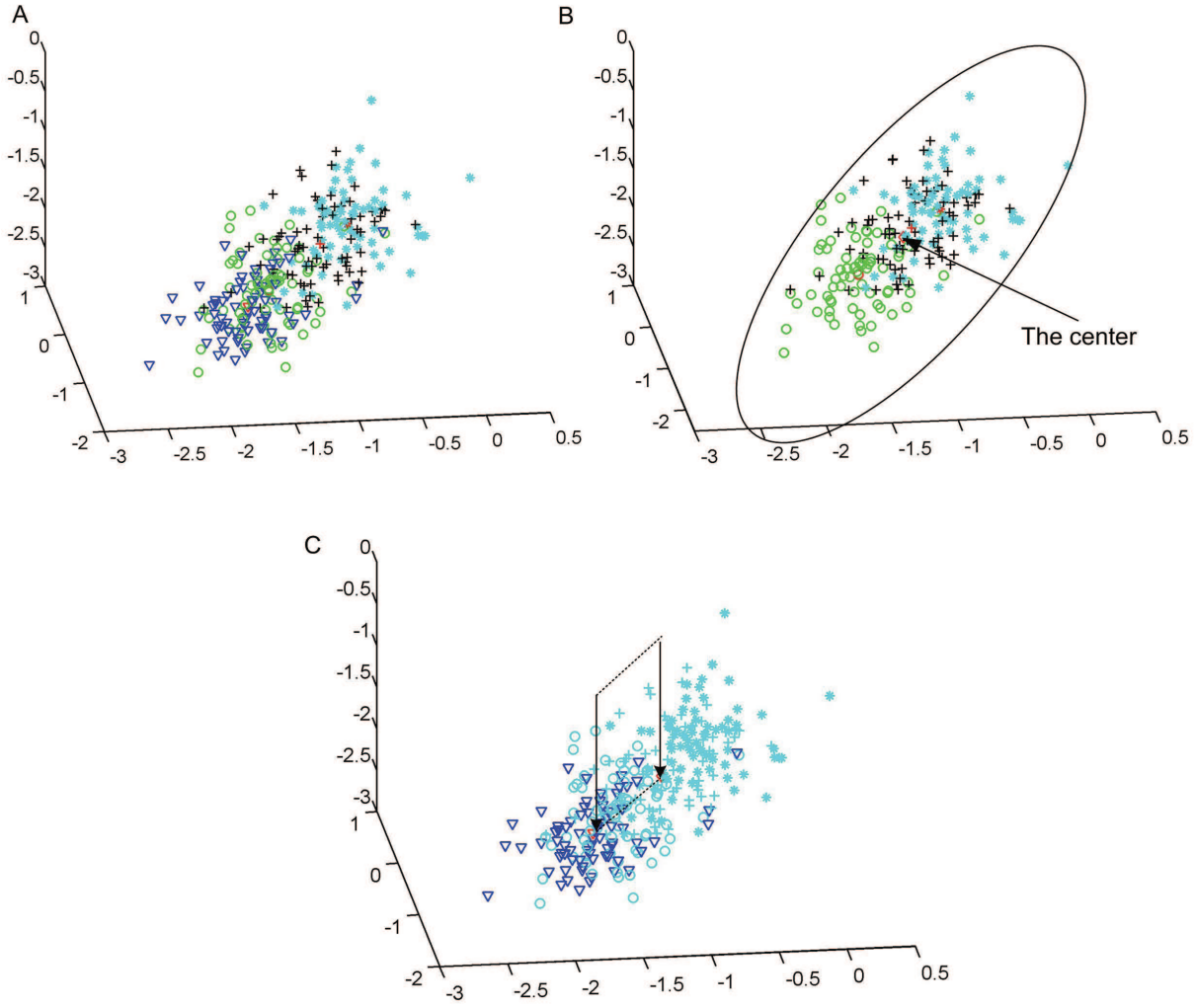


Fig. 7.3: Distribution of BCI features: a real example from BCI competition IV database IIa [39]. A. Distribution of the feature vector for different classes (Mardia's test, $P = 1.000$ for each class): left hand (∇), right hand ($\circ$), feet ($+$) and tongue ($*$). B. Set of feature vectors from all classes except left hand (Mardia's test [115], $P = 1.000$): the arrow shows the center of the set. C. Distance between the center of one class (blue color) and the center of the set of all the other classes (cyan color).

where $\mathscr{A}$ is the set of all different (ω^*, τ^*), the operator $UNIQUE$ eliminates possible time-frequency area repetitions in the set of $\{(\omega^*, \tau^*)_i \mid i \in 1, ..., I\}$. Then, the feature vector $[BP_{C3}, BP_{Cz}, BP_{C4}]$ is extracted from the time-frequency areas in $\mathscr{A}$ for classification. The classifier we used in this chapter is multi-class LDA from BioSig toolbox [176].

Table 7.1 shows an example of characteristic time-frequency areas for different classes for a subject from BCI competition IV dataset IIa [39]. In this example, the classes of right hand and feet share the same characteristic time-frequency area (21-29Hz, 0.5-3.5s after the cue on-set). In this case, the repetitions should be eliminated. Thus, the features will be only extracted from three different characteristic time-frequency areas: 1) 8-16Hz, 0.9-3.9s after the cue on-set, 2) 21-29Hz, 0.5-3.5s after the cue on-set and 3) 21-29Hz, 0.9-3.9s after the cue on-set. Thus, totally $3 \times 3 = 9$ features are used for classification.

Table 7.1: An example of characteristic time-frequency areas for different classes for a subject from BCI competition IV dataset IIa [39]. The time-frequency areas shared by different classes are shaded in gray.

	Left hand	Right hand	Feet	Tongue
Freq.(Hz)	8-16	21-29	21-29	21-29
Time(s)	0.9-3.9	0.5-3.5	0.5-3.5	0.9-3.9

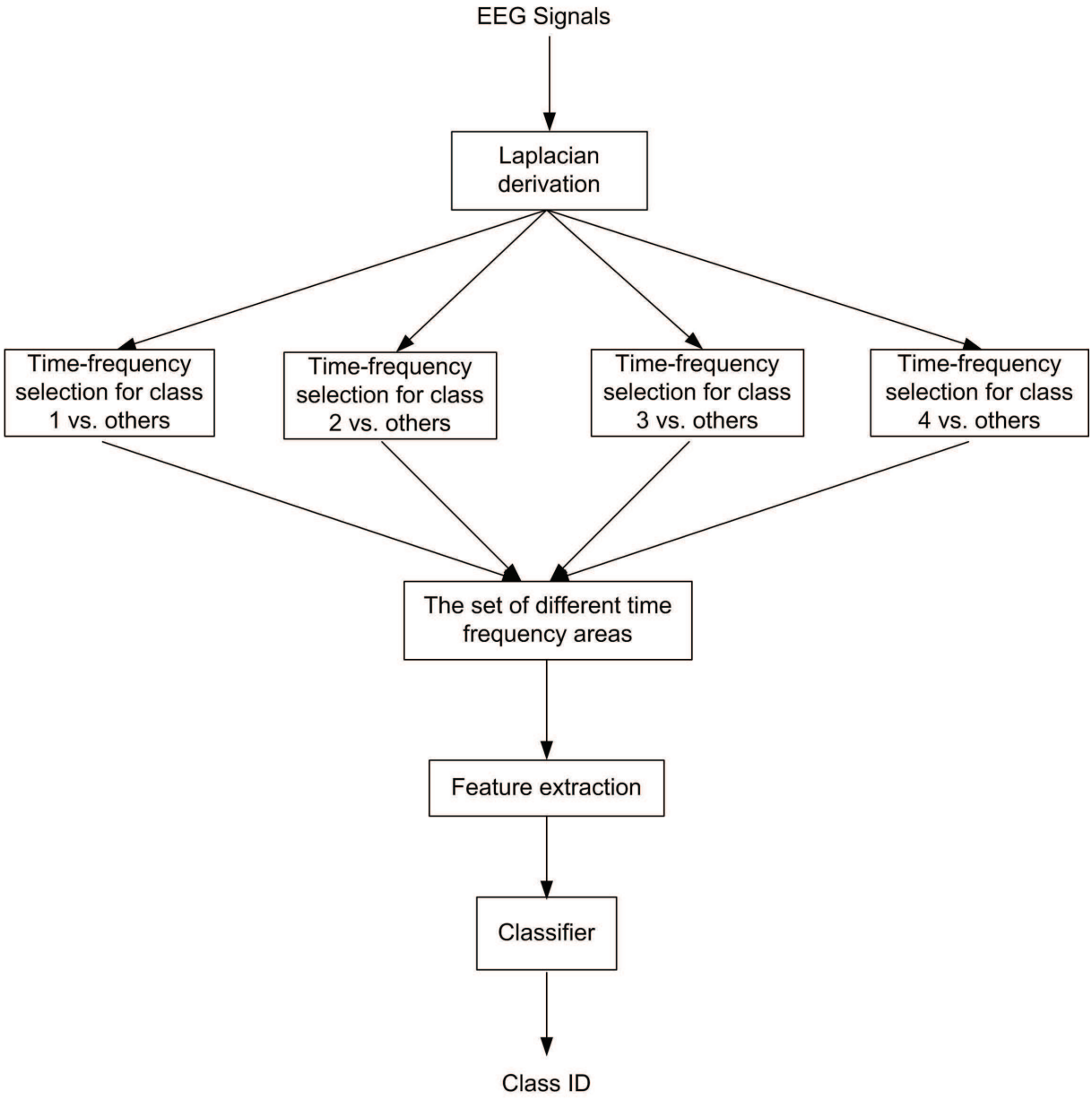


Fig. 7.4: Scheme of Multi-class F score based time-frequency selection (MFTFS).

7.3 Data description and preprocessing

In this chapter, two multi-class datasets from the BCI competition were used. Both datasets contain four classes of motor imagery data: left hand, right hand, foot and tongue.

7.3.1 BCI competition IV dataset IIa

BCI competition IV dataset IIa [39] was recorded from nine subjects using 22 EEG electrodes. EOG data were recorded using three EOG electrodes for removing the EOG artifacts. The placement of EEG electrodes is shown in Figure 7.5. This dataset contains one training session (288 trials) and one independent evaluation session (288 trials, 72 trials for each class). In each trial, the subjects were required to perform motor imagery during four seconds after the cue on-set.

As the training and testing sessions are recorded in two different days, this dataset presents the challenge of "session-to-session transfer". Thus, this dataset can be used to evaluate the robustness of a method to unpredictable changes of signal. For details about this dataset, please refer to Chapter 4.

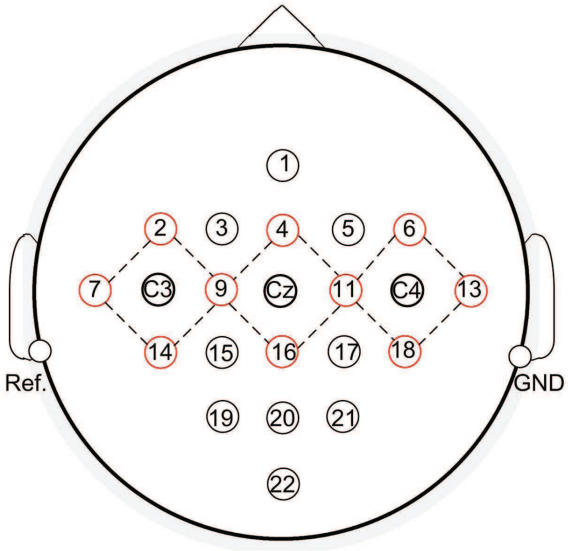


Fig. 7.5: Placement of 22 electrodes for the BCI competition IV dataset IIa [39]. The Laplacian derivation is performed at C3, Cz and C4 based on the center and four nearest neighboring electrodes (marked in red).

7.3.2 BCI competition III dataset IIIa

BCI competition III dataset IIIa [179] was recorded from three subjects (denoted "k3", "k6", "l1") using 60 EEG electrodes. The placement of EEG electrodes is shown in Figure 7.7. No EOG electrodes were placed for removing the EOG artifacts. Trials with artifacts, which were visually identified, are marked out by the property HDR.ArtifactSelection ($HDR.ArtifactSelection = 1$ indicates that the trial was contaminated by artifacts, $HDR.ArtifactSelection = 0$ indicates that the trial was relatively "clean" without visible artifacts). Both training and testing data are recorded in the same session with the same number of trials, and randomly mixed by data providers. The number of training/testing data for each subject is listed in Table 7.2, and also the percentage of artifact contaminated (AC) trials in the training data.

The timing of the experimental paradigm is shown in Figure 7.6. In each trial, the subjects were required to sat quietly for the first two seconds. Then, an acoustic stimulus was given, with a cross presented in the center of the screen, to indicate the beginning of the trial. A visual cue in form of an arrow pointing either to left, right, up or down (corresponding to left hand, right hand, both feet and tongue) was given at 3s and stayed on the screen for one second. The subjects were asked to imagine the corresponding movement during four seconds after the cue on-set. The cues indicating different motor imagery tasks were displayed in a randomized order.

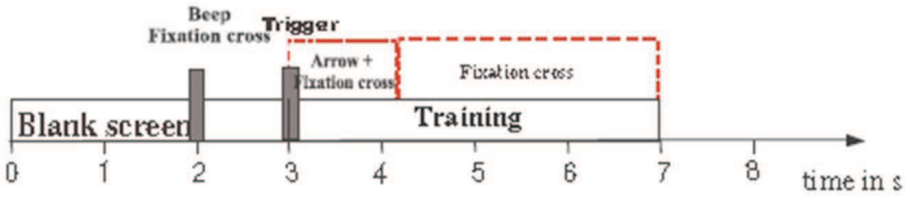


Fig. 7.6: Timing of the experimental paradigm for dataset IIIa [179].

As the AC trials are marked out in this dataset, we tested our method using only "clean" trials, and using all training trials (containing both "clean" and AC trials), to evaluate the robustness of our method to artifacts.

Table 7.2: Number of training (clean trials and artifact contaminated (AC) trials) and testing trials, and percentage of artifact contaminated (AC) trials in the training data from BCI competition III dataset IIIa.

Subject	Training trials				Testing trials
	Total	Clean trials	AC trials	Percentage of AC trials	
k3	180	149	41	17%	180
k6	120	92	28	23%	120
l1	120	84	36	30%	120

7.3.3 Preprocessing

For both datasets, the Laplacian derivation (Chapter 2, Section 2.6.3) was performed at C3, Cz and C4 based on the center and four nearest neighboring electrodes (see Figures 7.5 and 7.7) [75].

The parameters for time-frequency selection are the same for these two datasets, since the classes (i.e. left hand, right hand, both feet and tongue) and timing (i.e. four seconds after cue on-set) of motor imagery in these two datasets are the same. For each Laplacian channel, 5th order Butterworth filters are applied to compute 19 successive 4Hz-wide frequency bands of signals ($F = 4$Hz, $F_s = 1$Hz): 8-12Hz, 9-13Hz, 10-14Hz, ..., 26-30Hz, and 15 successive 8Hz-wide frequency bands of signals ($F = 8$Hz, $F_s = 1$Hz): 8-16Hz, 9-17Hz, 10-18Hz, ..., 22-30Hz. Then 36 overlapping time segments in each frequency band were obtained through 2s, 2.5s and 3s-wide (i.e. $T = 2$, 2.5 and 3s, respectively) sliding windows (12 segments for each sliding window) with 0.2s-step (i.e. $T_s = 0.2$) moving from

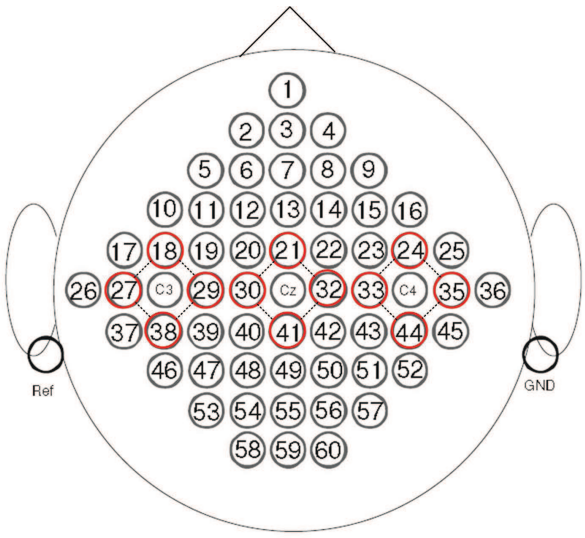


Fig. 7.7: Placement of 60 electrodes on BCI competition III dataset IIIa [179]. Laplacian derivation is performed at C3, Cz and C4 based on the center and four nearest neighboring electrodes (marked in red).

0.5s after the cue on-set. Therefore, there are $(19+15)\times 36 = 1224$ time-frequency areas for estimating the characteristic time-frequency area for each class. These parameters are the same as in the previous chapter. They have yielded good performances in our previous work [236]. Although multi-class problems are not the same as two-class problems, this setting of parameters generally covers different frequency bands (4Hz and 8Hz-wide) in the range of 8-30Hz where ERD/ERS typically occurs, and different time segments (2s, 2.5s and 3s-wide) in the whole period of motor imagery.

7.4 Results

7.4.1 BCI competition IV dataset IIa

The OVR-based F score selected characteristic time-frequency areas for different classes are listed in Table 7.3 for each subject. For each class of motor imagery, its characteristic time-frequency area varies with subjects, indicating the necessity of subject-specific estimation in multi-class time-frequency selection. Table 7.4 lists the percentage of times that a time-frequency area of a given size in time and frequency is selected. The characteristic frequency bands are mainly in the range of β band (12-30Hz) with a relatively broad frequency width (8Hz width), in particular for motor imagery of feet and tongue. The characteristic time segment has typically a long length (3s length). Thus, discriminating one class from all the others usually needs a relatively large time-frequency area for extracting useful information.

In this competition, five winners have reported their results on the website. Key information on the methods used by these winners is given in Table 7.5 (for more detailed information, see [194]), to compare with our method MFTFS. Only the results of the 3^{rd} and 5^{th} place winners are obtained by using reduced channels, others are yielded by using all 22 channels [194].

Table 7.3: Characteristic time-frequency areas for different classes on BCI competition IV dataset IIa. The time-frequency areas shared by different classes are shaded in gray.

Subject		Left hand	Right hand	Feet	Tongue
1	Freq.(Hz)	12-20	12-20	12-20	12-20
	Time(s)	0.7-3.7	0.9-3.9	1.1-4.1	0.9-3.9
2	Freq.(Hz)	15-23	11-19	21-29	22-30
	Time(s)	1.5-4.5	0.9-3.9	2.3-5.3	2.7-5.7
3	Freq.(Hz)	20-28	21-29	18-26	19-27
	Time(s)	0.7-3.7	0.7-3.7	1.1-4.1	0.7-3.7
4	Freq.(Hz)	22-30	14-18	22-30	14-22
	Time(s)	0.5-3.5	1.3-4.3	1.7-4.7	0.5-3.5
5	Freq.(Hz)	10-18	22-30	26-30	10-18
	Time(s)	0.5-3.5	0.5-3.0	2.7-5.7	0.9-3.4
6	Freq.(Hz)	13-21	13-21	14-22	13-21
	Time(s)	1.1-4.1	0.5-3.0	0.5-3.5	0.9-3.9
7	Freq.(Hz)	8-16	21-29	21-29	21-29
	Time(s)	0.9-3.9	0.5-3.5	0.5-3.5	0.9-3.9
8	Freq.(Hz)	8-16	11-19	12-20	9-17
	Time(s)	0.7-3.7	0.5-3.0	1.3-4.3	0.5-3.5
9	Freq.(Hz)	14-22	23-27	22-30	12-20
	Time(s)	0.5-2.5	0.5-2.5	0.9-3.4	0.5-3.5

Table 7.4: Percentage of times that a time-frequency area of a given size in time and frequency is selected (BCI competition IV dataset IIa).

	2s	2.5s	3s	Total
4Hz	2.8%	0	2.8%	5.6%
8Hz	2.8%	11.1%	80.6%	94.4%
Total	5.6%	11.1%	83.4%	100%

Table 7.6 compares the classification performances in session-to-session transfers of the MFTFS and of other methods in the BCI competition. The classification performance is evaluated by *kappa coefficient*, κ (see [178] or Chapter 2 Section 2.5.2 for details). The 1^{st} place winner yields the best performance among all players and our method for all subjects except Subjects 1, 2 and 9, because it used the selected frequency band and all 22 channels. For Subjects 1 and 9, the best performance is achieved by the 2^{nd} place winner using a board frequency band (8-30Hz) with CSP, which is a little better than the 1^{st} place winner. The reason is that at least one motor imagery task has elicited a broad band ERD/ERS for those subjects. For example, the motor imagery of right hand elicited a right brain (C3) dominant broad band ERD for Subject 1 (see Figure 7.8), which makes the best performance come from using a broad frequency band (8-30Hz) for this subject. For Subject 2, although the 1^{st} place winner's result is not better than our method, it generates the best result among all winners. Generally speaking, the 1^{st} place winner generates very good results (the best performance for 6 subjects, the second place performance for 3 subjects). For Subject 5, the result yielded by the 1^{st} place winner ($\kappa = 0.40$) is far better than all others. These results indicate that using selected frequency band and all channels are useful for generating a good performance. Thus, if

Table 7.5: Comparison between our method (MFTFS) and the methods used by BCI competition winners (BCI competition IV dataset IIa).

Method	Frequency band	Channels	Features	Classifier
MFTFS	Selected by F score	3 Laplacian Channels (C3, Cz and C4)	BP	Multi-class LDA
1^{st} [45, 194]	Selected by Filter Bank CSP (FBCSP)	All channels (22 channels)	CSP	Naive Bayes Parzen Window classifier
2^{nd} [194]	8-30Hz	All channels (22 channels)	CSP	Bayesian classifier
3^{rd} [194]	8-25Hz	Channels selected by the recursive channel elimination [97]	CSP	SVMs
4^{th} [194]	Selected based on classification accuracy	All channels (22 channels)	NTSPP+CSP	LDAs + SVMs
5^{th} [194]	8-25Hz	Channels manually selected around C3 and C4	CSP	SVMs

we have to reduce some channels, selecting optimal time-frequency areas may become an important solution for improving the classification results.

Table 7.6: Comparison of the classification performances (evaluated with *kappa coefficient*) in session-to-session transfers of the MFTFS and of other methods based on reduced channels (3^{nd} and 5^{th} place winners) and all channels (1^{st}, 2^{nd} and 4^{th} place winners), respectively, on BCI competition IV dataset IIa. The best performances among methods based on reduced channels are shaded in gray. The best performances among all methods are in bold.

	Subjects									
	1	2	3	4	5	6	7	8	9	Mean
MFTFS	0.47	**0.46**	0.31	0.22	0.13	0.25	0.52	0.38	0.48	0.36
3^{nd}	0.38	0.18	0.48	0.33	0.07	0.14	0.29	0.49	0.44	0.31
5^{th}	0.41	0.17	0.39	0.25	0.06	0.16	0.34	0.45	0.37	0.29
1^{st}	0.68	0.42	**0.75**	**0.48**	**0.40**	**0.27**	**0.77**	**0.75**	0.61	**0.57**
2^{nd}	**0.69**	0.34	0.71	0.44	0.16	0.21	0.66	0.73	**0.69**	0.52
4^{th}	0.46	0.25	0.65	0.31	0.12	0.07	0.00	0.46	0.42	0.30

As our method is based on reduced channels (only 3 Laplacian channels), it is fair to further compare our results with those also obtained by reduced channels in the competition (3^{rd} and 5^{th} place winners), in order to valid the contribution of time-frequency optimization. Among all methods based on reduced channels, MFTFS generated the best mean classification performance (mean *kappa coefficient*, $\bar{\kappa} = 0.36$). Although the difference between our results and these two winners are not statistically significant due to the lim-

ited number of subjects, MFTFS outperforms both winners (3^{rd} place winner: $\bar{\kappa} = 0.31$, 5^{rd} place winner: $\bar{\kappa} = 0.29$) and yields the best individual results for 6 out of 9 subjects (except Subjects 3, 4 and 8). This result indicates that using MFTFS can improve the classification results of BCI based on reduced channels in session-to-session transfers for most subjects. Compared to the methods using all channels, MFTFS generates better performances than the 4^{th} place winner, who used a selected frequency band and all 22 channels for most subjects (except Subjects 3, 4 and 8). For Subject 2, MFTFS even generates better performance than the 1st place winner. These results indicate that it is possible to reduce channels but improve the classification performance by using MFTFS in some cases. Our method generates relatively worse performances than CSP-based methods on Subject 3, 4 and 8 because of the poor quality of signals recorded from these three subjects. As a result, they need either more channels (1^{st}, 2^{nd} and 4^{th} place winners) or a broad band (3^{rd} and 5^{th} place winners) to provide more information for signal enhancement. Time-frequency visualization of ERD/ERS for Subject 3 is shown in Figure 7.9, which is generated by the BioSig Toolbox [176] (for details, see [152] or Chapter 6). For this subject, both hand movements elicit discriminative ERD: left hand movement generates right brain (C4) dominant ERD, while right hand movement generates left brain (C3) dominant ERD. As a result, F score based time-frequency selection yields very good performance in the discrimination between left hand and right hand movements ($\kappa = 0.96$), which is presented in the previous chapter. However, the tongue movement does not elicit the typical discriminative ERS in hand representative areas, so that the multi-class performance deteriorates ($\kappa = 0.31$). All winners in this competition have used CSP algorithms to enhance the signals. As we has discussed in Chapter 2 Section 2.6.5, CSP algorithm is powerful to improve the signal quality. Thus, CSP-based methods generate better results than us for these three subjects. However, CSP is typically based on multi-channel setting, so the results of using CSP with all channels are much better than those of using CSP with reduced channels (see Table 7.6).

7.4.2 BCI competition III dataset IIIa

Table 7.7 lists the classification results (also evaluated by *kappa coefficient*, κ) when using all training trials (All) and using only the "clean" trials. As there are only three subjects in this dataset, we examine the results one subject by one subject. For Subject $k3$, using

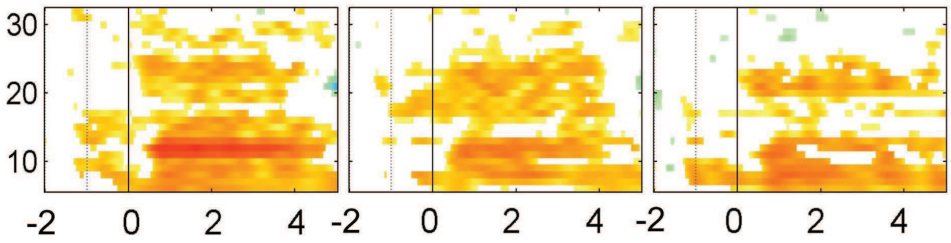


Fig. 7.8: Time-frequency visualization of right hand motor imagery elicited ERD/ERS for Subject 1, using the time interval between -2s (i.e. 2s before cue on-set) and -1s (1s before cue on-set) as the baseline. This visualization was generated by the BioSig Toolbox [176], using overlapping 2Hz bands (step = 1Hz) in the frequency range between 6 and 32Hz (for details, see [152] or Chapter 6).

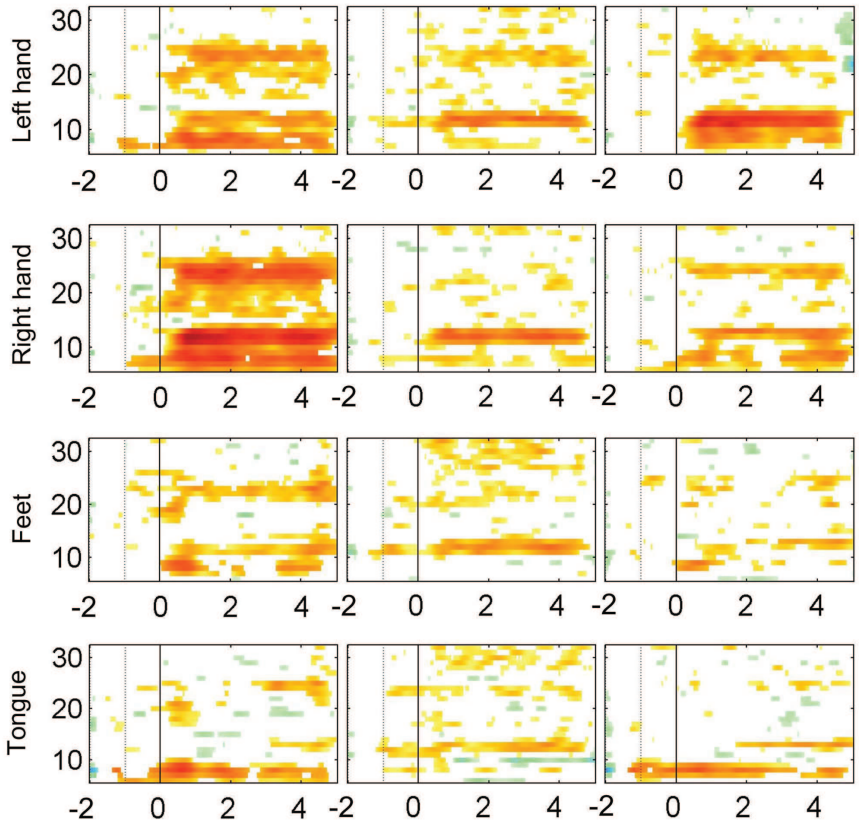


Fig. 7.9: Time-frequency visualization of ERD/ERS for Subject 3, using the time interval between -2s (i.e. 2s before cue on-set) and -1s (1s before cue on-set) as the baseline. This visualization was generated by the BioSig Toolbox [176], using overlapping 2Hz bands (step = 1Hz) in the frequency range between 6 and 32Hz (for details, see [152] or Chapter 6).

only the "clean" trials ($\kappa = 0.69$) yields a better classification result than using all training trials ($\kappa = 0.63$). However, for Subjects $k6$ and $l1$, the classification results obtained by using only the "clean" trials (Subject $k6$: $\kappa = 0.70$, Subject $l1$: $\kappa = 0.60$) are worse than those obtained by using all training trials (Subject $k6$: $\kappa = 0.71$, Subject $l1$: $\kappa = 0.72$). Generally speaking, artifacts will deteriorate the classification results, while the number of trials for training will also affect the results. Usually, a small training dataset may not be able to generate a good classification result. However, removing artifact-contaminated (AC) trials will definitely reduce the number of trials for training. As a result, a trade-off should be made, which depends on the sensitivity of the method to noise and the amount of training data. For a method which is sensitive to noise, adding AC trials to the training dataset can only deteriorate the classification result. As a result, using only the "clean" trials will always yield better results than using all training trials. On the contrary, for a method which is robust to noise, adding AC trials may improve the classification results. In Table 7.7, we can see that using all training trials ($\bar{\kappa} = 0.69$) generates a better mean classification result than using only "clean" trials ($\bar{\kappa} = 0.63$), indicating that MFTFS is robust to noise. In this dataset, subject $l1$ has the least clean training trials (84 trials, see Table 7.2). If only using clean trials, the classification result for $l1$ ($\kappa = 0.60$) is the worst among all the subjects. To study the effect of adding AC

trials on the performance of MFTFS, we add different amounts of AC data to training data on this subject, so as to get a curve of classification performance with respect to different numbers of training data with different amounts of noise (see Figure 7.10). AC trials are randomly selected to constitute the different amounts of noise. The curve provides the average κ values over 20 runs (standard deviation ≤ 0.06). From Figure 7.10, we can see that when the number of training trials is less than 114 (AC trials/training trials = 30/84), adding AC trials can improve the classification performance. When the number of training trials is larger than 114, adding AC trials will deteriorate the classification performance. The best performance ($\kappa = 0.77$) is achieved with 114 training trials (AC trials/training trials = 30/84) for this subject. We also give a curve for Subject $k3$ (Figure 7.11), since the classification result deteriorates when using all training trials for this subject (see Table 7.7). In Figure 7.11, we can see that adding AC trials cannot improve the classification results. The best performance is achieved when using only clean trials. The reason is that this subject has a large number of training trials (141 trials, see Table 7.2). Usually, a training dataset should have five to ten times as many trials as the dimensionality of features to guarantee a good performance of a classifier [83, 164]. For a four-class problem, the dimensionality of MFTFS selected BP features are more than $3 \times 4 = 12$, so that 141 trials are enough to feed a classifier. In this case, adding AC trials may deteriorate the classification performance with MFTFS, since it mainly introduces artifacts. To sum up, using MFTFS, when the number of clean training trials is not sufficient, it is possible to improve classification results by adding AC trials. On the contrary, if the number of clean training trials is large enough, adding AC trials is not helpful.

Table 7.7: Evaluation of the MFTFS sensitivity (evaluated with *kappa coefficient*) to noise on BCI competition III dataset IIIa.

	Subjects			
	k3	k6	l1	Mean
All	0.63	0.71	0.72	0.69
Clean	0.69	0.70	0.60	0.66

Table 7.8 lists the characteristic time-frequency areas for different classes, obtained by using all training trials (All) and only the "clean" trials, respectively. From Table 7.8, we can see that the characteristic time-frequency areas shift for some classes for each subject between using all training trials and only the "clean" trials. Thus, MFTFS can adapt to the amount of training data and noise to find the optimal time-frequency areas for extracting discriminative features.

In the competition, only the classification results of the first three winners on this dataset are reported. Key information on the methods used by the three winners are given in Table 7.9, in order to compare with MFTFS. All these methods have employed all 60 channels provided in the dataset. The 1^{st} place winner performed time-frequency selection by Fisher ratios [114]; however, the detailed information on this time-frequency selection is not available in the literature. The 2^{nd} place winner applied a multi-class CSP algorithm (see Chapter 4 and [60]) on the broad band (8-30Hz) without time-frequency selection. The 3^{rd} place winner performed "infomax" independent component analysis (ICA) [112] for seeking the interesting independent component and then calculated amplitude spectra using Welch method [222] with 5 overlapping Hanning windows. The

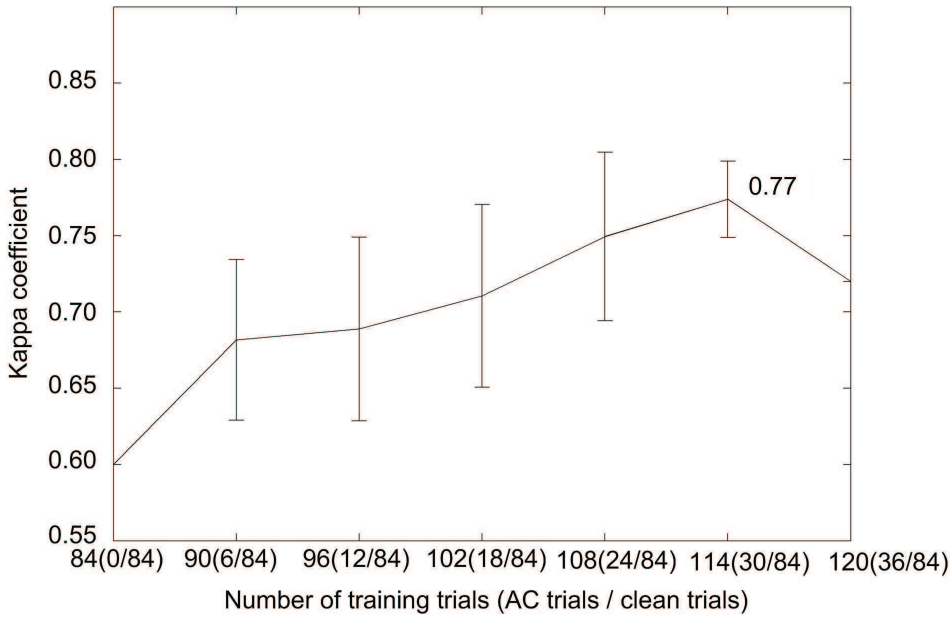


Fig. 7.10: Curve of classification performance (measured by κ values) with respect to different numbers of training data with different amounts of noise for Subject $l1$. Results are averaged over 20 runs. Error bars show standard deviations of κ values.

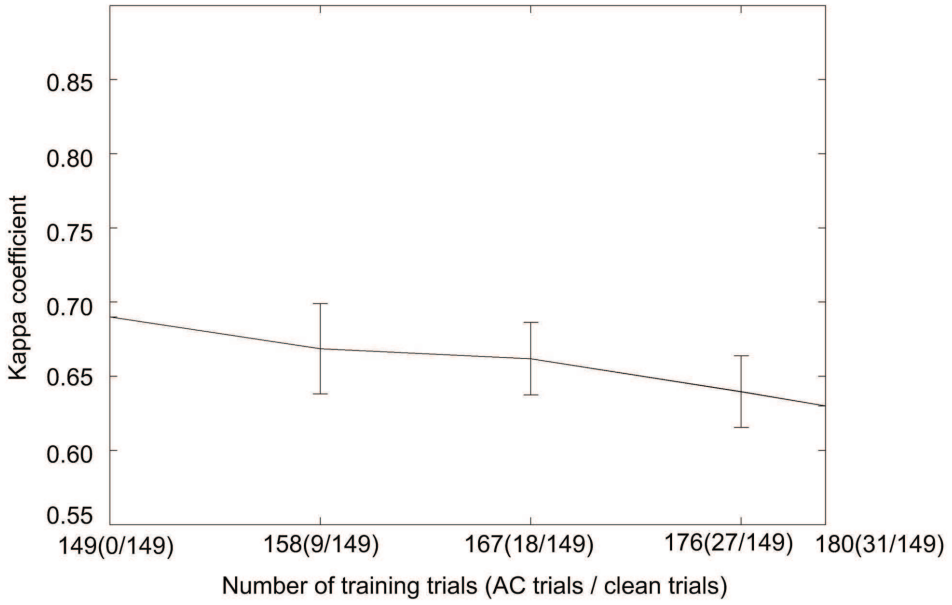


Fig. 7.11: Curve of classification performance (measured by κ values) with respect to different number of training data with different amounts of noise for Subject $k3$. Results are averaged over 20 runs. Error bars show standard deviations of κ values.

features were obtained by linear principal component analysis (PCA) [87]. For more detailed information about these methods, please refer to [35]. As the AC trials are marked out in this dataset, researchers can choose to use all training trials or only clean training trials, though all winners did not report whether they have used all training trials or only

Table 7.8: Characteristic time-frequency areas for different classes on BCI competition III dataset IIIa, obtained by using all training trials (All) and only the "clean" trials, respectively.

Subject			Left hand	Right hand	Feet	Tongue
k3	All	Freq.(Hz)	12-20	13-21	13-21	12-20
		Time(s)	0.9-3.9	1.9-4.9	2.1-5.1	0.7-3.7
	Clean	Freq.(Hz)	12-20	12-20	13-21	12-20
		Time(s)	2.5-5.0	2.1-5.1	2.1-5.1	0.7-3.7
k6	All	Freq.(Hz)	22-30	26-30	22-30	19-27
		Time(s)	0.5-3.5	0.9-3.9	2.7-5.7	2.7-5.7
	Clean	Freq.(Hz)	22-30	26-30	22-30	18-26
		Time(s)	0.5-3.5	0.9-3.9	2.3-5.3	2.7-5.7
l1	All	Freq.(Hz)	20-28	12-20	22-30	12-20
		Time(s)	2.7-5.7	1.1-4.1	2.5-5.5	1.1-4.1
	Clean	Freq.(Hz)	22-30	13-17	22-30	12-20
		Time(s)	0.7-2.7	2.3-4.8	2.7-5.7	2.7-5.7

Table 7.9: Comparisons between different methods on BCI competition III dataset IIIa.

Method	Time-frequency areas	Channels	Features	Classifier
MFTFS	Selected by F score	3 Laplacian Channels (C3, Cz and C4)	BP	Multi-class LDA
1^{st} [35]	Selected by Fisher ratios [114]	All channel (60 channels)	CSP	Multiple SVMs
2^{nd} [35]	8-30Hz	All channel (60 channels)	CSP	SVM+kNN+LDA
3^{rd} [35]	Selected by infomax ICA [112] and PCA	All channel (60 channels)	Amplitude spectra [222]	Multiple SVMs

clean trials in the competition. In this case, we use only clean trials for Subject $k3$, and all training trials for Subjects $k6$ and $l1$. The reason for this choice was discussed in the previous paragraph. The comparisons of the classification performances between the MFTFS and all other methods are provided in Table 7.10. Although the classification performances obtained by using MFTFS are worse than the 1^{st} place winner for all subjects (see Table 7.10), MFTFS uses far less electrodes (15 electrodes) than all methods (60 electrodes) and yields a better mean classification performance ($\bar{\kappa} = 0.71$) than the 2^{nd} ($\bar{\kappa} = 0.69$) and 3^{rd} ($\bar{\kappa} = 0.63$) place winners. For Subjects $k6$ and $l1$, the results obtained by MFTFS (Subject $k6$: $\bar{\kappa} = 0.71$, Subject $l1$: $\bar{\kappa} = 0.72$) are better than the 2^{nd} (Subject $k6$: $\bar{\kappa} = 0.43$, Subject $l1$: $\bar{\kappa} = 0.71$) and 3^{rd} place (Subject $k6$: $\bar{\kappa} = 0.41$, Subject $l1$: $\bar{\kappa} = 0.52$) winners. Although for Subject $k3$, our result is not as good as other methods using all 60 channels, this result ($\kappa = 0.69$) is close to the results of our method on the

other subjects, indicating that the performance of our method is generally stable. As we have discussed before, using a full EEG cap is helpful for improving the classification on some subjects if a full cap is acceptable in practice. Here, the number of electrodes used in MFTFS is close to the one in the commercial BCI system Emotiv EPOC (see Chapter 2 Section 2.3.2). Thus, the number of electrodes we used is reasonable and acceptable for general applications (e.g. in a game environment).

Table 7.10: Comparisons of the classification performances (evaluated with *kappa coefficient*) between the MFTFS and Other Methods on BCI competition III dataset IIIa.

	Subjects			
	k3	k6	l1	Mean
$MFTFS$	0.69	0.71	0.72	0.71
1^{st}	0.82	0.76	0.80	0.79
2^{nd}	0.90	0.43	0.71	0.69
3^{rd}	0.95	0.41	0.52	0.63

7.5 Discussion and future work

This chapter extended the F score based time-frequency optimization to the multi-class case through one-versus-rest strategy. The method was applied with three Laplacian channels C3, Cz and C4 for a four-class BCI problem and tested on two standard multi-class datasets (BCI competition IV dataset IIa and BCI competition III dataset IIIa). The experimental results on BCI competition IV dataset IIa showed that our method yielded better results than other methods based on reduced channels in session-to-session transfers. The experimental results on BCI competition III dataset IIIa showed that adding the artifact-contaminated trials may not deteriorate the performance of our method, indicating that our method is robust to noise. The classification performance of our method is better than the 2^{nd} place winner on this dataset but employed far less electrodes (only 15 electrodes) than all other methods (60 electrodes). Compared to existing commercial BCI products, the number of electrodes we used is still reasonable and acceptable for general applications (e.g. in a game environment).

In conclusion, our method is robust to unpredictable signal changes in session-to-session transfer and artifacts. It can be used to help reducing the number of electrodes in motor imagery BCI, without a large loss of classification performance. However, as this method is based on 3C setup (C3, Cz and C4), its reliability is based on the precise placement of electrodes, which we have discussed in Chapter 5.

To overcome this problem, we suggest a novel design of Laplacian derivation based 3C setup EEG device for general BCI users (see Figure 7.12). The user can choose the right size of the device based on the size of the head. The ear contacts and EOG sensors can easily help the user to put the EEG device in the right position. The electrodes are fixed in the device, so as to prevent a possible large shift of electrode location. This device can be combined with a normal earphone, so as to play the audio cues in BCI experiences through the ear contacts. The EOG sensors are used to detect the ocular activities, so as to remove EOG artifacts or/and use EOG signals as a control signal like what we have

discussed in Chapter 3. A prototype of this design could be made in cooperation with a EEG device company and tested in real applications in the future.

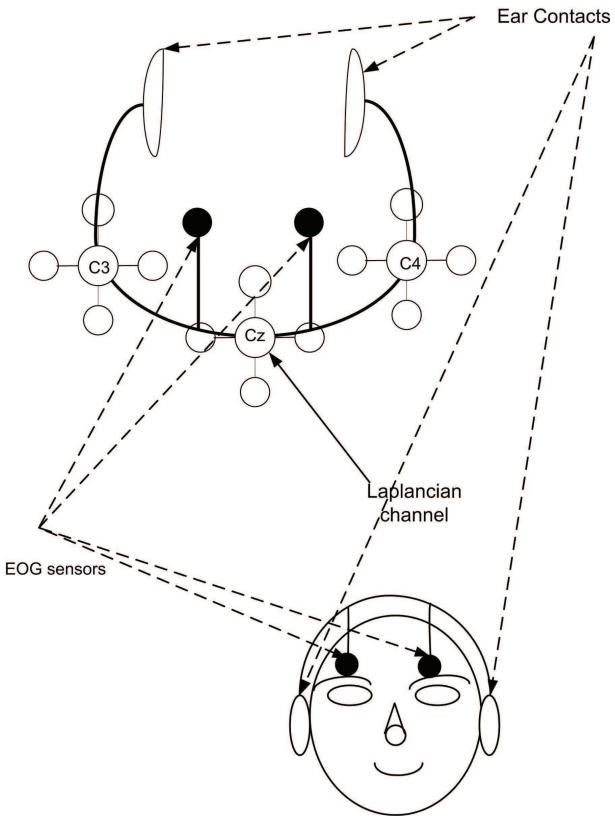


Fig. 7.12: Novel design of Laplacian derivation based 3C setup.

Chapter 8

Conclusion and Perspectives

8.1 Conclusion

Although modern BCI techniques show a relative maturity compared to the past decades, several obstacles still impede their widespread adoption, in particular for large public applications. In this thesis, we proposed a hybrid BCI (hBCI) and several efficient approaches to address some of the challenges in this context.

EEG technique is often used in BCIs for measuring brain activities, because of its inexpensive, non-invasive and good time resolution. However, EEG signals are much weaker compared to other physiological signals, such as heart rates, EOG signals, and often contaminated by these physiological signals. Most researches have focused on how to remove these physiological "artifacts" from EEG signals, in order to obtain the "clean" EEG signal for BCI controls. This research direction is consistent with the traditional definition of BCI that only brain signals can be used in BCI controls. However, this definition, to some extents, ignores the core value of BCI techniques in the context of large public applications. For non-medical application, BCI techniques aim to provide users a hand-free control, so as to increase their freedom in human-computer interaction. But the traditional definition excluding non-brain controls in a BCI system impairs users' freedom of using all possible signals for the control. In our proposed hBCI, users can use brain activity (EEG signals) and eyes movements/blinks (EOG signals) to realize human-computer interactions, where EOG serves as a control signal instead of being considered as artifacts; and users can control the system whenever they want just by opening their eyes. This design provides users more freedom to realize the control through all possible physiological functions and to switch the control system independently. It turns "artifacts" from ocular activities into useful control signals in BCI experiences, which provides an alternative solution to deal with physiological "artifacts". As both EOG and EEG data can be measured by an EEG recording system, this design allows using only one recording system with few electrodes, so as to reduce the hardware cost of a hBCI system. However, we only performed the experiments on five healthy subjects in this thesis. Thus, the robustness of proposed hBCI still needs to be tested on more subjects and further improved before putting this system into real applications.

Most work of this thesis focused on developing machine learning methods in order to make BCI systems efficient, economic and easy to use.

Motor imagery BCI is a popular type of EEG-based BCI, which identifies subject's motor intentions through classifying the spatial relevant EEG patterns (ERD/ERS). In practice, raw scalp EEG signals have a poor signal quality due to the volume conduction effect, which deteriorates the classification performance of motor imagery BCI. To solve this

problem, spatial filtering techniques have been proposed in the literature. The most effective one is the common spatial patterns (CSP) algorithm. However, the performance of CSP usually depends on preprocessing (EEG time segment and frequency band selection), electrode selection and number of paired filters. Although many methods were proposed to select the electrodes and improve the preprocessing procedure for CSP, how to find the optimal number of paired filters in CSP is still an open question. To address this problem, we proposed an automatic method relying on *Rayleigh quotient* to estimate the optimal number of filters for each subject. Based on an existing dataset (BCI competition IV dataset IIa), we studied the effect of this parameter on the classification performances. The estimated subject-specific optimal values yielded better performances on testing data than the recommended value in the literature. In the view of feature extraction, this method also provides a simple solution for feature dimensionality reduction. It reduces the number of CSP features to an optimal number by selecting a few key CSP filters. However, it cannot reduce the number of electrodes in a BCI system.

The CSP algorithm typically needs a multi-channel setting, which may be a drawback for daily-use BCIs in signal acquisition. Based on the same dataset, three subsets of channels selected by prior knowledge of brain functional areas were tested with the CSP algorithm to study the channel selection and reduction challenges, comparing with a SVM-based channel selection method, called recursive channel elimination (RCE). The classification results obtained by two subsets selected based on prior information of functional areas of the brain are generally better than those obtained with a subset of channels selected by RCE, indicating the potential interest of using some prior knowledge in BCI challenges. However, the classification results obtained using reduced channels are not as good as those obtained when using a full cap with the CSP algorithm. We further explored some non-CSP based approaches, aiming to reduce the number of electrodes without a significant drop of classification accuracy. This issue was addressed in two directions in this thesis.

The first direction is to find the optimal subset of electrodes from a full EEG cap. Unlike existing methods simply choosing electrodes based on spatial information, disregarding the potential impact of temporal information, we proposed a novel subject-specific channel selection method to realize the parametrization of both time segment and channel positions. This method is based on a novel criterion, F score, derived from *Fisher's* discriminant analysis. Different from the classical Fisher's criterion, which is typically for measuring the discrimination power of a single feature, the F score is proposed to estimate the discrimination power of a group of features (a feature vector). The experimental results showed that the optimal subset of channels varied with time segments, indicating the necessity of considering time information in channel selection. Based on our proposed method, the optimal combination of time segment and subset of channels can be learned for each subject. This method efficiently reduced the number of channels (from 118 channels to no more than 11), and shortened the training time, without a significant decrease of classification accuracy on a 118-channel dataset (BCI competition III dataset IVa). Compared to existing commercial BCI products (e.g. Emotiv), the number of electrodes we used is reasonable and acceptable for general applications (e.g. in a game environment). However, it needs to be mentioned that the optimal combination of time segment and subset of channels varies with subjects. In this case, a subject-specific optimization is necessary. Although subject-specific optimization can provide a precise estimation in finding the interesting channels for classification, it will unavoidably increase the hardware

cost of a device due to at least two reasons. On the one hand, a full cap is necessary for recording the data to select the interesting channels at the first time use. On the other hand, a BCI device with a subject-specific montage of electrodes has to be produced based on the channel selection result for each subject. Although the user may benefit from this personalized design, mass production becomes impossible for device producers, which finally increases the cost of a BCI device and therefore prevents the wide adoption of BCI. So it is a big limitation of this direction.

The second direction is to improve the classification performances with a few electrodes placed based on prior knowledge of brain function areas. Typically, some researchers may simply use three bipolar or Laplacian channels located around C3, Cz and C4 (according to the standard 10-20 EEG recording system) to record motor imagery data. This simple placement can easily reduce the number of electrodes but may not always yield good classification performances. To improve classification performances, we proposed a novel method to identify subject-specific time-frequency characteristics, so as to extract effective band power features. We adapted the proposed F score to time-frequency selection problems. We also proposed a novel alternative criterion based on domain specific knowledge (such as location of brain activity during a motor imagery) to address the specific problem of the time-frequency selection. Both criteria led to good performances on two standard datasets (BCI competition IV dataset IIa and IIb) for the discrimination between imagination of right and left hand movements, using less electrodes (only two bipolar channels C3 and C4) than other methods. The time-frequency selection method based on F score was further extended to multi-class problems through one-versus-rest (OVR) strategy. The method was applied with three Laplacian channels C3, Cz and C4 for a four-class BCI problem (left hand, right hand, foot and tongue) and tested on two standard multi-class datasets (BCI competition IV dataset IIa and BCI competition III dataset IIIa). Our method is robust to unpredictable signal changes in session-to-session transfer and to artifacts. Although this direction does not directly reduce the number of electrodes by selecting the interesting channels, it does make BCI systems based on few electrodes become more reliable and easy to use. In this case, an unified hardware, like we suggested in Chapter 7 Figure 7.12, can be produced to record the data; and then to improve BCI performance by using the proposed time-frequency optimization method. Note that the number of electrodes in this approach is also reasonable and acceptable compared to the commercial BCI products.

Existing commercial BCI products require users to wait a few minutes for the calibration, so that the subject-specific parameters in the algorithm can be learned for on-line classification. Thus, the computational time for finding subject-specific parameters should also be considered when designing algorithms for BCIs for large public application, since general users may lose their interests if the calibration time is too long. All the algorithms proposed in this thesis only need a few minutes for finding the subject-specific parameters, which is also reasonable and acceptable compared to the existing commercial products.

Finally, all the algorithms have been tested on standard datasets, in order to be compared with state-of-the-art methods. The experimental results show that the proposed methods increase the efficiency of the system (e.g. reduce the hardware cost in multi-functional system, decrease the number of electrodes, shorten the training time) and/or improve the classification performances, so as to make the BCI systems efficient, economic and easy to use.

8.2 Perspectives

The applications of BCI are not only limited to providing personal assistance for disabled people, but they are more and more promising in human-computer interaction field, for servicing healthy people. The potential applications may include remotely controlling a robot to work in a dangerous environment, helping drivers to take some emergency actions to prevent a car accident, and realizing some hand-free controls in playing games. Although this thesis has proposed several solutions for some key problems, improvements and new innovations are still needed to further solve the existing challenges.

8.2.1 Signal processing and cloud computing

BCI systems are driven by brain signals. Among existing brain signals, EEG is the most promising one for daily use BCIs, in particular for non-medical purposes. However, challenges in EEG processing processing still hinder us to develop a reliable system.

On the one hand, EEG signals are sensitive to artifacts. Those artifacts can come from different sources including physiological and environment noise. Although some artifacts like signals from ocular activities can also be used as control signals in hybrid BCI, denoising techniques are still need to be further developed to improve the SNR of EEG signals. To eliminate the artifacts by requiring the subject to sit quietly and to keep claim during the control is not a good idea for general public applications, in particular for neuro-games. Different types of artifact sources, such as yawning, speaking, chewing, laughing, head movement and body movement, should be examined to learn the properties of artifacts, in order to remove them or turn them into useful control signals.

On the other hand, automatic adaptive algorithms should be further developed to overcome individual differences and data evolution. Note that the personalized design of hardware is usually expensive, which may not be interesting for general healthy users and BCI device producers. As a result, the subject-specific parameterization in signal processing and data classification is very desirable for improving BCI performances based on a few fixed electrodes. More reliable and fast algorithms should be developed to further improve the subject-specific parameterization processes. It may be interesting to employ some more machine learning methods based on SVM [66], Riemannian geometry [17] and non-negative matrix factorization [98] in the future.

To address these problems, using only the standard datasets in BCI competitions and some self-recording data are not enough. The datasets in BCI competitions only contain limited amount of EEG signals (i.e. one to three sessions) from only a few subjects (i.e. one to nine subjects). Although some datasets provide artifact-contaminated data, the artifacts are mainly from ocular activities. Other artifacts, like the artifacts from head movement and yawning, which are likely to occur in daily BCI experiences, are not presented in these datasets. Although self-recording data may compensate some limitations of standard datasets, such as providing various types of artifacts, they are not publicly accessible. To develop reliable methods, a big database containing a large amount of BCI data from a large number of subjects are required. Various types of artifacts should be presented in this database. Moreover, the data should come from different labs, and be recorded with various EEG devices. Such a database has not been built in the past decades possibly because of lacking an effective network and challenges

in big data storage. Now, cloud collaboration may help solve this challenge, so as to build a publicly accessible big BCI database and a collaborating network. Figure 8.1 presents a cloud for sharing BCI data and developing reliable algorithms, BCI cloud. In this cloud, researchers can share their data, storage and computational resources, in order to develop and test the novel algorithms with big data. Based on this cloud, more reliable algorithms will be developed to address the problems in BCI signal processing, in particular for solving the problems of artifacts, individual differences and data evolution. For example, when large data from different subjects are accessible, it may be possible to use modern statistical techniques to address the data evolution problem, and to identify the similarity and difference between subjects, which may accelerate the subject-specific parameterization processes in real applications.

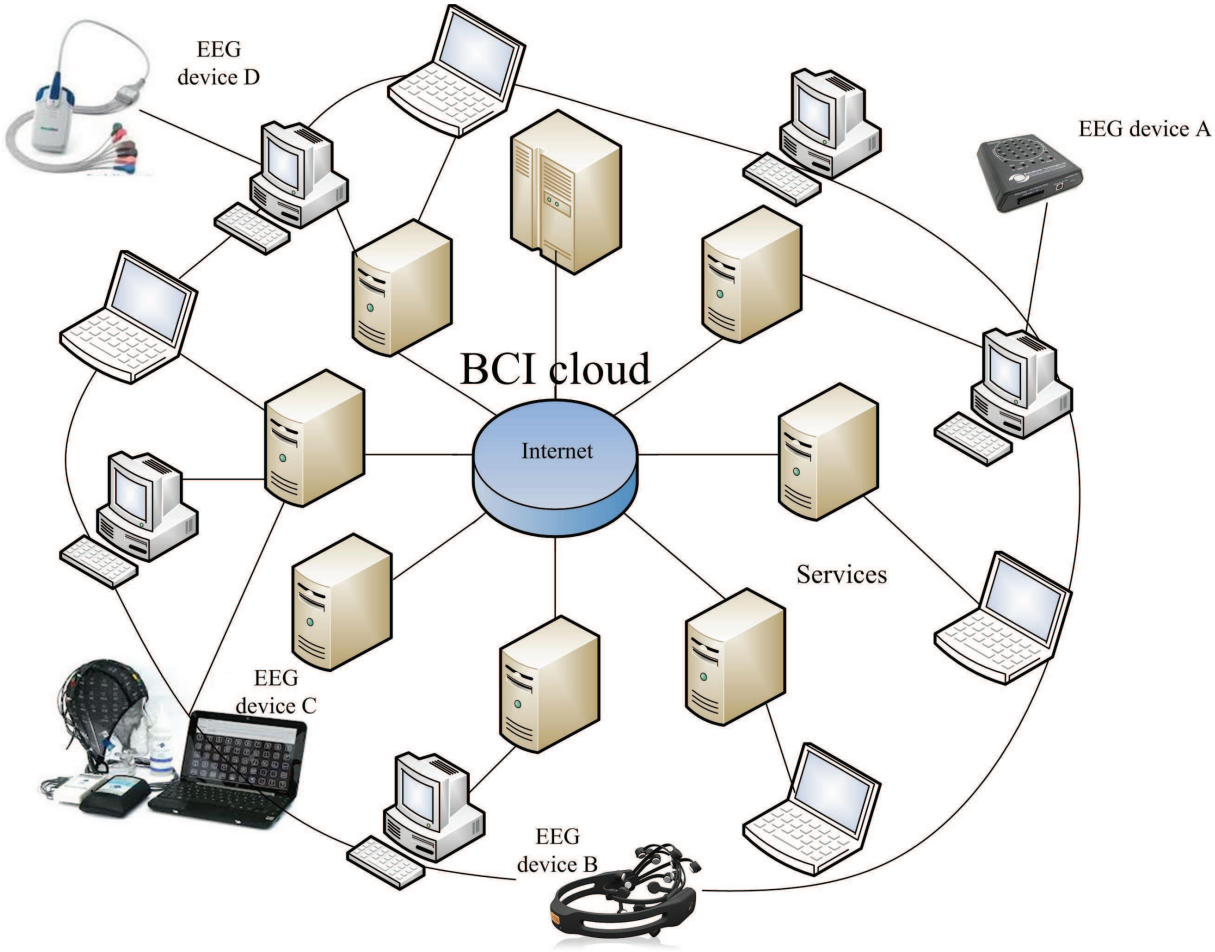


Fig. 8.1: BCI cloud: a cloud for sharing BCI data and developing reliable algorithms.

8.2.2 Brain sensors and wireless system

EEG data acquisition highly depends on the electrodes. Although the number of electrodes can be reduced to the minimum, placing the electrodes is still a problem. Traditional "wet" electrodes need a conductive gel to reduce the skin-electrode impedance. Although dry sensors are produced by several hardware producers, such as Neurosky and g.tec, the quality of the signal is not as good as with traditional electrodes. All algorithms proposed in this thesis were tested on data recorded with "wet" electrodes. In the future,

they should be further tested on data recorded using dry sensors. Recently, researchers in Freer Logic[1] have claimed that they are able to record brain activities through a device placed on the arm (see Figure 8.2). Although this technique still needs to be proved, it does attract some big companies in industry and also provides us a new scope in monitoring brain activities. Moreover, signal processing methods according to this new technique are also worth to be developed, so as to accelerate the speed of innovation in BCI fields.

Several recently launched EEG devices, such as Neurosky, Freer Logic and Emotiv, are wireless. Compared to traditional EEG devices, they are more portable. Directly transferring the brain signals through wireless network may suffer from interference and distortion. Typically, brain signals are processed in a small chip inside the device before data transmission. Most BCI algorithms including ours are mainly developed and tested on a personal computer. In the future, they should be optimized and adapted to single-chip microcomputers, so as to work with the low-power chips inside the EEG devices.

Fig. 8.2: Freer Logic device. Figure is from http://www.freerlogic.com/.

8.2.3 Integrated Information theory for BCI

BCI systems rely on the brain signal decoding. Human brain is a very complex system that we are far from fully understand. Existing methods are based on black-box analysis, so that we can only predict subject's intentions through summarizing the relationship between a mental task (input) and its corresponding changes in brain signals (output). In practice, the training data with true labels are used for summarizing this relationship (training the classifier). This decoding strategy is based on the assumption that there is a bijection between input and output. Although each part of the body has its corresponding functional area in the brain, this bijection may not always work due to the complexity of brain. Until now, the functions of invasive motor imagery BCI mainly focus on identifying hand, foot and tongue motor imagery. It is easy to separate the right hand from left hand motor imagery. However, it is difficult to identify whether the subject wants to open or close his hand, since both mental tasks can elicit an ERD in the hand representation area

[1] http://www.freerlogic.com/

of brain. Furthermore, challenges also come from identifying conscious brain activity from unconscious one. For example, the subject tends to move his right hand for realizing some controls, but he unconsciously moved his feet during the control. For the existing BCI techniques, it is difficult for the system to understand what the subject really wants.

Nowadays, a novel theory, integrated information [198, 199] may provide us potential solutions for addressing these challenges. Integrated information theory considers consciousness as the capacity of a system to integrate information, so as to define a parameter ϕ based on the information theory to quantify the consciousness. Then, it identifies different kinds of consciousness (vision, audition, body movement) by creating a multi-dimensional space called qualia space (Q-space). Details about this theory are available in [200]. Some refined versions of ϕ now can be applied to real neuronal data for measuring the level of consciousness [15, 18]. It might be interesting to introduce this theory into BCI field for brain signal decoding. By measuring the level of consciousness, it may be possible to identify conscious brain control from unconscious one. In the future, the theory in Q-space can be further developed to identify more detailed mental tasks, so as to solve problems such as separating closing hand from opening hand.

Appendix A

EEG experiment information sheet for participants

The informed consent used in the experiments is made according to WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects[1]. A copy of this consent is provided below.

[1] http://www.wma.net/en/30publications/10policies/b3/

BCI Research: EEG Experiment
INFORMATION SHEET FOR PARTICIPANTS

I have been asked to participate in a Brain-computer interface (BCI) research study conducted by _____ as a subject.

My participation in this study is entirely voluntary.

It is recommended that I read the information below and ask questions about anything I do not understand before deciding whether or not to participate.

PURPOSE OF THE STUDY

I understand that this experiment is designed to do some BCI research. I understand that the entire process will involve a scalp EEG recording session of up _____minutes/hours, which will include _____tasks.

PROCEDURES

If I volunteer to participate in this study, I will be asked to undergo _____channel topographic EEG acquisition for approximately _____ minutes/hours. During this time, I will try my best to anticipate according to instructions.

POTENTLAL RISKS AND DISCOMFORTS

I understand that EEG acquisition requires the placement of electrodes on the scalp for the purpose of recording an EEG data. There are few risks associated with this procedure. There is a remote possibility of skin irritation from the electrode cream used to attach electrodes. Techniques used to attach electrodes have been used at numerous research institutions for many years with no significantly negative side effects reported.

I understand that I can remove the electrodes at any time if I so desire and there is no risk of electroshock from this procedure. Before conducting any EEG study, I understand that the research group have try their best to avoid any psychological, legal, or financial risks for this participating in the experiment, but as always, there may be possible unforeseeable risks that have not been identified.

No personal information about me during the research project will be disclosed to anyone outside the project's

team members (TII group, WHIST Lab and TAO group) without written permission. When the results of the research are published and discussed in conferences, no information will be included that reveals my identity. Any photographs, videos, or audiotape records will be used for educational purposes. My identity will be protected or disguised. All personal information collected during this study will be stored in the research database at the TII group, WHIST Lab and TAO group. Records will be coded and anonymized to prevent access by any unauthorized personnel.

RIGHTS OF RESEARCH SUBJECTS

I may withdraw my consent at any time and discontinue participation without penalty. I am not waiving any legal rights or remedies because of my participation in this research study.

SIGNATURE OF RESEARCH SUBJECT OR LEGAL REPRESENTATIVE

I have read and understand the information provided above. I have been given an opportunity to ask questions and all of my questions have been answered to my satisfaction.

BY SIGNING THIS FORM, I WILLINGLY AGREE TO PARTICIPATE IN THE RESEARCH IT DESCRIBES.

Name of Subject

_____

Signature of Subject **Date**

SIGNATURE OF INVESTIGATOR

I have explained the research to the subject or his/her legal representative, and answered all of his/her questions. I believe that he/she understands the information described in this document and freely consents to participate.

Name of Investigator

_____

Names of Labs/research groups

TII group, TSI, Télécom ParisTech/CNRS UMR 5141 LTCI
TAO group, INRIA Saclay/CNRS/LRI, Université Paris-Sud
WHIST Lab, France

Signature of Investigator **Date**

Appendix B

Paired *t*-test

In this thesis, experimental results were submitted to a statistical hypothesis test, to check whether the outcomes of a study would lead to a rejection of a null hypothesis for a pre-determined threshold probability, the significance level [168]. The null hypothesis typically refers to a general or default position, for example, there is no difference between two populations, as was the case in most of our experiments.

Paired t-test is a kind of statistical hypothesis test used to compare the mean values of two paired populations, where each sample in one population can be paired with the corresponding sample in the other population. The number of samples in two populations should be same. For example, we want to compare the classification performances (i.e. kappa coefficient, κ) of two methods, A and B, on a dataset containing nine subjects (see Table B.1). For each subject, we have the classification results obtained by methods A and B, and they are paired.

Table B.1: Classification performances (i.e. kappa coefficient, κ) of two methods, A and B, on a dataset containing nine subjects

	Subjects								
	1	2	3	4	5	6	7	8	9
A	0.68	0.42	0.75	0.48	0.40	0.27	0.77	0.75	0.61
B	0.69	0.34	0.71	0.44	0.16	0.21	0.66	0.73	0.69

In paired t-test, the null hypothesis is that the mean difference between paired populations is zero. Let us denote by $X = [x_1, x_2, ..., x_N]$ and $Y = [y_1, y_2, ..., y_N]$ the two paired populations. To test whether the null hypothesis is true, the procedure is as follows:

1. Calculate the signed difference between the two populations on each pair:

$$d_i = y_i - x_i, i = 1, 2, ...N \tag{B.1}$$

2. Calculate the mean difference, μ_d and the standard deviation of the difference σ_d

3. Calculate the standard error of the mean difference:

$$SE(\mu_d) = \frac{\sigma_d}{\sqrt{N}} \tag{B.2}$$

4. Dividing the mean μ_d by the standard error $SE(\mu_d)$ of the mean yields a test statistic, t_s:

$$t_s = \frac{\mu_d}{SE(\mu_d)} \tag{B.3}$$

which follows a t-distribution [168] with $N - 1$ degrees of freedom.

5. The probability that the null hypothesis is true, P value, is obtained by checking the t-distribution with df and t_s (see Figure B.2)

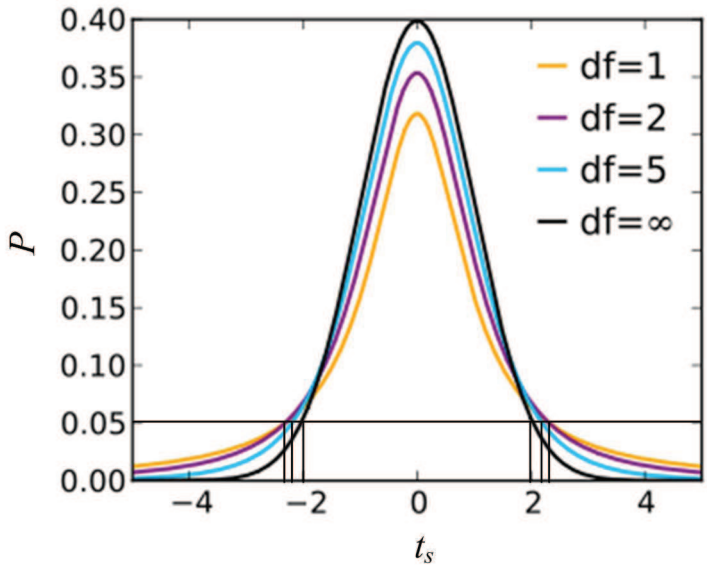


Fig. B.1: t-distribution for different degrees of freedom df. The horizontal line indicates the significance level of 0.05, and the vertical lines show the values of t_s at the significance level of 0.05 for different degrees of freedom.

Typically, the significance level is set to 0.05. Thus, the null hypothesis is rejected when the P value is less than 0.05.

Using the example in Table B.1, the signed difference for each pair is computed and given in Table B.2.

Table B.2: Signed difference for each pair

	Subjects								
	1	2	3	4	5	6	7	8	9
d_i	0.01	-0.08	-0.04	-0.04	-0.24	-0.06	-0.11	-0.02	0.08

Calculating the mean value and standard deviation of the differences gives:

$$\mu_d = -0.0556, \text{ and } \sigma_d = 0.088$$

Then,

$$SE(\sigma_d) = \frac{\sigma_d}{\sqrt{9}} = 0.0293$$

So we have $t_s = -1.8929$ with $df = 8$. Checking the t-distribution with $df = 8$ (see Figure B.2), we can see $P > 0.05$. Therefore, the null hypothesis is true for this example, and we can conclude that there is no significant difference between the classification results obtained by methods A and B.

For more details about paired t-test, please refer to [184]; and for other statistical hypothesis tests, see [168].

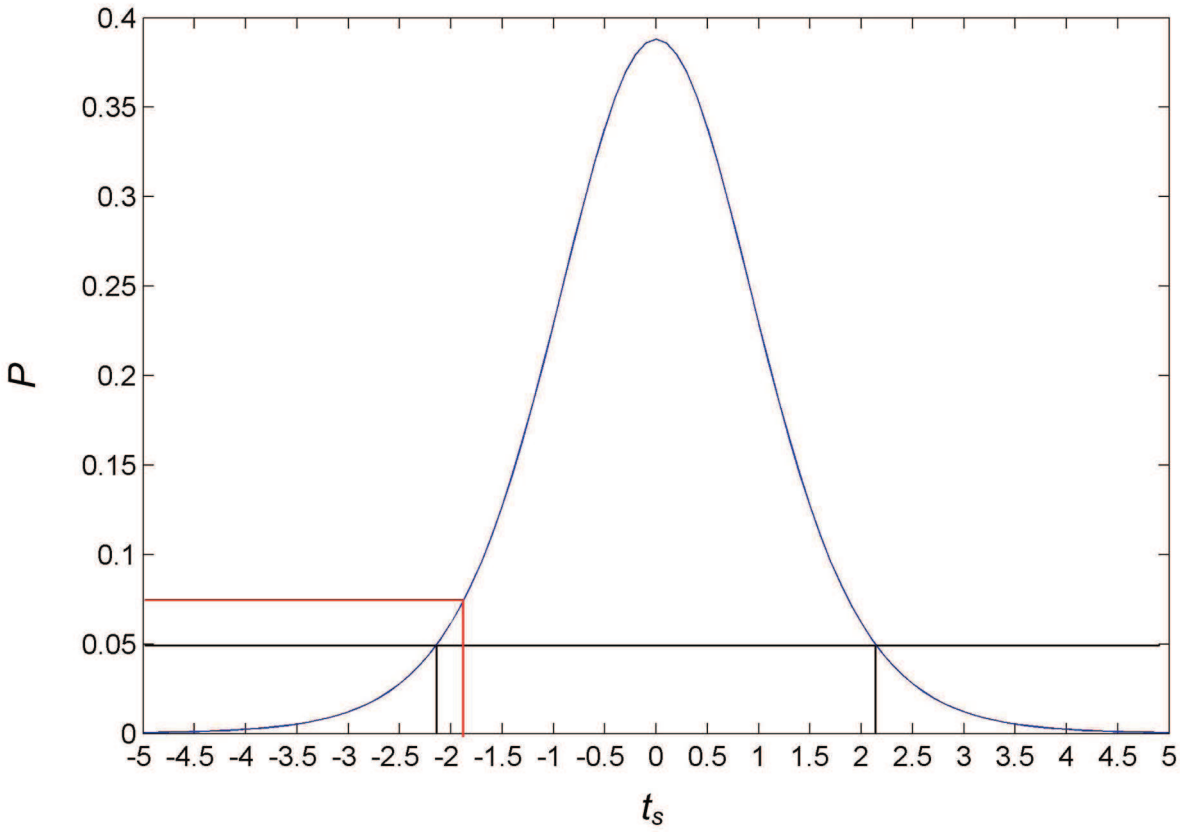


Fig. B.2: t-distribution for degree of freedom $df = 8$. The black horizontal line indicates the significance level of 0.05, and the black vertical lines show the values of t_s at the significance level of 0.05. The red line shows the P value for $t_s = -1.8929$.

Appendix C

Publications

1. Y Yang, S Chevallier, J Wiart and I Bloch, Time-frequency optimization for discrimination between imagination of right and left hand movements based on two bipolar electroencephalography channels, Submission to *Pattern Recognition* [J].

2. Y Yang, O Kyrgzov, S Chevallier, J Wiart and I Bloch, Subject-specific channel selection using time information for motor imagery BCIs, Submission to *IEEE transactions on Biomedical Engineering* [J].

3. Y Yang, O Kyrgzov, I Bloch and Joe Wiart, Subject-specific channel selection for classification of motor imagery electroencephagraphic data. In *38th International Conference on Acoustics, Speech, and Signal Processing* (*ICASSP 2013*), Vancouver, Canada, May 2013, pp. 1277-1280 [C]

4. Y Yang, J Wiart and I Bloch, Towards next generation human-computer interaction – Brain-computer interfaces: applications and challenges, In *1st International Symposium of Chinese Computer Human Interaction* (*Chinese CHI 2013*), Paris, France, April, 2013 [C].

5. O Kyrgyzov, I Bloch, Y Yang, Joe Wiart, A Souloumiac, Data ranking and clustering via normalized graph cut based on asymmetric affinity, Submission to *17th International Conference on Image Analysis and Processing* (*ICIAP 2013*) [C]

6. Y Yang, S Chevallier, J Wiart and I Bloch, Time-frequency selection in two bipolar channels for improving the classification of motor imagery EEG, In *Proceedings of 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*(*IEEE EMBC 2012*), San Diego, CA, USA, August 2012, pp. 2744-2747 [C].

7. Y Yang, S Chevallier, J Wiart, I Bloch, Automatic selection of the number of spatial filters for motor-imagery BCI, In *Proceedings of 20th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning* (*ESANN 2012*), Bruges, Belgium, 25-27 April, 2012, pp. 109-114 [C].

8. Y Yang, S Chevallier, J Wiart and I Bloch, A self-paced hybrid BCI based on EEG and EOG, In *Proceedings of 3rd Tools for Brain-Computer Interaction Workshop* (*TOBI 2012*), Wurzburg, Germany, March 2012, pp. 33-34 [C].

Bibliography

[1] B. Z. Allison, C. Brunner, V. Kaiser, G. R. Müller-Putz, C. Neuper, and G. Pfurtscheller. Toward a hybrid brain–computer interface based on imagined movement and visual attention. *Journal of Neural Engineering*, 7:026007, 2010. 25

[2] B. Z. Allison and J. A. Pineda. ERPs evoked by different matrix sizes: Implications for a brain computer interface (BCI) system. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11(2):110–113, 2003. 27

[3] B. Z. Allison and J. A. Pineda. Effects of SOA and flash pattern manipulations on ERPs, performance, and preference: implications for a BCI system. *International Journal of Psychophysiology*, 59(2):127–140, 2006. 28

[4] B. Z. Allison, E. W. Wolpaw, and J. R. Wolpaw. Brain-computer interface systems: progress and prospects. *Expert Review of Medical Devices*, 4(4):463–474, 2007. 24, 25, 27, 28, 29, 60

[5] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang. Filter bank common spatial pattern algorithm on BCI Competition IV datasets 2a and 2b. *Frontiers in Neuroscience*, 6:39:1–9, 2012. 17, 18, 48, 50, 110, 123, 126, 127, 128

[6] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan. Filter bank common spatial pattern (FBCSP) in brain-computer interface. In *IEEE International Joint Conference on Neural Networks (IJCNN 2008)*, pages 2390–2397, 2008. 48, 50

[7] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan. Robust filter bank common spatial pattern (RFBCSP) in motor-imagery-based brain-computer interface. In *31th IEEE Annual International Conference of the Engineering in Medicine and Biology Society (EMBC 2009)*, pages 578–581, 2009. 50

[8] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan. Mutual information-based selection of optimal spatial-temporal patterns for single-trial EEG-based BCIs. *Pattern Recognition*, 45:2137–2144, 2011. 17, 50, 110, 111, 127

[9] K. Ansari-Asl, G. Chanel, and T. Pun. A channel selection method for EEG classification in emotion assessment based on synchronization likelihood. In *15th European Signal Processing Conference (EUSIPCO 2007)*, pages 1241–1245, 2007. 45

[10] M. Arvaneh, C. Guan, K. K. Ang, and C. Quek. Optimizing the channel selection and classification accuracy in EEG-based BCI. *IEEE Transactions Biomedical Engineering*, 58(6):1865–1873, 2011. 17, 18, 127, 128

[11] C. Babiloni, F. Babiloni, F. Carducci, F. Cincotti, C. D. Percio, D. V. Moretti, and P. M. Rossini. Quantitative EEG: modeling time, space, and phase of brain oscillatory activity. *Supplements to Clinical Neurophysiology*, 54:284–288, 2002. 39

[12] F. Babiloni, L. Bianchi, F. Semeraro, J. R. Millán, J. Mouriño, A. Cattini, S. Salinari, M. G. Marciani, and F. Cincotti. Mahalanobis distance-based classifiers are able to recognize EEG patterns by using few EEG electrodes. In *23rd IEEE Annual International Conference of Engineering in Medicine and Biology Society (EMBS 2001)*, volume 1, pages 651–654, 2001. 42, 92

[13] F. Babiloni, F. Cincotti, L. Bianchi, G. Pirri, J. del R Millán, J. Mouriño, S. Salinari, and M. Grazia Marciani. Recognition of imagined hand movements with low resolution surface Laplacian and linear classifiers. *Medical Engineering & Physics*, 23(5):323–328, 2001. 40

[14] F. Babiloni, F. Cincotti, F. Carducci, P. M. Rossini, and C. Babiloni. Spatial enhancement of EEG data by surface Laplacian estimation: the use of magnetic resonance imaging-based head models. *Clinical Neurophysiology*, 112(5):724–727, 2001. 34

[15] D. Balduzzi and G. Tononi. Integrated information in discrete dynamical systems: motivation and theoretical framework. *PLoS Computational Biology*, 4(6):e1000091, 2008. 154

[16] A. Barachant, T. Aksenova, and S. Bonnet. Filtrage spatial robuste à partir d'un sous-ensemble optimal d'électrodes en BCI EEG. In *XXIIe colloque GRETSI (traitement du signal et des images)*, 2009. 34, 52

[17] A. Barachant and S. Bonnet. Channel selection procedure using Riemannian distance for BCI applications. In *5th International IEEE/EMBS Conference on Neural Engineering (NER 2011)*, pages 348–351, 2011. 10, 34, 52, 54, 92, 95, 108, 151

[18] A. B. Barrett and A. K. Seth. Practical measures of integrated information for time-series data. *PLoS Computational Biology*, 7(1):e1001052, 2011. 154

[19] E. Basar, C. Basar-Eroglu, S. Karakas, and M. Schurmann. Gamma, alpha, delta, and theta oscillations govern cognitive processes. *International Journal of Psychophysiology*, 39(2-3):241–248, 2001. 60

[20] E. Basar, M. Schurmann, C. Basar-Eroglu, and S. Karakas. Alpha oscillations in brain functioning: an integrative theory. *International Journal of Psychophysiology*, 26(1-3):5–29, 1997. 60

[21] A. Bashashati, M. Fatourechi, R. K. Ward, and G. E. Birch. User customization of the feature generator of an asynchronous brain interface. *Annals of Biomedical Engineering*, 34(6):1051–1060, 2006. 43

[22] J. D. Bayliss, S. A. Inverso, and A. Tentler. Changing the P300 brain computer interface. *CyberPsychology & Behavior*, 7(6):694–704, 2004. 20

[23] P. Berg and M. Scherg. Dipole models of eye movements and blinks. *Electroencephalography and Clinical Neurophysiology*, 79(1):36–44, 1991. 66

[24] M. Besserve, J. Martinerie, and L. Garnero. Improving quantification of functional networks with EEG inverse problem: Evidence from a decoding point of view. *NeuroImage*, 55(4):1536–1547, 2011. 26

[25] F. Beverina, G. Palmas, S. Silvoni, F. Piccione, and S. Giove. User adaptive BCIs: SSVEP and P300 based interfaces. *PsychNology Journal*, 1(4):331–354, 2003. 28

[26] N. Birbaumer. Slow cortical potentials: Plasticity, operant control, and behavioral effects. *The Neuroscientist*, 5(2):74–78, 1999. 27

[27] N. Birbaumer, T. Elbert, A. G. M. Canavan, and B. Rockstroh. Slow potentials of the cerebral cortex and behavior. *Physiological Reviews*, 70:1–41, 1990. 27

[28] N. Birbaumer, N. Ghanayim, T. Hinterberger, I. Iversen, B. Kotchoubey, A. Kübler, J. Perelmouter, E. Taub, and H. Flor. A spelling device for the paralysed. *Nature*, 398(6725):297–298, 1999. 27

[29] N. Birbaumer, A. Kubler, N. Ghanayim, T. Hinterberger, J. Perelmouter, J. Kaiser, I. Iversen, B. Kotchoubey, N. Neumann, and H. Flor. The thought translation device (TTD) for completely paralyzed patients. *IEEE Transactions on Rehabilitation Engineering*, 8(2):190–193, 2000. 27, 30

[30] C.M. Bishop. *Neural networks for pattern recognition*. Clarendon press Oxford, 1995. 52

[31] B. Blankertz. BCI competition IV, 2008. `http://www.bbci.de/competition/iv/`. 120

[32] B. Blankertz, G. Curio, and K. R. Müller. Classifying single trial EEG: Towards brain computer interfacing. *Advances in Neural Information Processing Systems*, 1:157–164, 2002. 35, 47

[33] B. Blankertz, G. Dornhege, M. Krauledat, K. R. Müller, V. Kunzmann, F. Losch, and G. Curio. The Berlin brain-computer interface: EEG-based communication without subject training. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(2):147–152, 2006. 12, 34, 51, 52, 53, 54, 96

[34] B. Blankertz, M. Kawanabe, R. Tomioka, F. Hohlefeld, V. Nikulin, and K. R. Müller. Invariant common spatial patterns: Alleviating nonstationarities in brain-computer interfacing. *Advances in Neural Information Processing Systems*, 20, 2008. 43

[35] B. Blankertz, K. R. Müller, D. J. Krusienski, G. Schalk, J. R. Wolpaw, A. Schlögl, G. Pfurtscheller, J. R. Millán, M. Schroder, and N. Birbaumer. The BCI competition III: Validating alternative approaches to actual BCI problems. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(2):153–159, 2006. 35, 101, 144, 145

[36] B. Blankertz, M. Tangermann, C. Vidaurre, S. Fazli, C. Sannelli, S. Haufe, C. Maeder, L. Ramsey, I. Sturm, G. Curio, et al. The Berlin brain–computer interface: non-medical uses of BCI technology. *Frontiers in Neuroscience*, 4, 2010. 1, 19, 33

[37] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K. R. Müller. Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal Processing Magazine*, 25(1):41–56, 2008. 5, 6, 7, 15, 16, 40, 41, 42, 47, 76, 78, 82, 85, 92, 113, 114, 117

[38] M. Brown, M. Marmor, E. Zrenner, M. Brigell, and M. Bach. ISCEV standard for clinical electro-oculography (EOG) 2006. *Documenta Ophthalmologica*, 113(3):205–212, 2006. 62, 67

[39] C. Brunner, M. Naeem, R. Leeb, B. Graimann, and G. Pfurtscheller. 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, 2007. 6, 16, 50, 80, 82, 117, 120, 133, 134, 135, 136

[40] P. Brunner, L. Bianchi, C. Guger, F. Cincotti, and G. Schalk. Current trends in hardware and software for brain–computer interfaces (BCIs). *Journal of Neural Engineering*, 8:025001, 2011. 32

[41] A. Bulling, J. A. Ward, H. Gellersen, and G. Troster. Eye movement analysis for activity recognition using electrooculography. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(4):741–753, 2011. 64, 66, 72

[42] H. Cecotti. A Self-Paced and Calibration-Less SSVEP-Based Brain–Computer Interface Speller. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 18(2):127–133, 2010. 28

[43] H. Cecotti. Spelling with non-invasive brain-computer interfaces–current and future trends. *Journal of Physiology-Paris*, 2011. 32

[44] R. D. Chervin, J. W. Burns, N. S. Subotic, and C. Roussi. System and method for analysis of respiratory cycle-related EEG changes in sleep-disordered breathing, March 13 2007. US Patent 7,190,995. 65

[45] Z. Y. Chin, K. K. Ang, C. Wang, C. Guan, and H. Zhang. Multi-class filter bank common spatial pattern for four-class motor imagery BCI. In *31th IEEE Annual International Conference of the Engineering in Medicine and Biology Society (EMBC 2009)*, pages 571–574, 2009. 8, 50, 140

[46] F. Cincotti, D. Mattia, C. Babiloni, F. Carducci, L. Bianchi, J.R. Millán, J. Mouriño, S. Salinari, MG Marciani, and F. Babiloni. Classification of EEG mental patterns by using two scalp electrodes and Mahalanobis distance-based classifiers. *Methods of Information in Medicine - Methodik der Information in der Medizin*, 41(4):337–341, 2002. 42, 92

[47] J. Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46, 1960. 37

[48] D. Coyle, J. Garcia, A. R. Satti, and T. M. McGinnity. EEG-based continuous control of a game using a three channel motor imagery BCI: BCI game. In *IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB 2011)*, pages 1–7, 2011. 33

[49] R. J. Croft and R. J. Barry. EOG correction of blinks with saccade coefficients: a test and revision of the aligned-artefact average solution. *Clinical Neurophysiology*, 111(3):444–451, 2000. 62, 66

[50] R. J. Croft and R. J. Barry. Removal of ocular artifact from the EEG: a review. *Clinical Neurophysiology*, 30(1):5–19, 2000. 75

[51] R. J. Croft, J. S. Chandler, R. J. Barry, N. R. Cooper, and A. R. Clarke. EOG correction: A comparison of four methods. *Psychophysiology*, 42(1):16–24, 2005. 66

[52] E. A. Curran and M. J. Stokes. Learning to control brain activity: A review of the production and control of EEG components for driving brain-computer interface (BCI) systems. *Brain and Cognition*, 51(3):326–336, 2003. 1, 19

[53] B. Dal Seno, M. Matteucci, and L. Mainardi. Online detection of P300 and error potentials in a BCI speller. *Computational Intelligence and Neuroscience*, 2010:1, 2010. 27

[54] J. J. Daly, R. Cheng, J. Rogers, K. Litinas, K. Hrovat, and M. Dohring. Feasibility of a new application of noninvasive brain computer interface (BCI): a case study of training for recovery of volitional motor control after stroke. *Journal of Neurologic Physical Therapy*, 33(4):203–211, 2009. 20

[55] S. G. Danko. The reflection of different aspects of brain activation in the electroencephalogram: Quantitative electroencephalography of the states of rest with the eyes open and closed. *Human Physiology*, 32(4):377–388, 2006. 65

[56] J. de la Vega Arias, C. Hintermüller, and C. Guger. Generic brain-computer interface for social and human-computer interaction. In *5th International Conference on Advances in Computer-Human Interactions*, pages 145–149, 2012. 32

[57] J. Ding, G. Sperling, and R. Srinivasan. Attentional modulation of SSVEP power depends on the network tagged by the flicker frequency. *Cerebral Cortex*, 16(7):1016, 2006. 28

[58] E. Donchin, K. M. Spencer, and R. Wijesinghe. The mental prosthesis: assessing the speed of a P300-based brain-computer interface. *IEEE Transactions on Rehabilitation Engineering*, 8(2):174–179, 2002. 27

[59] G. Dornhege, B. Blankertz, G. Curio, and K. R. Müller. Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms. *IEEE Transactions on Biomedical Engineering*, 51(6):993–1002, 2004. 49

[60] G. Dornhege, B. Blankertz, G. Curio, and K. R. Müller. Increase information transfer rates in BCI by CSP extension to multi-class. *Advances in Neural Information Processing Systems*, 16:733–740, 2004. 6, 8, 76, 79, 143

[61] G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio, and K. R. Müller. Combined optimization of spatial and temporal filters for improving brain-computer interfacing. *IEEE Transactions on Biomedical Engineering*, 53(11):2274–2281, 2006. 48

[62] J. Farquhar, N. J. Hill, T. N. Lal, and B. Schölkopf. Regularised CSP for sensor selection in BCI. In *Proceedings of the 3rd International Brain-Computer Interface Workshop and Training Course*, pages 14–15. Verlag der Technischen Universität Graz, 2006. 42, 52

[63] L. A. Farwell and E. Donchin. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*, 70(6):510–523, 1988. 27, 28, 30

[64] M. Fatourechi, A. Bashashati, R. K. Ward, G. E. Birch, et al. EMG and EOG artifacts in brain computer interface systems: A survey. *Clinical Neurophysiology*, 118(3):480–494, 2007. 42, 59, 60

[65] A. Furdea, S. Halder, D. J. Krusienski, D. Bross, F. Nijboer, N. Birbaumer, and A. Kubler. An auditory oddball (P300) spelling system for brain-computer interfaces. *Psychophysiology*, 46(3):617–625, 2009. 28

[66] D. Garrett, D. A. Peterson, C. W. Anderson, and M. H. Thaut. Comparison of linear, nonlinear, and feature selection methods for EEG signal classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11(2):141–144, 2003. 46, 47, 151

[67] B. Graimann, J. E. Huggins, S. P. Levine, and G. Pfurtscheller. Visualization of significant ERD/ERS patterns in multichannel EEG and ECoG data. *Clinical Neurophysiology*, 113(1):43–47, 2002. 112, 120

[68] B. Graimann and G. Pfurtscheller. Quantification and visualization of event-related changes in oscillatory brain activity in the time-frequency domain. *Progress in Brain Research*, 159:79–97, 2006. 24, 33, 38, 41

[69] C. Guger, S. Daban, E. Sellers, C. Holzner, G. Krausz, R. Carabalona, F. Gramatica, and G. Edlinger. How many people are able to control a P300-based brain-computer interface (BCI)? *Neuroscience Letters*, 462(1):94–98, 2009. 28

[70] C. Guger, G. Krausz, B. Z. Allison, and G. Edlinger. Comparison of dry and gel based electrodes for P300 brain–computer interfaces. *Frontiers in Neuroscience*, 6:60:1–7. 32

[71] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik. Gene selection for cancer classification using support vector machines. *Machine Learning*, 46(1):389–422, 2002. 56, 88

[72] D. Heingartner. Mental block. *IEEE Spectrum*, 46(1):42–43, 2009. 32

[73] T. Hinterberger, J.M. Houtkooper, and B. Kotchoubey. Effects of feedback control on slow cortical potentials and random events. In *Parapsychological Association Convention*, pages 39–50, 2004. 27

[74] B. Hjorth. EEG analysis based on time domain properties. *Electroencephalography and Clinical Neurophysiology*, 29(3):306–310, 1970. 45

[75] B. Hjorth. An on-line transformation of EEG scalp potentials into orthogonal source derivations. *Electroencephalography and Clinical Neurophysiology*, 39(5):526–530, 1975. 40, 137

[76] C. Hondrou and G. Caridakis. Affective, natural interaction using EEG: Sensors, application and future directions. *Artificial Intelligence: Theories and Applications*, pages 331–338, 2012. 30

[77] J. Hori, K. Sakano, and Y. Saitoh. Development of a communication support device controlled by eye movements and voluntary eye blink. *IEICE Transactions on Information and Systems*, 89(6):1790–1797, 2006. 60

[78] P. Horki, T. Solis-Escalante, C. Neuper, and G. Müller-Putz. Combined motor imagery and SSVEP based BCI control of a 2 DoF artificial upper limb. *Medical and Biological Engineering and Computing*, pages 1–11, 2011. 40

[79] Neurosky Inc. Neurosky's esense meters and detection of mental state, 2009. *http://company.neurosky.com/files/neurosky_esense_whitepaper.pdf*. 30

[80] N. F. Ince, S. Arica, and A. Tewfik. Classification of single trial motor imagery EEG recordings with subject adapted non-dyadic arbitrary time–frequency tilings. *Journal of Neural Engineering*, 3:235, 2006. 50, 51, 110

[81] N. F. Ince, F. Goksu, A. H. Tewfik, and S. Arica. Adapting subject specific motor imagery EEG patterns in space-time-frequency for a brain computer interface. *Biomedical Signal Processing and Control*, 4(3):236–246, 2009. 45, 51

[82] A. Jaimes and N. Sebe. Multimodal human-computer interaction: A survey. *Computer Vision and Image Understanding*, 108(1-2):116–134, 2007. 60

[83] A. Jain and D. Zongker. Feature selection: Evaluation, application, and small sample performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2):153–158, 1997. 46, 95, 143

[84] H. H. Jasper. The ten-twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology*, 10(2):371–375, 1958. 31, 38, 92, 100

[85] R. A. Johnson and D. W Wichern. *Applied multivariate statistical analysis*. Prentice hall Englewood Cliffs, NJ, 1992. 133

[86] S. J. Johnstone, R. Blackman, and J. M. Bruggemann. Eeg from a single-channel dry-sensor recording device. *Clinical EEG and Neuroscience*, 43(2):112–120, 2012. 30

[87] I Jolliffe. *Principal component analysis*. Wiley Online Library, 2005. 144

[88] J. Kalcher and G. Pfurtscheller. Discrimination between phase-locked and non-phase-locked event-related EEG activity. *Electroencephalography and Clinical Neurophysiology*, 94(5):381–384, 1995. 20

[89] S. P. Kelly, E. Lalor, R. B. Reilly, and J. J. Foxe. Independent brain computer interface control using visual spatial attention-dependent modulations of parieto-occipital alpha. In *2nd International IEEE EMBS Conference on Neural Engineering*, pages 667–670, 2005. 60

[90] S. P. Kelly, E. C. Lalor, C. Finucane, G. McDarby, and R. B. Reilly. Visual spatial attention control in an independent brain-computer interface. *IEEE Transactions on Biomedical Engineering*, 52(9):1588–1596, 2005. 29

[91] J. Kennedy and R. C. Eberhart. A discrete binary version of the particle swarm algorithm. In *IEEE International Conference on Systems, Man, and Cybernetics and Computational Cybernetics and Simulation (SMC 1997)*, volume 5, pages 4104–4108, 1997. 54

[92] Z. J. Koles. The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG. *Electroencephalography and Clinical Neurophysiology*, 79(6):440–447, 1991. 78

[93] A. Kostov and M. Polak. Parallel man-machine training in development of EEG-based cursor control. *IEEE Transactions on Rehabilitation Engineering*, 8(2):203–205, 2000. 29

[94] R. Krepki, B. Blankertz, G. Curio, and K. R. Müller. The Berlin brain-computer interface (BBCI)–towards a new communication channel for online control in gaming applications. *Multimedia Tools and Applications*, 33(1):73–90, 2007. 20, 21, 33

[95] A. Kubler, A. Furdea, S. Halder, E.M. Hammer, F. Nijboer, and B. Kotchoubey. A brain–computer interface controlled auditory event-related potential (P300) spelling system for locked-in patients. *Annals of the New York Academy of Sciences*, 1157(Disorders of Consciousness):90–100, 2009. 28

[96] T. N. Lal, T. Hinterberger, G. Widman, M. Schröder, J. Hill, W. Rosenstiel, C. Elger, B. Schölkopf, and N. Birbaumer. Methods towards invasive human brain computer interfaces. *Advances in Neural Information Processing Systems*, 17. 35, 55

[97] T. N. Lal, M. Schröder, T. Hinterberger, J. Weston, M. Bogdan, N. Birbaumer, and B. Schölkopf. Support vector channel selection in BCI. *IEEE Transactions on Biomedical Engineering*, 51(6):1003–1010, 2004. 10, 34, 51, 55, 56, 57, 77, 88, 92, 95, 108, 140

[98] H. Lee, J. Yoo, and S. Choi. Semi-supervised nonnegative matrix factorization. *IEEE Signal Processing Letters*, 17(1):4–7, 2010. 151

[99] R. Leeb, D. Friedman, G. R. Müller-Putz, R. Scherer, M. Slater, and G. Pfurtscheller. Self-paced (asynchronous) BCI control of a wheelchair in virtual environments: a case study with a tetraplegic. *Computational Intelligence and Neuroscience*, 2007:7, 2007. 29

[100] R. Leeb, F. Lee, C. Keinrath, R. Scherer, H. Bischof, and G. Pfurtscheller. Brain–computer communication: motivation, aim, and impact of exploring a virtual apartment. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 15(4):473–482, 2007. 16, 17, 36, 50, 110, 117, 118, 120, 131

[101] R. Leeb, H. Sagha, R. Chavarriaga, and J. R. Millán. A hybrid brain–computer interface based on the fusion of electroencephalographic and electromyographic activities. *Journal of Neural Engineering*, 8:025011, 2011. 25, 59

[102] R. Leeb, R. Scherer, F. Lee, H. Bischof, and G. Pfurtscheller. Navigation in virtual environments through motor imagery. In *9th Computer Vision Winter Workshop, CVWW*, volume 4, pages 99–108, 2004. 33

[103] S. Lemm, B. Blankertz, G. Curio, and K. R. Müller. Spatio-spectral filters for improving the classification of single trial EEG. *IEEE Transactions on Biomedical Engineering*, 52(9):1541–1548, 2005. 48

[104] E. C. Leuthardt, G. Schalk, J. R. Wolpaw, J. G. Ojemann, and D. W. Moran. A brain–computer interface using electrocorticographic signals in humans. *Journal of Neural Engineering*, 1:63, 2004. 1, 19, 25

[105] H. T. Lin and C. J. Lin. A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods. Technical report, Department of Computer Science, National Taiwan University. 47

[106] G. Liu, G. Huang, J. Meng, and X. Zhu. A frequency-weighted method combined with common spatial patterns for electroencephalogram classification in brain–computer interface. *Biomedical Signal Processing and Control*, 5(2):174–180, 2010. 48

[107] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi. A review of classification algorithms for EEG-based brain–computer interfaces. *Journal of Neural Engineering*, 4:R1–R14, 2007. 43, 45, 46, 47, 78, 94, 95, 111

[108] B. Lou, B. Hong, X. Gao, and S. Gao. Bipolar electrode selection for a motor imagery based brain–computer interface. *Journal of Neural Engineering*, 5:342–349, 2008. 16, 38, 42, 51, 110, 114, 131

[109] A. Luo and T. J. Sullivan. A user-friendly SSVEP-based brain–computer interface using a time-domain classifier. *Journal of Neural Engineering*, 7:026010, 2010. 28

[110] J. Lv and M. Liu. Common spatial pattern and particle swarm optimization for channel selection in BCI. In *3rd IEEE International Conference on Innovative Computing Information and Control (ICICIC'08)*, pages 457–457, 2008. 5, 35, 52, 54, 55, 78

[111] A. Magliero, T.R. Bashore, M.G.H. Coles, and E. Donchin. On the dependence of P300 latency on stimulus evaluation processes. *Psychophysiology*, 21(2):171–186, 1984. 27

[112] S Makeig, A. J. Bell, T. P. Jung, and T. J. Sejnowski. Independent component analysis of electroencephalographic data. *Advances in Neural Information Processing Systems*, pages 145–151, 1996. 143, 145

[113] J. Malmivuo and R. Plonsey. *Bioelectromagnetism: principles and applications of bioelectric and biomagnetic fields*. Oxford University Press, USA, 1995. 38

[114] K. Z. Mao. RBF neural network center selection based on Fisher ratio class separability measure. *IEEE Transactions on Neural Networks*, 13(5):1211–1217, 2002. 143, 145

[115] K. V Mardia. Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57(3):519–530, 1970. 133, 134

[116] D. J. McFarland, C. W Anderson, K-R Müller, A. Schlögl, and D. J. Krusienski. Bci meeting 2005-workshop on BCI signal processing: feature extraction and translation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(2):135–138, 2006. 20

[117] D. J. McFarland, L. M. McCane, S. V. David, and J. R. Wolpaw. Spatial filter selection for eeg-based communication. *Electroencephalography and Clinical Neurophysiology*, 103(3):386–394, 1997. 38, 39, 40

[118] D. J. McFarland, L. M. McCane, and J. R. Wolpaw. EEG-based communication and control: short-term role of feedback. *IEEE Transactions on Rehabilitation Engineering*, 6(1):7–11, 1998. 29

[119] D. J. McFarland, L. A. Miner, T. M. Vaughan, and J. R. Wolpaw. Mu and beta rhythm topographies during motor imagery and actual movements. *Brain Topography*, 12(3):177–186, 2000. 20, 29

[120] D. J. McFarland, W. A. Sarnacki, T. M. Vaughan, and J. R. Wolpaw. Brain-computer interface (BCI) operation: signal and noise during early training sessions. *Clinical Neurophysiology*, 116(1):56, 2005. 43

[121] D. J. McFarland and J. R. Wolpaw. Brain-computer interface operation of robotic and prosthetic devices. *Computer*, 41(10):52–56, 2008. 24, 29

[122] P. Meinicke, M. Kaper, F. Hoppe, M. Heumann, and H. Ritter. Improving transfer rates in brain computer interfacing: a case study. *Advances in Neural Information Processing Systems*, pages 1131–1138, 2003. 28

[123] J. Mellinger, G. Schalk, C. Braun, H. Preissl, W. Rosenstiel, N. Birbaumer, and A. Kubler. An MEG-based brain-computer interface (BCI). *Neuroimage*, 36(3):581–593, 2007. 1, 19, 25

[124] J. Meng, G. Liu, G. Huang, and X. Zhu. Automated selecting subset of channels based on CSP in motor imagery brain-computer interface system. In *IEEE International Conference on Robotics and Biomimetics (ROBIO 2009)*, pages 2290–2294, 2009. 34, 51, 52, 53, 54

[125] J. R. Millán and J. Mouriño. 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, 2003. 60

[126] J. R. Millán, R. Rupp, G. R. Müller-Putz, R. Murray-Smith, C. Giugliemma, M. Tangermann, C. Vidaurre, F. Cincotti, A. Kübler, and R. Leeb. Combining brain-computer interfaces and assistive technologies: state-of-the-art and challenges. *Frontiers in Neuroscience*, 4:1–15, 2010. 59

[127] J.R. Millán. On the need for on-line learning in brain-computer interfaces. In *IEEE International Joint Conference on Neural Networks (IJCNN 2004)*, volume 4, pages 2877–2882, 2004. 36

[128] M. Moore Jackson and R. Mappus. Applications for brain-computer interfaces. *Brain-Computer Interfaces*, pages 89–103, 2010. 32

[129] S. T. Morgan, J. C. Hansen, and S. A. Hillyard. Selective attention to stimulus location modulates the steady-state visual evoked potential. *Proceedings of the National Academy of Sciences of the United States of America*, 93(10):4770, 1996. 28

[130] T. R. Mullen and Q. Huang. Brainwave-facilitated presenter feedback mechanism, November 30 2007. US Patent App. 11/947,908. 29

[131] T. R. Mullen and Q. Huang. Brain-wave aware sleep management, March 30 2010. US Patent 7,689,274. 29

[132] K. R. Müller, M. Krauledat, G. Dornhege, G. Curio, and B. Blankertz. Machine learning techniques for brain-computer interfaces. *Biomedical Engineering-Biomedizinische Technik*, 49(1):11–22, 2004. 46, 93, 94, 108

[133] M. M. Müller, T. W. Picton, P. Valdes-Sosa, J. Riera, W. A. Teder-Sälejärvi, and S. A. Hillyard. Effects of spatial selective attention on the steady-state visual evoked potential in the 20-28 Hz range. *Cognitive Brain Research*, 6(4):249–261, 1998. 28

[134] T. Müller, T. Ball, R. Kristeva-Feige, T. Mergner, and J. Timmer. Selecting relevant electrode positions for classification tasks based on the electro-encephalogram. *Medical and Biological Engineering and Computing*, 38(1):62–67, 2000. 53

[135] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg. Designing optimal spatial filters for single-trial EEG classification in a movement task. *Clinical Neurophysiology*, 110(5):787–798, 1999. 7, 34, 41, 43, 76, 82, 85, 116

[136] G. R. Müller-Putz, V. Kaiser, T. Solis-Escalante, and G. Pfurtscheller. Fast set-up asynchronous brain-switch based on detection of foot motor imagery in 1-channel EEG. *Medical and Biological Engineering and Computing*, 48(3):229–233, 2010. 40

[137] M. Naeem, C. Brunner, R. Leeb, B. Graimann, and G. Pfurtscheller. Seperability of four-class motor imagery data using independent components analysis. *Journal of Neural Engineering*, 3:208, 2006. 36

[138] M. Naeem, C. Brunner, and G. Pfurtscheller. Dimensionality reduction and channel selection of motor imagery electroencephalographic data. *Computational Intelligence and Neuroscience*, 2009:1–8, 2009. 24, 33, 38, 77, 80

[139] N. J. D. Nagelkerke. A note on a general definition of the coefficient of determination. *Biometrika*, 78(3):691–692, 1991. 71, 72

[140] A. Nijholt, D. P. O. Bos, and B. Reuderink. Turning shortcomings into challenges: Brain-computer interfaces for games. *Entertainment Computing*, 1(2):85–94, 2009. 19

[141] A. Nijholt and D. Tan. Brain-computer interfacing for intelligent systems. *IEEE Intelligent Systems*, 23(3):72–79, 2008. 33

[142] A. W. North. Accuracy and precision of electro-oculographic recording. *Investigative Ophthalmology & Visual Science*, 4(3):343, 1965. 64, 65

[143] Q. Novi, C. Guan, T. H. Dat, and P. Xue. Sub-band common spatial pattern (SBCSP) for brain-computer interface. In *3rd International IEEE/EMBS Conference on Neural Engineering (CNE'07)*, pages 204–207, 2007. 34, 48, 49

[144] P. L. Nunez, R. Srinivasan, A. F. Westdorp, R. S. Wijesinghe, D. M. Tucker, R. B. Silberstein, and P. J. Cadusch. EEG coherency I: statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales. *Electroencephalography and Clinical Neurophysiology*, 103(5):499–515, 1997. 16, 114

[145] B. Obermaier, C. Guger, C. Neuper, and G. Pfurtscheller. Hidden Markov models for online classification of single trial eeg data. *Pattern recognition letters*, 22(12):1299–1309, 2001. 45

[146] A. Osman and A. Robert. Time-course of cortical activation during overt and imagined movements. volume 1, pages 1842–1852, 2001. 50

[147] J. W. Osselton. Acquisition of EEG data by bipolar unipolar and average reference methods: a theoretical comparison. *Electroencephalography and Clinical Neurophysiology*, 19(5):527–528, 1965. 110

[148] S. Palva and J. M. Palva. New vistas for α-frequency band oscillations. *Trends in Neurosciences*, 30(4):150 – 158, 2007. 60, 62

[149] S. Parini, L. Maggi, A. C. Turconi, and G. Andreoni. A robust and self-paced BCI system based on a four class SSVEP paradigm: algorithms and protocols for a high-transfer-rate direct brain communication. *Computational Intelligence and Neuroscience*, 2009:1–11, 2009. 26

[150] M. J. Peters and H. J. Wieringa. The influence of the volume conductor on electric source estimation. *Brain Topography*, 5(4):337–345, 1993. 15, 76, 113

[151] G. Pfurtscheller, B. Z. Allison, C. Brunner, G. Bauernfeind, T. Solis-Escalante, R. Scherer, T. O. Zander, G. Mueller-Putz, C. Neuper, and N. Birbaumer. The hybrid BCI. *Frontiers in Neuroscience*, 2(3):1–12, 2010. 25, 59, 60

[152] G. Pfurtscheller, C. Brunner, A. Schlögl, and F. H. Lopes da Silva. Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks. *NeuroImage*, 31(1):153–159, 2006. 16, 29, 30, 42, 80, 106, 112, 114, 130, 141, 142

[153] G. Pfurtscheller, R. Leeb, C. Keinrath, D. Friedman, C. Neuper, C. Guger, and M. Slater. Walking from thought. *Brain Research*, 1071(1):145–152, 2006. 33

[154] G. Pfurtscheller and F. H. Lopes da Silva. Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical Neurophysiology*, 110(11):1842–1857, 1999. 29, 33, 38, 39, 43, 76, 112

[155] G. Pfurtscheller and C. Neuper. Future prospects of ERD/ERS in the context of brain–computer interface (BCI) developments. *Progress in Brain Research*, 159:433–437, 2006. 20

[156] G. Pfurtscheller, C. Neuper, D. Flotzinger, and M. Pregenzer. EEG-based discrimination between imagination of right and left hand movement. *Electroencephalography and Clinical Neurophysiology*, 103(6):642–651, 1997. 15, 16, 110, 113, 114, 121

[157] G. Pfurtscheller, T. Solis-Escalante, R. Ortner, P. Linortner, and GR Müller-Putz. 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, 2010. 26

[158] T. T. H. Pham, R. J. Croft, and P. J. Cadusch. Is ocular voltage propagation to the electroencephalogram frequency dependent? *Psychophysiology*, 46(5):949–956, 2009. 66

[159] T. T. H. Pham, R. J. Croft, P. J. Cadusch, and R. J. Barry. A test of four EOG correction methods using an improved validation technique. *International Journal of Psychophysiology*, 79(2):203–210, 2010. 62

[160] T. W. Picton. The P300 wave of the human event-related potential. *Journal of Clinical Neurophysiology*, 9(4):456, 1992. 20, 27

[161] J. A. Pineda. The functional significance of mu rhythms: translating "seeing" and "hearing" into "doing". *Brain Research Reviews*, 50(1):57–68, 2005. 20

[162] J. Polich and C. Margala. P300 and probability: comparison of oddball and single-stimulus paradigms. *International Journal of Psychophysiology*, 25(2):169–176, 1997. 27

[163] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller. Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Transactions on Rehabilitation Engineering*, 8(4):441–446, 2000. 34

[164] S. J. Raudys and A. K. Jain. Small sample size effects in statistical pattern recognition: Recommendations for practitioners. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(3):252–264, 1991. 46, 143

[165] D. Ravden and J. Polich. On P300 measurement stability: habituation, intra-trial block variation, and ultradian rhythms. *Biological Psychology*, 51(1):59–76, 1999. 28

[166] G. Rebolledo-Mendez, I. Dunwell, E. Martínez-Mirón, M. Vargas-Cerdán, S. De Freitas, F. Liarokapis, and A. García-Gaona. Assessing Neurosky's usability to detect attention levels in an assessment exercise. *Human-Computer Interaction, New Trends*, pages 149–158, 2009. 30

[167] D. Regan. Steady-state evoked potentials. *JOSA*, 67(11):1475–1489, 1977. 28

[168] John A Rice. *Mathematical statistics and data analysis*. Duxbury press, 2007. 159, 160

[169] P. J. Rousseeuw and K. Van Driessen. A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, pages 212–223, 1999. 50

[170] M. Sande Lemos and B. J. Fisch. The weighted average reference montage. *Electroencephalography and Clinical Neurophysiology*, 79(5):361–370, 1991. 34

[171] G. Schalk, J. Kubanek, K. J. Miller, N. R. Anderson, E. C. Leuthardt, J. G. Ojemann, D. Limbrick, D. Moran, L. A. Gerhardt, and J. R. Wolpaw. Decoding two-dimensional movement trajectories using electrocorticographic signals in humans. *Journal of Neural Engineering*, 4:264, 2007. 24

[172] R. Scherer, F. Lee, A. Schlögl, R. Leeb, H. Bischof, and G. Pfurtscheller. Toward self-paced brain–computer communication: navigation through virtual worlds. *IEEE Transactions on Biomedical Engineering*, 55(2):675–682, 2008. 26, 29

[173] R. Scherer, G. R. Müller-Putz, and G. Pfurtscheller. Self-initiation of EEG-based brain–computer communication using the heart rate response. *Journal of Neural Engineering*, 4:L23–L29, 2007. 25, 26, 59

[174] R. Scherer, A. Schlögl, F. Lee, H. Bischof, J. Janša, and G. Pfurtscheller. The self-paced Graz brain-computer interface: methods and applications. *Computational Intelligence and Neuroscience*, 2007:9–9, 2007. 20, 26

[175] A. Schlögl. *The electroencephalogram and the adaptive autoregressive model: theory and applications*. Shaker Germany, 2000. 45

[176] A. Schlögl and C. Brunner. BioSig: a free and open source software library for BCI research. *Computer*, 41(10):44–50, 2008. 43, 83, 100, 106, 120, 134, 141, 142

[177] A. Schlögl, C. Keinrath, D. Zimmermann, R. Scherer, R. Leeb, and G. Pfurtscheller. A fully automated correction method of EOG artifacts in EEG recordings. *Clinical Neurophysiology*, 118(1):98–104, 2007. 42, 43, 44, 60

[178] A. Schlögl, J. Kronegg, J. E. Huggins, and S. G. Mason. *Toward brain-computer interfacing*, chapter Evaluation Criteria for BCI Research, pages 327–360. MIT Press, 2007. 37, 121, 123, 139

[179] A Schlögl and G Pfurtscheller. Dataset IIIa: 4-class EEG data, 2005. 136, 137, 138

[180] M. Schröder, T. N. Lal, T. Hinterberger, M. Bogdan, N. J. Hill, N. Birbaumer, W. Rosenstiel, and B. Schölkopf. Robust EEG channel selection across subjects for brain-computer interfaces. *EURASIP Journal on Applied Signal Processing*, 2005(19):3103–3112, 2005. 55, 56, 88

[181] S. J. Segalowitz and K. L. Barnes. The reliability of ERP components in the auditory oddball paradigm. *Psychophysiology*, 30(5):451–459, 1993. 28

[182] E. W. Sellers, D.J. Krusienski, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw. A P300 event-related potential brain-computer interface (BCI): the effects of matrix size and inter stimulus interval on performance. *Biological Psychology*, 73(3):242–252, 2006. 27, 28

[183] E. W. Sellers, A. Kubler, and E. Donchin. 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, 2006. 28, 74

[184] S. Senn and W. Richardson. The first t-test. *Statistics in Medicine*, 13(8):785–803, 1994. 160

[185] F. Sharbrough, G. E. Chatrian, R. P. Lesser, H. Lüders, M. Nuwer, and T. W. Picton. American electroencephalographic society guidelines for standard electrode position nomenclature. *Electroencephalography and Clinical Neurophysiology*, 8:200–202, 1991. 110, 111

[186] J. C. Shaw. Correlation and coherence analysis of the EEG: a selective tutorial review. *International Journal of Psychophysiology*, 1(3):255–266, 1984. 43

[187] R. B. Silberstein, P. L. Nunez, A. Pipingas, P. Harris, and F. Danieli. Steady state visually evoked potential (SSVEP) topography in a graded working memory task. *International Journal of Psychophysiology*, 42(2):219–232, 2001. 30

[188] R. Sitaram, H. Zhang, C. Guan, M. Thulasidas, Y. Hoshi, A. Ishikawa, K. Shimizu, and N. Birbaumer. Temporal classification of multichannel near-infrared spectroscopy signals of motor imagery for developing a brain-computer interface. *NeuroImage*, 34(4):1416–1427, 2007. 1, 19, 25

[189] T. Solis-Escalante, G. Müller-Putz, and G. Pfurtscheller. Overt foot movement detection in one single Laplacian EEG derivation. *Journal of Neuroscience Methods*, 175(1):148–153, 2008. 40, 130

[190] W. Sommer and S. Schweinberger. Operant conditioning of P300. *Biological Psychology*, 33(1):37–49, 1992. 28

[191] W. A. Stephenson and F. A. Gibbs. A balanced non-cephalic reference electrode. *Electroencephalography and Clinical Neurophysiology*, 3(2):237–240, 1951. 38

[192] V. Surakka, M. Illi, and P. Isokoski. Gazing and frowning as a new human–computer interaction technique. *ACM Transactions on Applied Perception*, 1(1):40–56, 2004. 25

[193] M. Tangermann, M. Krauledat, K. Grzeska, M. Sagebaum, B. Blankertz, C. Vidaurre, and K. R. Müller. Playing pinball with non-invasive BCI. *Advances in Neural Information Processing Systems*, 21:1641–1648, 2009. 24

[194] M. Tangermann, K-R Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. J. Miller, and G. Mueller-Putz. Review of the BCI competition IV. *Frontiers in Neuroscience*, 6:55, 2012. 17, 86, 87, 88, 123, 126, 128, 138, 140

[195] D. M. J. Tax and R. P. W. Duin. Using two-class classifiers for multiclass classification. In *International Conference on Pattern Recognition*, volume 2, pages 124–127. IEEE, 2002. 130, 132

[196] M. Teplan. Fundamentals of EEG measurement. *Measurement Science Review*, 2(2):1–11, 2002. 38

[197] R. W. Thatcher, C. Biver, J. F. Gomez, D. North, R. Curtin, R. A. Walker, and A. Salazar. Estimation of the EEG power spectrum using MRI relaxation time in traumatic brain injury. *Clinical Neurophysiology*, 112(9):1729–1745, 2001. 39

[198] G. Tononi. An information integration theory of consciousness. *BMC Neuroscience*, 5(1):42, 2004. 154

[199] G. Tononi. Consciousness as integrated information: a provisional manifesto. *The Biological Bulletin*, 215(3):216–242, 2008. 154

[200] G. Tononi, M. Velmans, and S. Schneider. *The information integration theory of consciousness*. Blackwell: Malden, MA, USA, 2007. 154

[201] Y. Tsaig and D. L. Donoho. Breakdown of equivalence between the minimal L1-norm solution and the sparsest solution. *Signal Processing*, 86(3):533–548, 2006. 52

[202] D. Valbuena, M. Cyriacks, O. Friman, I. Volosyak, and A. Graser. Brain-computer interface for high-level control of rehabilitation robotic systems. In *10th IEEE International Conference on Rehabilitation Robotics (ICORR 2007)*, pages 619–625, 2008. 28

[203] A. Vallabhaneni, T. Wang, and B. He. Brain-computer interface. *Neural Engineering*, pages 85–121, 2005. 1, 19

[204] B. van de Laar, B. Reuderink, D. P. O. Bos, and D. Heylen. Evaluating user experience of actual and imagined movement in BCI gaming. *International Journal of Gaming and Computer-Mediated Simulations (IJGCMS)*, 2(4):33–47, 2010. 33

[205] J. G. van Dijk, J. F. V. Caekebeke, A. Jennekens-Schinkel, and A. H. Zwinderman. Background EEG reactivity in auditory event-related potentials. *Electroencephalography and Clinical Neurophysiology*, 83(1):44–51, 1992. 39

[206] J. van Erp, F. Lotte, and M. Tangermann. Brain-computer interfaces: Beyond medical applications. *Computer*, 45(4):26–34, 2012. 32

[207] M. van Gerven and O. Jensen. Attention modulations of posterior alpha as a control signal for two-dimensional brain-computer interfaces. *Journal of Neuroscience Methods*, 179(1):78–84, 2009. 60

[208] V. Vapnik. Pattern recognition using generalized portrait method. *Automation and Remote Control*, 24:774–780, 1963. 47

[209] M. N. Verbaten, J. N. R Beaujon, and W. Sjouw. EEG alpha rhythm, ocular activity and basal skin resistance. *Acta Psychologica*, 39(2):153–160, 1975. 69

[210] F. B. Vialatte, M. Maurice, J. Dauwels, and A. Cichocki. Steady-state visually evoked potentials: Focus on essential paradigms and future perspectives. *Progress in Neurobiology*, 90(4):418–438, 2010. 28

[211] J. J. Vidal. Toward direct brain-computer communication. *Annual Review of Biophysics and Bioengineering*, 2(1):157–180, 1973. 19

[212] J. J. Vidal. Real-time detection of brain events in EEG. *Proceedings of the IEEE*, 65(5):633–641, 1977. 39

[213] C. Vidaurre, N. Kramer, B. Blankertz, and A. Schlögl. Time domain parameters as a feature for EEG-based brain-computer interfaces. *Neural Networks*, 22(9):1313–1319, 2009. 15, 44, 45, 47, 93, 100, 112, 113

[214] C. Vidaurre, A. Schlögl, R. Cabeza, R. Scherer, and G. Pfurtscheller. A fully on-line adaptive BCI. *IEEE Transactions on Biomedical Engineering*, 53(6):1214–1219, 2006. 36

[215] R. N. Vigário. Extraction of ocular artefacts from EEG using independent component analysis. *Electroencephalography and Clinical Neurophysiology*, 103(3):395–404, 1997. 39, 60

[216] S. Vorobyov and A. Cichocki. Blind noise reduction for multisensory signals using ICA and subspace filtering, with application to EEG analysis. *Biological Cybernetics*, 86(4):293–303, 2002. 39

[217] M. Vourkas, S. Micheloyannis, and G. Papadourakis. Use of Ann and Hjorth parameters in mental-task discrimination. In *1st International Conference on Advances in Medical signal and Information Processing*, pages 327–332, 2000. 45

[218] T. Wang, J. Deng, and B. He. Classifying EEG-based motor imagery tasks by means of time-frequency synthesized spatial patterns. *Clinical Neurophysiology*, 115(12):2744–2753, 2004. 50

[219] T. Wang and B. He. An efficient rhythmic component expression and weighting synthesis strategy for classifying motor imagery EEG in a brain–computer interface. *Journal of Neural Engineering*, 1:1, 2004. 50

[220] Y. Wang, S. Gao, and X. Gao. Common spatial pattern method for channel selelction in motor imagery based brain-computer interface. In *27th IEEE Annual International Conference of the Engineering in Medicine and Biology Society (EMBS 2005)*, pages 5392–5395, 2006. 10, 34, 52, 53, 87, 92, 95, 108

[221] N. Weiskopf, F. Scharnowski, R. Veit, R. Goebel, N. Birbaumer, and K. Mathiak. Self-regulation of local brain activity using real-time functional magnetic resonance imaging (fMRI). *Journal of Physiology-Paris*, 98(4-6):357–373, 2004. 1, 19, 25

[222] P Welch. The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15(2):70–73, 1967. 143, 145

[223] J. Weston, A. Elisseeff, B. Schölkopf, and M. Tipping. Use of the zero norm with linear models and kernel methods. *The Journal of Machine Learning Research*, 3:1439–1461, 2003. 55

[224] J. Weston and C. Watkins. Support vector machines for multi-class pattern recognition. In *European Symposium on Artificial Neural Networks (ESANN)*, volume 4, pages 219–224, 1999. 132

[225] J. R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, and T. M. Vaughan. Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 113(6):767–791, 2002. 20, 24, 25, 28, 59, 60

[226] J. R. Wolpaw and D. J. McFarland. Multichannel EEG-based brain-computer communication. *Electroencephalography and Clinical Neurophysiology*, 90(6):444–449, 1994. 1, 19

[227] J. R. Wolpaw and D. J. McFarland. 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, 2004. 29

[228] J. R. Wolpaw, D. J. McFarland, G. W. Neat, and C. A. Forneris. An EEG-based brain-computer interface for cursor control. *Electroencephalography and Clinical Neurophysiology*, 78(3):252–259, 1991. 24, 29

[229] J. R. Wolpaw, D. J. McFarland, and T. M. Vaughan. Brain-computer interface research at the Wadsworth Center. *IEEE Transactions on Rehabilitation Engineering*, 8(2):222–226, 2000. 24

[230] K. F. K. Wong, A. Galka, O. Yamashita, and T. Ozaki. Modelling non-stationary variance in EEG time series by state space GARCH model. *Computers in Biology and Medicine*, 36(12):1327–1335, 2006. 46

[231] R. F. Woolson and W. R. Clarke. *Statistical methods for the analysis of biomedical data*. Wiley New York, 1987. 5, 79

[232] B. Wu, F. Yang, J. Zhang, Y. Wang, X. Zheng, and W. Chen. A frequency-temporal-spatial method for motor-related electroencephalography pattern recognition by comprehensive feature optimization. *Computers in Biology and Medicine*, 42:353–363, 2012. 44, 48, 110

[233] W. Wu, X. Gao, and S. Gao. One-versus-the-rest (OVR) algorithm: An extension of common spatial patterns (CSP) algorithm to multi-class case. In *27th Annual International Conference of the Engineering in Medicine and Biology Society (EMBS 2005)*, pages 2387–2390, 2006. 6, 79

[234] N. Yamawaki, C. Wilke, Z. Liu, and B. He. An enhanced time-frequency-spatial approach for motor imagery classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(2):250–254, 2006. 33, 50, 110

[235] Y. Yang, S. Chevallier, J. Wiart, and I. Bloch. Automatic selection of the number of spatial filters for motor-imagery BCI. In *20th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2012)*, pages 109–114, 2012. 18, 128

[236] Y. Yang, S. Chevallier, J. Wiart, and I. Bloch. Time-frequency selection in two bipolar channels for improving the classification of motor imagery EEG. In *34th IEEE Annual International Conference of Engineering in Medicine and Biology Society (EMBC 2012)*, pages 2744–2747, 2012. 92, 111, 114, 120, 123, 138

[237] Y. Yang, J. Wiart, and I. Bloch. Towards next generation human-computer interaction - brain-computer interfaces: applications and challenges. In *1st International Symposium of Chinese Computer Human Interaction (Chinese CHI 2013)*. 20

[238] X. Yong, R. K. Ward, and G. E. Birch. Robust common spatial patterns for EEG signal preprocessing. In *30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS 2008)*, pages 2087–2090, 2008. 50

[239] X. Yong, R. K. Ward, and G. E. Birch. Sparse spatial filter optimization for EEG channel reduction in brain-computer interface. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2008)*, pages 417–420, 2008. 34, 42, 52, 53

[240] B. Zhang, J. Wang, and T. Fuhlbrigge. A review of the commercial brain-computer interface technology from perspective of industrial robotics. In *IEEE International Conference on Automation and Logistics (ICAL)*, pages 379–384. IEEE, 2010. 29, 32