

ERROR-RELATED EEG POTENTIALS IN BRAIN-COMPUTER INTERFACES

THÈSE N° 3928 (2007)

PRÉSENTÉE LE 25 OCTOBRE 2007

À LA FACULTÉ DES SCIENCES ET TECHNIQUES DE L'INGÉNIEUR

Laboratoire de l'IDIAP

SECTION DE GÉNIE ÉLECTRIQUE ET ÉLECTRONIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Pierre FERREZ

ingénieur en sciences des matériaux diplômé EPF
de nationalité suisse et originaire de Bagnes (VS)

acceptée sur proposition du jury:

MER J. M. Sallese, président du jury
Prof. J. del R. Millán, directeur de thèse
Prof. F. Babiloni, rapporteur
Prof. A. Billard, rapporteur
Prof. K.-R. Müller, rapporteur



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Lausanne, EPFL

2007

*... à Albert et Emile,
mes grands-pères que je n'ai pas eu le bonheur de connaître.*

Acknowledgements

Acknowledgements are surely not the easiest part of a Ph.D. thesis to write, but they are definitely the most read part, since people usually inspect them to check if their name is mentioned or not... Therefore, I decided for your convenience to start this thesis with the Acknowledgements, so that the “inspection” is easier for you...

The cover of this thesis only shows my name. However, several persons contributed at different levels to its realization.

First of all, with an obvious and total lack of originality, I want to express my deepest gratitude to my Ph.D. supervisor, José Millán, who gave me the opportunity to discover the challenging world of “Brain-Computer Interfaces” and to work towards my Ph.D. degree so close to the best ski slopes in the universe... His extensive knowledge, his long experience and his numerous relations allowed me to conduct this research in excellent conditions. I further thank Aude Billard, Fabio Babiloni and Klaus-Robert Müller for accepting to be members of my Ph.D. thesis Jury.

Partial funding was provided by the Swiss National Science Foundation through the National Center of Competence in Research on “Interactive Multimodal Information Management” (IM2) and by the European Information Society Technologies (IST) Programme FET Project FP6-003758 on “Mental Augmentation through determination of Intended Action” (MAIA). I am grateful to both institutions for having made this research possible. I further thank all MAIA partners for their collaboration. In particular, I would like to thank Febo Cincotti at the IRCCS Fondazione Santa Lucia in Rome for his help with the CCD inverse model.

I am very thankful to many IDIAPers for their support, in particular my brother Jean-Albert, the BCIers and the System Admins. Discussions with Ferran Galán on the white board were maybe not always highly productive but at least always loud and animated... Tristan Carron was always ready to help debugging Windows computers or emptying beers at Marcel’s... Laurent Uldry has contributed to this thesis by accomplishing his diploma project under my supervision. Finally, I would like to thank all IDIAPers who kindly accepted to participate in my investigations.

A good working environment is mandatory if one hopes to find his own way to the Holly Grail that is a Ph.D. degree. Nevertheless, being able to “switch off my brain” (not too hard for a BCIer...) was also mandatory for my own sake. This notably includes crazy telemark sessions in Verbier and in the far North of Sweden, cheese eating contests in Bruson and memorable drinking sessions in company of Mrs. Guinness and Mr. Sinatra... I could mention many people here, but I am sure that the “regulars” all know who they are...

Last but not least, I wish to thank my entire extended family for providing me with a loving environment. In particular, I wish to thank my parents, Colette and Jean Ferrez. They raised me, taught me, supported me and loved me.

Finally, there are a few famous quotes I want to share with you. They somehow show the difficulty of explaining complex scientific concepts or achievements in simple terms and also how scientific advancement can be scary... I find them really funny and I hope they will at least make you smile. . .

Gravitation is not responsible for people falling in love.

Two things are infinite: the universe and human stupidity; and I'm not sure about the the universe.

If A is a success in life, then A equals x plus y plus z. Work is x; y is play; and z is keeping your mouth shut.

When you are courting a nice girl, an hour seems like a second. When you sit on a red-hot cinder, a second seems like an hour. That's relativity.

You see, wire telegraph is a kind of a very, very long cat. You pull his tail in New York and his head is meowing in Los Angeles. Do you understand this? And radio operates exactly the same way: you send signals here, they receive them there. The only difference is that there is no cat.

Albert Einstein

... comme vous me voyez, je suis incapable de reconnaître un rayon laser d'une corde à linge ordinaire, ou un chien qui pète d'un avion qui renifle. Et pourtant, la science, c'est pas de la merde, justement ! Le savant le savait bien, lui, que sans la science l'homme ne serait qu'un stupide animal sottement occupé à s'adonner aux vains plaisirs de l'amour dans les folles prairies de l'insouciance, alors que la science, et la science seule, a su lui apporter patiemment, au fil des siècles, le parcmètre automatique et l'horloge pointeuse sans lesquels il n'est pas de bonheur terrestre possible. C'est quand même grâce aux progrès fantastiques de la science que désormais nous savons que, quand on plonge un corps dans une baignoire, le téléphone sonne. C'est grâce aux progrès fantastiques de la science que désormais l'homme peut se rendre, en moins de trois heures, de Moscou à Varsovie. Et si y avait pas la science, si y avait pas la science, malheureux colportes, boursoufflés d'ingratitude aveugle et d'ignorance crasse, si y avait pas la science, combien d'entre nous pourraient profiter de leur cancer pendant plus de cinq ans ? Et n'est-ce pas le triomphe absolu de la science que d'avoir permis qu'aujourd'hui, sur la seule décision d'un vieillard californien impuissant, ou d'un fossile ukrainien encore plus gâteux que l'autre, l'homme puisse en une seconde faire sauter quarante fois sa planète, sans bouger les oreilles ! C'est pas moi qui le dis, c'est Fucius, croyez-moi, il avait oublié d'être con. Fucius disait : "Une civilisation sans la science, c'est aussi absurde qu'un poisson sans bicyclette."

Pierre Desproges

Résumé

Les personnes souffrant de lourds handicaps physiques tels que les lésions de la moelle épinière ou la sclérose latérale amyotrophique, mais ayant conservé des fonctions cérébrales intactes, sont en quelque sorte prisonnières de leur propre corps. Elles ont besoin d'autres moyens de communication et de contrôle pour interagir avec leur entourage dans la vie de tous les jours. Ces alternatives sont censées améliorer la qualité de vie de ces personnes en leur donnant l'opportunité de retrouver une partie de leur indépendance. Ces solutions se doivent donc d'être fiables et ergonomiques afin de pouvoir être utilisées avec succès par des personnes à motricité fortement réduite. Durant les deux dernières décennies, de nombreuses études ont proposé l'activité électroencéphalographique (EEG) comme base pour une interaction directe entre le cerveau et un ordinateur. Les interfaces cerveau-ordinateur (BCIs) basées sur l'EEG procurent de nouveaux moyens de contrôle et de communication aux personnes handicapées et sont des alternatives prometteuses aux méthodes invasives. Cependant, comme tout autre mode d'interaction basé sur des données physiologiques, (activité musculaire, parole, expression gestuelle, etc.), les BCIs sont sujets aux erreurs lors de la reconnaissance de l'intention du sujet, et ces erreurs peuvent être fréquentes. En effet, même les sujets bien entraînés atteignent rarement 100% de succès. Cependant, une caractéristique unique de l'activité cérébrale réside dans le fait qu'elle contient non seulement des informations à partir desquelles il est possible de dériver des commandes de contrôle pour diriger un système, mais également des informations relatives à l'état mental du sujet qui sont essentielles à une interaction fructueuse, et tout ceci avec des délais de l'ordre de la milliseconde.

Un de ces états, qui a été proposé par différents groupes comme outil pour améliorer les performances des BCIs, est la conscience de l'occurrence d'une erreur. Cependant, ces groupes proposent d'utiliser des potentiels d'erreur (ErrP) consécutifs à des erreurs commises par le sujet lui-même. Cette thèse décrit un nouveau type d'ErrP, nommé *interaction ErrP*, qui sont présents dans l'EEG directement après une erreur de l'interface et non plus du sujet lui-même. De plus, ces ErrP sont détectés de manière satisfaisante au niveau de chaque décision. En effet, les taux de classification des réponses correctes et erronées sont en moyenne de 80%. Il devient dès lors possible d'introduire une sorte de procédure de vérification automatique dans le BCI : après la transformation de l'intention du sujet en une commande de contrôle, le BCI présente cette commande mais ne la transmettra pas si des ErrP sont générés. Les résultats expérimentaux présentés dans cette thèse confirment l'amélioration des performances du BCI avec ce nouveau protocole. De plus, cette procédure s'avère tout particulièrement bénéfique pour les utilisateurs débutants qui n'atteignent que très rarement de bonnes performances. En effet, la suppression de tout ou partie des commandes erronées augmente la confiance de l'utilisateur et accélère ainsi son apprentissage du contrôle du BCI.

Le second point exploré dans cette thèse est le problème de l'intégration de la détection des ErrP dans un BCI. En effet, le fait de présenter visuellement l'intention du sujet telle

que reconnue par le BCI avant d'éventuellement la transmettre au système contrôlé est synonyme d'informations supplémentaires à traiter pour le sujet et peut considérablement ralentir l'interaction puisque l'introduction d'un filtrage automatique des erreurs interfère fortement avec le BCI. Cependant, cette étude montre la faisabilité de *simultanément* détecter les réponses incorrectes du système et classer des tâches d'imagination motrice pour le contrôle d'un système, et ce de manière satisfaisante. L'intégration de cette procédure automatique de détection des erreurs améliore grandement les performances du BCI.

Un autre aspect de cette thèse est l'étude du bénéfice potentiel apporté par l'exploitation de connaissances neurophysiologiques pour améliorer la classification des ErrP, ainsi que plus généralement les performances du BCI. Des études récentes ont montré que les ErrP sont très probablement générés dans une aire cérébrale nommée cortex cingulaire antérieur (ACC). Cette hypothèse est vérifiée à l'aide du modèle inverse *sLORETA*. En effet, la localisation obtenue pour les ErrP montre de nets focus d'activité dans l'ACC ainsi que dans l'aire prémotrice supplémentaire (pre-SMA). Les résultats de localisation obtenus avec le modèle *cortical current density* (CCD) sont très similaires et de plus ce modèle surpasse les EEG pour la classification des ErrP. Grâce à sa stabilité, ce modèle est susceptible d'être utilisé avec succès dans le domaine des BCIs. Le modèle *ELECTRA* est également testé, mais les résultats de classification et de localisation sont dans ce cas moins encourageants.

Plus généralement, cette étude suggère qu'il est possible de détecter dans l'EEG des états cognitifs et émotionnels de haut niveau (par opposition, et en addition aux commandes motrices) comme l'urgence, la fatigue, la frustration, la confusion ou l'attention, qui sont cruciales pour une interaction fructueuse. En effet, la détection rapide de ces états va conduire à de véritables interfaces adaptatives qui vont évoluer en temps réel en fonction des changements de l'état cognitif et émotionnel/affectif de l'utilisateur.

Mots-clés

Interface cerveau-ordinateur (BCI), Electroencéphalogramme (EEG), Potentiels d'erreur (ErrP), Sélection d'éléments, Classification, Bit rate, Modèles inverses, Cortex cingulaire antérieur (ACC), Aire prémotrice supplémentaire (pre-SMA).

Abstract

People with severe motor disabilities (spinal cord injury (SCI), amyotrophic lateral sclerosis (ALS), etc.) but with intact brain functions are somehow prisoners of their own body. They need alternative ways of communication and control to interact with their environment in their everyday life. These new tools are supposed to increase their quality of life by giving these people the opportunity to recover part of their independence. Therefore, these alternative ways have to be reliable and ergonomic to be successfully used by disabled people. Over the past two decades, numerous studies proposed electroencephalogram (EEG) activity for direct brain-computer interaction. EEG-based brain-computer interfaces (BCIs) provide disabled people with new tools for control and communication and are promising alternatives to invasive methods. However, as any other interaction modality based on physiological signals and body channels (muscular activity, speech and gestures, etc.), BCIs are prone to errors in the recognition of subject's intent, and those errors can be frequent. Indeed, even well-trained subjects rarely reach 100% of success. In contrast to other interaction modalities, a unique feature of the *brain channel* is that it conveys both information from which we can derive mental control commands to operate a brain-actuated device as well as information about cognitive states that are crucial for a purposeful interaction, all this on the millisecond range.

One of these states is the awareness of erroneous responses, which a number of groups have recently proposed as a way to improve the performance of BCIs. However, most of these studies propose the use of error-related potentials (ErrP) following an error made by the subject himself. This thesis first describes a new kind of ErrP, the so-called *interaction ErrP*, that are present in the ongoing EEG following an error of the interface and no longer errors of the subject himself. More importantly, these ErrP are satisfactorily detected no more in grand averages but at the level of single trials. Indeed, the classification rates of both error and correct single trials based on error-potentials detection are on average 80%. At this level it becomes possible to introduce a kind of automatic verification procedure in the BCI: after translating the subject's intention into a control command, the BCI provides a feedback of that command, but will not transfer it to the device if ErrP follow the feedback. Experimental results presented in this thesis confirm that this new protocol greatly increases the reliability of the BCI. Furthermore, this tool turns out to be of great benefit especially for beginners who normally reach moderate performances. Indeed, filtering out wrong responses increases the user's confidence in the interface and thus accelerates mastering the control of the brain-actuated device.

The second issue explored in this thesis is the practical integration of ErrP detection in a BCI. Indeed, providing a first feedback of the subject's intent, as recognized by the BCI, before eventually sending the command to the controlled device, induces additional information to process by the subject and may considerably slow down the interaction since the introduction of an automatic response rejection strongly interferes with the BCI. However, this study shows the

feasibility of *simultaneously* and satisfactorily detecting erroneous responses of the interface and classifying motor imagination for device control at the level of single trials. The integration of an automatic error detection procedure leads to great improvements of the BCI performance.

Another aspect of this thesis is to investigate the potential benefit of using neurocognitive knowledge to increase the classification rate of ErrP, and more generally the performance of the BCI. Recent findings have uncovered that ErrP are most probably generated in a deep fronto-central brain area called anterior cingulate cortex (ACC). This hypothesis is verified using a well-known inverse model called *sLORETA*. Indeed, the localization provided for ErrP shows clear foci of activity both in the ACC and the pre-supplementary motor area (pre-SMA). The localization results using the *cortical current density (CCD)* model are very similar and more importantly, this model outperforms EEG for ErrP classification. Thanks to its stability, this model is likely to be successfully used in a BCI framework. The *ELECTRA* model for estimating local field potentials is also tested, but classification and localization results using this method are not so encouraging.

More generally, the work described here suggests that it could be possible to recognize in real time high-level cognitive and emotional states from EEG (as opposed, and in addition, to motor commands) such as alarm, fatigue, frustration, confusion, or attention that are crucial for an effective and purposeful interaction. Indeed, the rapid recognition of these states will lead to truly adaptive interfaces that customize dynamically in response to changes of the cognitive and emotional/affective states of the user.

Keywords

Brain-computer interface (BCI), Electroencephalogram (EEG), Error-related potentials (ErrP), Feature selection, Single-trial classification, Bit rate, Inverse models, Anterior cingulate cortex (ACC), pre-supplementary motor area (pre-SMA).

Contents

1	Introduction	1
1.1	Motivation and aim	1
1.2	Structure	4
2	Background and state of the art	7
2.1	The human brain	7
2.1.1	The cerebral cortex	9
2.2	Measuring brain activity	11
2.2.1	Electroencephalogram (EEG)	13
2.2.1.1	Frequency range	14
2.2.1.2	Artifacts	15
2.3	Event-related potentials (ERP)	15
2.3.1	Error-related potentials (ErrP)	16
2.3.1.1	Choice reaction task	16
2.3.1.2	Reinforcement learning tasks	17
2.3.1.3	Observation tasks	17
2.3.1.4	Source localization	17
2.4	Estimation of intracranial activity from EEG	17
2.4.1	The inverse problem: A general approach	18
2.4.2	Distributed linear inverse estimation	18
2.4.3	Inverse solutions	19
2.4.3.1	CCD inverse model	19
2.4.3.2	sLORETA inverse model	21
2.4.3.3	ELECTRA inverse model	22
2.5	Brain-computer interfaces (BCIs)	23
2.5.1	Visual evoked potentials	24
2.5.2	P300 evoked potentials	24
2.5.3	Slow cortical potentials	24
2.5.4	Mu and beta rhythms	24
2.6	The IDIAP BCI	25
2.6.1	Spontaneous EEG and asynchronous operation	25
2.6.2	The machine learning way to BCI	26
2.6.3	Shared control	26
2.7	Data acquisition and processing	27
2.7.1	Hardware	27
2.7.2	Signal processing	28

2.7.3	Feature selection	28
2.7.3.1	Discriminant power (DP) algorithm	29
2.7.3.2	Modified DP algorithm	31
2.7.4	Statistical classifier	34
2.7.5	Measuring performance: The bit rate	35
2.8	The MAIA project	37
3	Preliminary results	39
3.1	Choice reaction task	39
3.1.1	Grand averages	40
3.1.2	Source localization	41
3.1.3	Classification	41
3.2	Reinforcement learning task	42
3.2.1	Grand averages	42
3.2.2	Source localization	43
3.2.3	Classification	43
4	First attempt: Progress bars	45
4.1	Experimental setup	45
4.2	Grand averages	46
4.3	Estimation of intracranial activity	48
4.4	Single-trial classification	49
4.5	Bit rate improvement	50
4.6	ErrP and Oddball N200 and P300	51
4.7	Ocular artifacts	51
4.8	Conclusion	52
5	Second protocol: Moving cursor	55
5.1	Experimental setup	55
5.2	Grand averages	56
5.3	Electrodes selection	60
5.4	Estimation of intracranial activity	61
5.5	Single trial classification	63
5.6	ErrP and Oddball N200 and P300	64
5.7	Ocular artifacts	64
5.8	Conclusion	65
6	Mental commands and ErrP detection	67
6.1	Experimental setup	67
6.2	Mental tasks classification	68
6.3	Error-related potentials	71
6.4	Estimation of intracranial activity	72
6.5	Conclusion	75

7	Online validation	77
7.1	Experimental setup	77
7.2	Performances	79
7.3	Feature stability	81
7.3.1	Motor imagery	81
7.3.2	Interaction ErrP	81
7.4	Conclusion	85
8	Integration of inverse models	87
8.1	Experimental setup	87
8.2	Grand averages	88
8.3	Cortical current density (CCD) model	91
8.4	Estimated local field potentials (eLFP)	93
8.5	Single trial classification	94
8.6	Source localization	97
8.7	Conclusion	97
9	Conclusion	101
9.1	Future directions	102
A	Brodmann areas	105
	Bibliography	107

List of Tables

2.1	Confidence intervals.	32
4.1	Recognition rates with the first protocol.	50
4.2	Gain in performance with the first protocol.	50
5.1	Recognition rates with the second protocol.	63
6.1	Motor imagery classification.	70
6.2	ErrP classification and performance increase.	73
7.1	Classification rates and performance increase.	79
7.2	Performances in term of number of steps.	80
8.1	Recognition rates using EEG, CCD and eLFP, 10 subjects.	95
8.2	Recognition rates using EEG, CCD and eLFP, 5 subjects	96

List of Figures

1.1	Non-invasive brain-computer interface (BCI).	2
1.2	Concept of the use of ErrP in a BCI.	3
2.1	Left view of the human brain.	8
2.2	Coronal section of the brain.	9
2.3	Primary motor cortex.	10
2.4	Sagittal section of the brain.	11
2.5	The first electroencephalogram.	13
2.6	Different types of sinusoidal EEG activity	14
2.7	CCD inverse model brain views.	20
2.8	ELECTRA solution space	22
2.9	EEG electrodes positions.	27
2.10	Comparison of different feature selection methods.	30
2.11	The basic DP algorithm.	31
2.12	DP algorithm sensitivity to noise	32
2.13	The modified DP algorithm.	33
2.14	Information transfer rate.	36
2.15	Brain-actuated wheelchair.	38
3.1	Response ErrP.	40
3.2	Localization of response ErrP.	41
3.3	Feedback ErrP.	43
3.4	Localization of feedback ErrP.	44
4.1	First graphical interface: Progress bars.	45
4.2	Interaction ErrP, 4 subjects.	47
4.3	Interaction ErrP, average.	48
4.4	Localization of interaction ErrP.	49
4.5	Ocular artifacts.	52
5.1	New graphical interface: Moving cursor.	56
5.2	Interaction ErrP, 5 subjects.	57
5.3	Scalp potentials topographies.	58
5.4	FCz and its 8 neighbors.	59
5.5	Discriminant power of electrodes for ErrP classification.	61
5.6	Localization of interaction ErrP.	62
5.7	ErrP with an error rate of 50%.	65

5.8	Ocular artifacts.	66
6.1	Timing of the protocol.	68
6.2	Discriminant power of frequencies.	69
6.3	Discriminant power of electrodes.	70
6.4	Interaction ErrP, 6 subjects.	71
6.5	Interaction ErrP, average.	72
6.6	Localization of interaction ErrP.	74
7.1	Timing of the online protocol.	78
7.2	Stability of features for the first subject.	82
7.3	Stability of features for the second subject.	83
7.4	Stability of features for ErrP detection.	84
8.1	Interaction ErrP, 10 subjects.	89
8.2	Scalp potentials topographies.	90
8.3	Most relevant solution points for the CCD, 10 subjects.	91
8.4	Most relevant solution points for the CCD, 5 subjects.	92
8.5	Most relevant solution points for eLFP, 10 subjects.	93
8.6	Most relevant solution points for eLFP, 5 subjects.	94
8.7	Average recognition rates, 10 subjects.	95
8.8	Average recognition rates, 5 subjects.	96
8.9	Source localization, first group	98
8.10	Source localization, second group	99

Chapter 1

Introduction

1.1 Motivation and aim

“To err is human, but to really foul things up you need a computer” This famous quote from Paul Ehrlich (German scientist who won the 1908 Nobel Prize in Physiology or Medicine, 1854-1915) may sound like a joke, but in a sense it can be seen as the starting point of this thesis. We all make mistakes, even when performing everyday’s tasks we are used to. In this case we usually quickly understand the problem and adapt our strategy to avoid repeating our errors, this is a learning process. When it comes to interactions with complex machines like computers, it’s sometimes much harder to isolate and understand quickly the problem and so we can get lost, since in some cases we may even not be responsible for the error. Many people work on making computers easier to use and more reliable, so that it’s becoming harder to “foul things up” using them. In the same idea, this thesis proposes a tool to improve performance and robustness of brain-computer interfaces (BCIs). However this tool could have great applications way beyond the specific field of BCIs.

Computers are basically machines following a list of instructions to manipulate data. Originally computers were the size of a large room, consuming as much power as several hundred of modern personal computers and mostly used by scientists for computation purposes. Nowadays, computers can be made small enough to fit into a wrist watch and be powered from a watch battery. With the emergence of communication-based technologies, personal computers and laptops as well as other devices like mobile phones have become major tools of the information age. Beside miniaturization, a crucial issue to transform original computers into user-friendly and entertaining devices is the ease of use. Today computers remain very complex machines, but it has become much easier to use them and to interact with them.

However, common human-computer interfaces presently available such as keyboards or mice directly depend on the activity of muscles and peripheral nerves. This dependence on movements make those traditional interfaces useless for people who are partially or totally paralyzed. People with severe motor disabilities (spinal cord injury (SCI), amyotrophic lateral sclerosis (ALS), etc.) need alternative ways of communication and control for their everyday life, not only to enjoy the use of a computer but also for basic tasks like switching the light on and off and more importantly for vital tasks like making an emergency phone call when needed.

The idea of driving complex devices such as robots no longer by physical control but only by *thinking* has been studied over the last few years. Invasive approaches involving intracranial electrodes implanted in the cortex have shown encouraging results with monkeys [Chapin

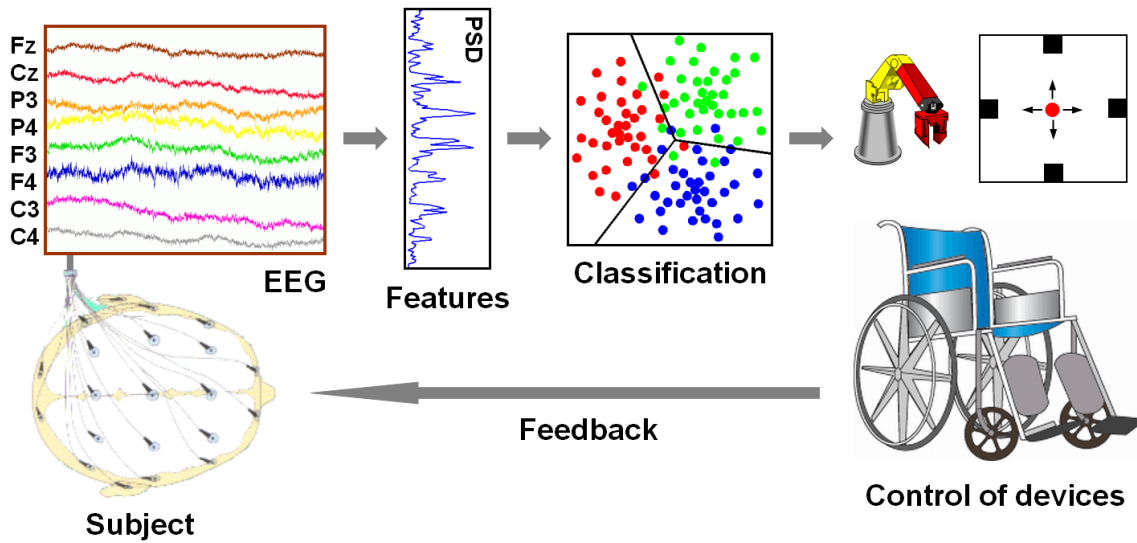


Figure 1.1: Non-invasive brain-computer interface (BCI).

EEG is recorded by means of electrodes placed on the scalp. Then some features are extracted from the EEG and sent to a classifier, whose response is translated into some action whose execution provides feedback to the user.

et al., 1999, Wessberg et al., 2000, Serruya et al., 2002, Taylor et al., 2002, Carmena et al., 2003, Nicolelis, 2003, Musallam et al., 2004]. However, non-invasive methods, such as those based on electroencephalogram (EEG), are necessary for humans because of ethical concerns and medical risks, even if EEG signals usually suffer from a reduced spatial resolution and decreased signal-to-noise ratio (SNR) due to measurements on the scalp [Pfurtscheller et al., 2000, Wolpaw et al., 2002, Millán, 2002, Millán et al., 2004a,b, Birbaumer et al., 1999].

A non-invasive brain-computer interface (BCI) is a system capable of translating the intention of a subject, as represented by his brain waves (EEG), into a control signal without using activity of any muscles or peripheral nerves. EEG is recorded by means of electrodes placed on the scalp. Then some features are extracted from the EEG and sent to a classifier, whose response is translated into some action whose execution provides feedback to the user, as shown in Figure 1.1. For instance, recent studies have shown that after a few days of training, subjects are able to control a miniature robot in an indoor environment with several rooms, corridors and doorways only using the signals derived from a EEG-based brain-computer interface [Millán et al., 2004a]. This same system was very recently used to drive a wheelchair so that a subject suffering from severe motor disabilities, but with intact brain capabilities, could greatly gain in autonomy [Philips et al., 2007]. Other applications, such as a virtual keyboard, could more generally provide an alternative way of communication with the outside world to people suffering from motor disabilities [Millán et al., 2004b, Scherer et al., 2004, Wolpaw et al., 2003].

Nevertheless, EEG-based communication systems suffer from the problem of errors in the recognition of subject's intent, and those errors can be frequent. In fact, even well-trained subjects rarely reach 100% of success. A possible way to reduce errors consists in a verification procedure whereby each output consists of two opposite trials, and success is required on both to validate the outcome [Wolpaw et al., 1998]. Even if this method greatly reduces the errors, it requires much more mental effort from the subject and reduces the communication rate.

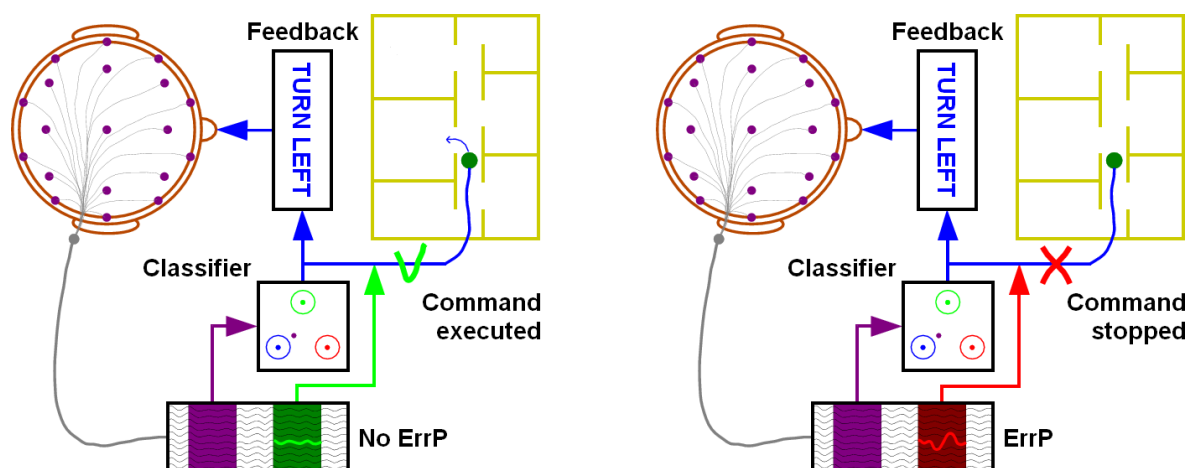


Figure 1.2: Concept of the use of ErrP in a BCI.

After translating the subject's intention into a control command, the BCI provides a feedback of that command. This command will be executed if no ErrP follow the feedback (left), but will be stopped if ErrP are present (right).

Another elegant approach recently introduced to improve the accuracy of EEG-based brain-computer interfaces consists in a verification procedure directly based on the subject's EEG signals [Schalk et al., 2000, Blankertz et al., 2002, 2003, Parra et al., 2003]. Since the late 1980s, different physiological studies have shown the presence of error-related potentials (ErrP) in the EEG recorded right after the occurrence of an error [Carter et al., 1998, Falkenstein et al., 2000, Vidal et al., 2000, Gehring and Fencsik, 2001, Coles et al., 2001, Rodriguez-Fornells et al., 2002, Ridderinkhof et al., 2003, Nieuwenhuis et al., 2001, Holroyd et al., 2003, Holroyd and Coles, 2002, Ullsperger and von Cramon, 2003, Gehring and Willoughby, 2002, Badgaiyan and Posner, 1998, Miltner et al., 1997, Luu et al., 2003]. Nevertheless, most of these studies show the presence of ErrP in typical choice reaction tasks, i.e. the subject is asked to respond as quickly as possible to a stimulus and ErrP arise following errors due to the subject's incorrect motor action (e.g., the subject pressed a key with the left hand when he should have responded with the right hand) [Carter et al., 1998, Nieuwenhuis et al., 2001]. More recently, other studies have also shown the presence of ErrP in typical reinforcement learning tasks where the subject is asked to make a choice and ErrP arise following the presentation of a stimulus that indicates incorrect performance (negative feedback) [Holroyd et al., 2003, Luu et al., 2003]. In both cases, (response ErrP and feedback ErrP), the anterior cingulate cortex (ACC) appears to be involved in the error-processing neural system leading to ErrP [Holroyd and Coles, 2002, Fiehler et al., 2004], although there exists disagreement on the exact role played by the ACC [Luu et al., 2003].

An important aspect of the described ErrP is that they always follow an error made by the subject himself. First the subject makes a selection, and then ErrP arise either simply after the occurrence of an error (choice reaction task) or after a feedback indicating the error (reinforcement learning task). Furthermore, in the case of a reinforcement learning task, the subject will try to adapt his strategy to minimize the occurrence of the negative feedback. However, in the context of a BCI, or human-computer interaction in general, the central question is:

Are ErrP also elicited when the error is made by the interface during the recognition of the subject's intent?

In order to consider the full implications of this question, let's imagine that the subject's intent is to make a robot reach a target to the left. What would happen if the interface fails to recognize the intended command and the robot starts turning in the wrong direction? Are ErrP still present even though the subject did not make any error but only perceives that the interface is performing wrongly?

The objective of this thesis is to investigate how ErrP could be used to improve the performance of a BCI. Thus, the first step is to explore whether or not ErrP also follow a feedback indicating incorrect responses of the interface and no longer errors of the subject himself. If ErrP are also elicited in this case, then we could integrate them in a BCI in the following way: after translating the subject's intention into a control command, the BCI provides a feedback of that command, but will not transfer it to the device if ErrP follow the feedback. Figure 1.2 illustrates this concept of the use of ErrP in a three-class BCI to control a mobile robot in an indoor environment [Millán et al., 2004a]. This new protocol should greatly increase the reliability of the BCI. Also, this tool should be of great benefit for beginners who normally reach moderate performances: filtering out wrong responses should increase the user's confidence in the interface and thus accelerate mastering the control of the brain-actuated device.

Of course, this protocol depends on the ability to detect ErrP no longer in averages of a large number of trials, but in each single trial using a short window following the feedback that shows the response of the classifier embedded in the BCI. This is the second step of the proposed thesis, namely the development of a robust classifier trained to differentiate correct and error trials. This, in turn, raises the issues of selecting the kind of feedback that best elicits ErrP and also exploring appropriate EEG preprocessing that enhances single-trial recognition of ErrP.

As already mentioned, a few studies propose the use of ErrP to improve the performances of a BCI [Schalk et al., 2000, Parra et al., 2003], but most of them [Blankertz et al., 2002, Parra et al., 2003] are in the framework of ErrP involving movements of the subject. These ErrP are found to be a potential tool, but due to their nature (i.e., the user responds to external stimuli) they cannot be integrated in the BCI. The study of Schalk et al. [Schalk et al., 2000] reports the presence of ErrP at the end of a complete trial (made of several single trials) in a bidirectional control of a cursor using a BCI when the cursor reaches the incorrect target. They did not try, however, to recognize automatically those errors. So to our knowledge, this thesis is the first attempt to detect ErrP after each decision of the system (each single trial) and to directly integrate them in the BCI.

1.2 Structure

This thesis is divided in two main parts. The first part contains a neurophysiological background and summarizes present BCI technology. It also describes the different algorithms and tools used in the thesis. The second part is devoted to experimental results, going from the first attempt to record error-related potentials to the online validation of their integration into the BCI and the attempt of using inverse solutions to increase recognition rates.

- Chapter 2** This chapter contains general information about the human brain and the different ways to record its activity, with a special focus on electroencephalogram (EEG). It also contains a description of event-related potentials (ERP) with a special focus on error-related potentials (ErrP). Then it provides an introduction on inverse solutions to estimate intracranial activity from scalp EEG and a description of the different kind of non-invasive brain-computer interfaces (BCIs). Finally, it contains a description of several algorithms used for data processing and a short introduction to the MAIA project.
- Chapter 3** This chapter presents preliminary results on error-related potentials. The main goal of these experiments was the validation of our EEG system and our data processing algorithms by replicating two different experiments available in the literature, one in a choice reaction task framework and one in a reinforcement learning task framework.
- Chapter 4** The first graphical interface that was used to record error-related potentials following an error made by the system in the recognition of the subject's intent is presented in this chapter. This chapter contains a description of the potentials, the single trial classification rates, the theoretical improvement of performance and clarifies issues such as ocular artifacts.
- Chapter 5** The protocol used in Chapter 4 was not optimal since it was very similar to an Oddball paradigm. The new protocol presented in this chapter avoids habituation to one of the stimuli and is also more realistic and engaging for the subject. This chapter roughly contains the same kind of analysis as Chapter 4.
- Chapter 6** In this chapter we report the feasibility of simultaneously and satisfactorily classifying motor imagination for the mental control of a brain actuated device and detecting erroneous responses of the interface to improve the BCI accuracy.
- Chapter 7** This chapter reports the online implementation of the off-line analysis made in Chapter 6. Two subjects participated in successful real time experiments where they were mentally controlling a cursor on a screen using motor imagery and the BCI was simultaneously detecting the presence of error-related potentials to filter out wrong decisions.
- Chapter 8** This chapter investigates the potential benefit of using neurocognitive knowledge to increase the classification rate of ErrP. We show that the CCD inverse model seems to be a very promising tool for BCIs.
- Chapter 9** In this final chapter we summarize and discuss the main results and achievements of this thesis and we introduce several possible future directions of investigation.

Chapter 2

Background and state of the art

This chapter contains a non-exhaustive background on the brain and its functions as well as on the different methods used to measure its activity, either neuronal or metabolic. Electroencephalography is described more precisely since electroencephalogram (EEG) is the most widely used input signal for non-invasive brain-computer interfaces (BCIs), as for example in the IDIAP BCI. This chapter also describes non-exhaustively event-related potentials (ERP) and in particular error-related potentials (ErrP) since the main goal of this work was the use of ErrP as a tool to improve the reliability of brain-computer interfaces. The second part of this chapter is dedicated to a theoretical framework of intracranial activity estimation from scalp EEG and its underlying concepts. Then this chapter describes different types of non-invasive BCIs according to the kind of signals they are based on, with an emphasis on the IDIAP BCI and its theoretical aspects. This thesis was conducted in the framework of the EU MAIA project so that this chapter finally contains a brief description of this project and its main achievements. Parts of the following descriptions are inspired by more complete information freely available on Wikipedia, the free encyclopedia¹.

2.1 The human brain

Anatomically the human brain is the central organ supervising the nervous system. The term *brain* often refers to the whole encephalon, i.e. the part of the central nervous system located in and protected by the skull, including for example the cerebellum. The brain controls and coordinates movements, behavior and homeostasis of internal functions such as heart beat, blood pressure and body temperature. It's located close to the primary sensory apparatus of vision, hearing, equilibrioception, taste and olfaction. The brain is a very complex organ, indeed the human brain contains more than 100 billion neurons, each linked to as many as 10,000 other neurons. The brain requires a lot of energy, about 20% of the oxygen caught by the lungs and 75% of the blood sugar produced by the liver is consumed by the brain. Neurons process information and are electrically active, this activity can be measured by different means as explained in Section 2.2. The brain consists of grey matter, the cell bodies of neurons, and of white matter, the fibers (axons) which connect neurons, the cerebral cortex is the grey matter outer layer of the brain. In humans and several other animals, the fissures (sulci) and convolutions (gyri) give the brain a wrinkled appearance. The brain is the site of intelligence and reason including

¹<http://en.wikipedia.org/>

cognition, perception, attention, memory and emotion. Control of posture and movements as well as learning (motor or cognitive) are examples of brain functions. The brain has a functional structure, different parts specifically control particular aspects of behavior and thought. This structure is not strict, complex functions such as memory cannot be assigned to a specific region of the brain. However it's possible to draw a map of the brain areas according to their cognitive functions, for example motor functions in the frontal part and vision in the posterior part. The brain has two hemispheres (left and right) separated by the medial longitudinal sulcus and the largest part of the human brain is the cerebrum. The cerebrum can be divided in four zones called lobes, as shown in Figure 2.1.

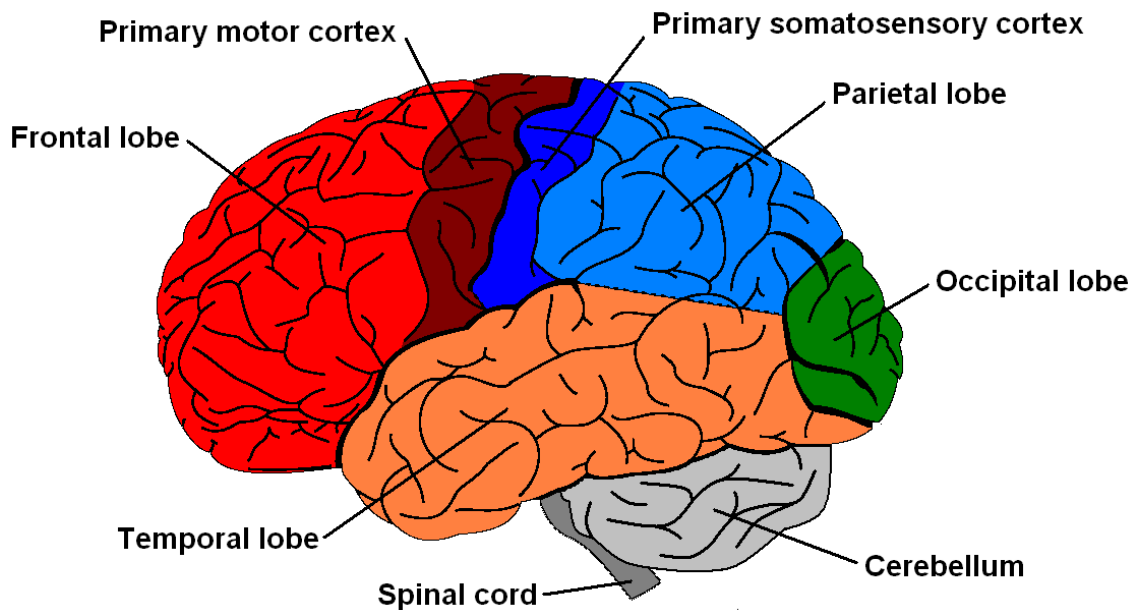


Figure 2.1: Left view of the human brain.

The cerebrum (main part of the brain) can be divided in four zones called lobes. The primary motor cortex is part of the frontal lobe and the primary somatosensory cortex is part of the parietal lobe.

- **Frontal lobes** are located at the front of each cerebral hemisphere. The frontal lobes are involved in various tasks like impulse control, judgment, language production, working memory, motor function, problem solving, sexual behavior, socialization, and spontaneity. The frontal lobes assist in planning, coordinating, controlling, and executing behavior.
- **Parietal lobes** are located posterior to the frontal lobes and above the occipital lobes. The parietal lobes play important roles in integrating sensory information from various parts of the body, and in the manipulation of objects. Portions of the parietal lobes are involved with visuospatial processing.
- **Temporal lobes** are located at the sides of the brain. The temporal lobes are responsible for auditory processing and are home to the primary auditory cortex. It is also heavily involved in semantics both in speech and vision. The temporal lobes contain the hippocampus that plays an essential role in the formation of new memories.

- **Occipital lobes** are the smallest lobes and are located in the rearmost portion of the brain. The occipital lobes are the visual processing center of the human brain, containing most of the anatomical region of the visual cortex.

2.1.1 The cerebral cortex

The human cerebral cortex is the folded grey tissue that covers the surface of each cerebral hemisphere as shown in Figure 2.2. It is responsible for language, music, calculation, imagining, thinking and planning. It controls our ability to move our limbs and any body part we can move deliberately. It determines our intelligence, emotions, and personality. It also processes sensory information for vision, hearing, and speech. Almost everything we do consciously depends on the cortex. The brain cortex can be divided in zones defined according to its cytoarchitecture (the arrangement of neuron in the cerebral cortex). These zones are called Brodmann areas (BA) and were originally defined in 1909 by German neurologist Korbinian Brodmann (1868-1918) and referred to by numbers from 1 to 52 [Brodmann, 1909]. Appendix A contains the list of the Brodmann areas for the human brain and their location in the cortex.

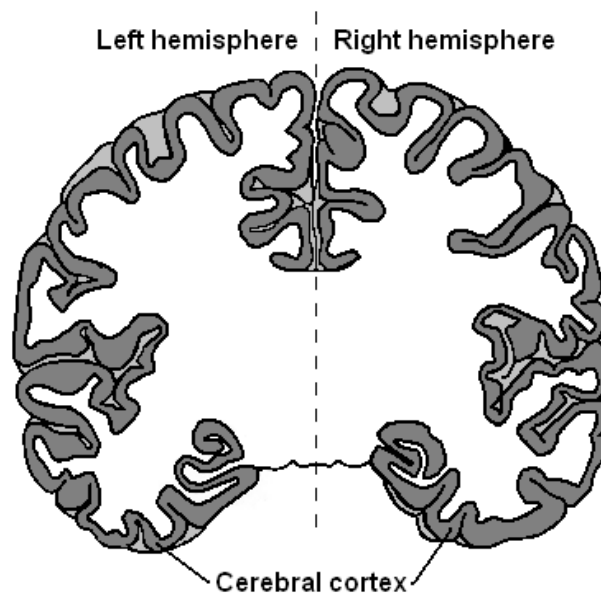


Figure 2.2: Coronal section of the brain.

The cerebral cortex is the outermost layer of the cerebrum and is 2-4 mm thick. It is covering the surface of each hemisphere and is composed of grey matter (cell bodies of neurons).

The motor cortex is a region of the cerebral cortex involved in the planning, control and execution of voluntary motor functions. The primary motor cortex (BA 4, cf. Figure 2.1, Figure 2.3 and Figure 2.4) is responsible for generating the neural impulses controlling execution of movements. The secondary motor cortex is responsible for transforming visual information into motor commands, for sensory guidance of movements and for planning and coordination of complex movements. Stimulation studies have shown that the activity of particular parts of the primary motor cortex causes movements of particular parts of the body. Figure 2.3 shows the localization of different parts of the body on the primary motor cortex according to the

observations of Penfield and Rasmussen [Penfield and Rasmussen, 1950]. Since these localizations are relatively precise and since motor areas are active during both movement preparation and movement imagination, motor-related tasks are widely used in non-invasive EEG-based brain-computer interfaces [Wolpaw et al., 2003, Pfurtscheller and Neuper, 2001, Kostov and Polak, 2000, Penny et al., 2000].

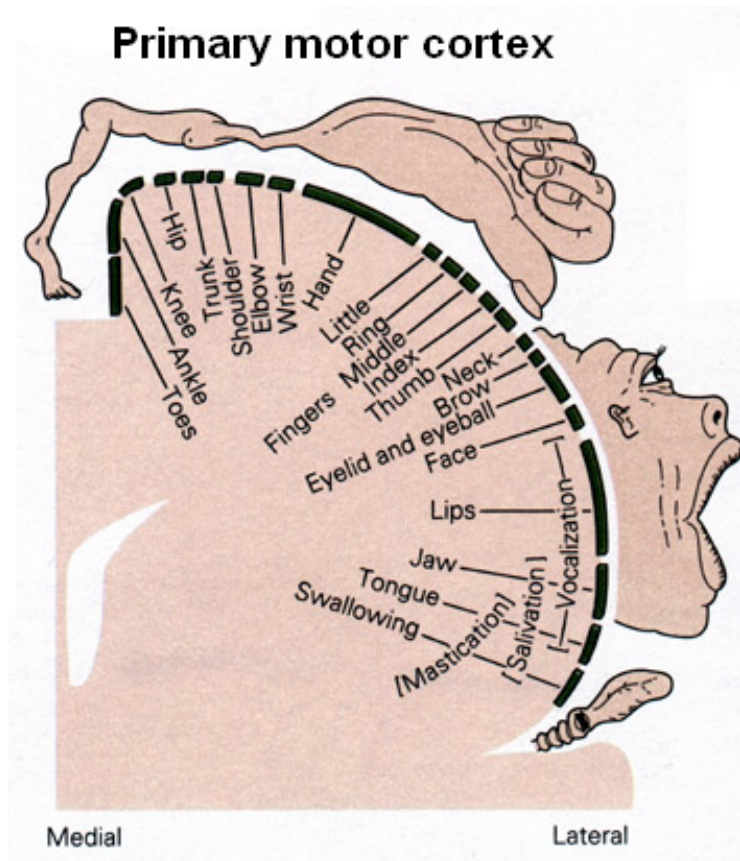


Figure 2.3: Primary motor cortex.

Localization of different parts of the body on the primary motor cortex according to the observations of Penfield and Rasmussen.

The pre-supplementary motor area (pre-SMA) is located above the ventral part of the anterior cingulate cortex (ACC) and is part of the motor cortex (cf. Figure 2.4). This area, part of the Brodmann area 6, is believed to play a role in the planning of complex, coordinated movements as well as in error processing [Holroyd and Coles, 2002, Fiehler et al., 2004, Hester et al., 2004].

The anterior cingulate cortex (ACC) is a deep cortical brain area as shown in Figure 2.4. It's located in both hemispheres around the corpus callosum, a white matter bundle that transmits neural signals between the left and right hemispheres. The anterior cingulate cortex includes the ventral (vACC, BA 24) and the dorsal (dACC, BA 24) areas of the cingulate cortex. Various functions have been ascribed to the ACC, including modulation of attention, monitoring competition, complex motor control, motivation, novelty, error detection and working memory. Several studies propose that the rostral anterior cingulate cortex (front) is related to emotions whereas the caudal anterior cingulate cortex (rear) is more related to cognition [Devinsky et al.,

1995, Carter et al., 1999, Bush et al., 2000]. Many studies also proposed the ACC as the brain's error processing center and thus error-related potentials (ErrP) are most probably generated in the ACC, although there exists disagreement on the exact role played by the ACC [Holroyd and Coles, 2002, Luu et al., 2003, Fiehler et al., 2004].

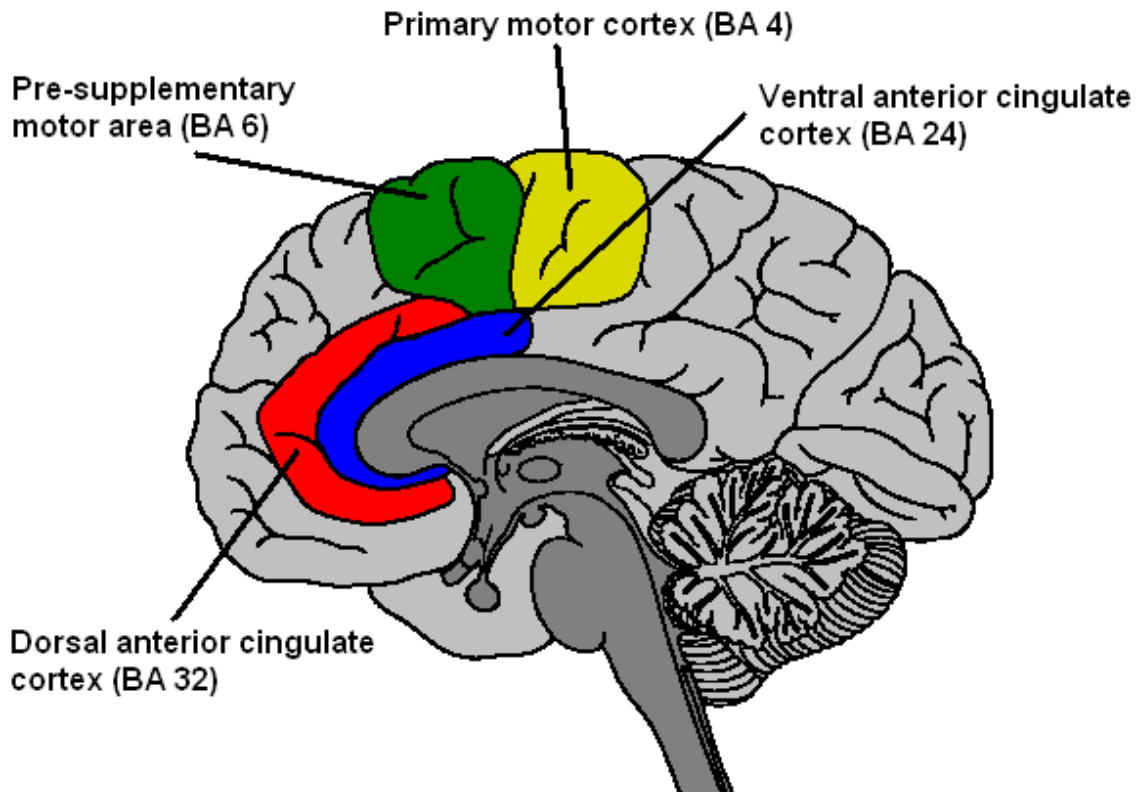


Figure 2.4: Sagittal section of the brain.

The anterior cingulate cortex (ACC) is a deep frontal area including the ventral and the dorsal ACC. The pre-supplementary motor area (pre-SMA) is fronto-central area part of the motor cortex and located anterior to the primary motor cortex.

2.2 Measuring brain activity

Brain activity measurement methods can be classified according to their invasiveness. The quality of the acquired signals usually increases with the invasiveness of the method since with invasive techniques the probe is closer to the source. Quality of non-invasive methods signals is poor since the skull acts as an attenuator of neural signals, thus filtering out high frequencies and lowering signal-to-noise ratio (SNR). Another way to classify the different available methods to measure the activity of the brain is the nature of the recording, namely neuronal or vascular (metabolic) activity. Neuronal activity can be measured in the range of milliseconds whereas the temporal resolution of vascular activity is much lower, it lies in the range of seconds. Several techniques are now briefly presented, starting with two invasive methods, namely single unit recordings and electrocorticogram (ECoG), followed by methods based on vascular activity (oxygen or sugar consumption), positron emission topography (PET), functional

magnetic response imaging (fMRI) and near infrared spectroscopy (NIRS). Finally two non-invasive methods based on neuronal activity will be presented, magnetoencephalogram (MEG) and electroencephalogram (EEG) with a particular focus on the latter.

- **Single unit recordings** The electrophysiological activity (action potentials) from a single or a reduced population of neurons can be recorded using this method. Electrodes (or micro-electrodes with a tip size of a few μm for single neuron recordings) are directly implanted in the cortex. Due to its high invasiveness, this technique is mainly used on animals (monkeys). This technique provides the best temporal and spatial resolution but beside the risk of the invasive approach, electrodes induce scars in the tissue so that quality of recordings decreases over time and neuronal tissue necrosis can follow electrode implantation.
- **Electrocorticogram (ECoG)** With this method, brain activity is measured by means of electrodes directly placed on the surface of the cortex, without penetrating into the brain. Very accurate electrical signals from neurons can be obtained by placing electrodes directly on the cortical grey matter. ECoG electrodes provide both good spatial and temporal resolution, but not as good as single unit recordings. This technique is widely used to localize areas of epileptic seizures and its main drawback is of course invasiveness.
- **Positron Emission Tomography (PET)** Three dimensional maps of functional processes in the brain can be obtained with this nuclear medical imaging technique. A short-lived radioactive tracer isotope is integrated into a metabolically active molecule, typically sugar, and is injected into the blood circulation. This radioactive isotope decays by emitting a positron, the antimatter counterpart of an electron. Therefore it is possible to measure metabolic activity of a brain area by detecting positron emission. Due to the relatively slow coupling between neuronal activity and metabolism (neurovascular coupling), the temporal resolution of this technique is usually low, typically lying in the range of seconds. This technique also presents a small risk for the subject since radioactive compounds are injected in the blood circulation.
- **Functional Magnetic Response Imaging (fMRI)** Red blood cells in local capillaries deliver oxygen to neurons by means of haemoglobin. There is an increased demand for oxygen when neuronal activity becomes more intense so that an increase of blood flow can be observed in the regions with higher neuronal activity. Oxygenated and deoxygenated haemoglobin have different magnetic properties and thus the blood has different magnetic responses according to its degree of oxygenation. Since blood oxygenation is correlated with neuronal activity, differences in magnetic response can be used to measure brain activity. The spatial resolution of fMRI is very high whereas due to the neurovascular coupling, its temporal resolution is usually low, typically lying in the range of seconds.
- **Near Infrared Spectroscopy (NIRS)** This method uses the interaction of the near infrared region of the electromagnetic field spectrum (from about 1000 nm to 2500 nm) with biological materials that show a relatively good transparency in this wavelength. Oxygenated and deoxygenated haemoglobin have different optical properties. As for fMRI, since blood oxygenation is correlated with neuronal activity, differences in optical

response can be used to measure brain activity. Due to the neurovascular coupling, this technique has a low temporal resolution and so far its spatial resolution is poor.

- **Magnetoencephalography (MEG)** Electrical currents generate a magnetic field so that the electrical (neuronal) activity of the brain can be observed by measuring the associated magnetic field emanating from the brain. This technique shows a very high temporal resolution, typically on the order of milliseconds. However MEG is technically very demanding since movements artifacts strongly contaminate recordings and thus huge equipments are required to obtain quality recordings.

2.2.1 Electroencephalogram (EEG)

Electroencephalography is the neurophysiological measurement of the electrical activity of the brain recorded by means of electrodes placed on the scalp. The resulting traces are known as electroencephalogram (EEG). EEG represent the electric signal emerging from a large population of neurons and reflects the synchronous activity of this population of neurons. A typical adult human EEG signal is about 20-100 μV when measured from the scalp. EEG is frequently used in experimentation because this technique is simple compared to other brain activity measurement techniques and the process is non-invasive to the subject. EEG is capable to detect changes in the brain electrical activity on the range of the millisecond. English physician Richard Caton already discovered the presence of electrical current in the brain in 1875, but it was not until 1924 that German neurologist Hans Berger made the first brain's electrical activity recording on graph paper as shown in Figure 2.5 [Berger, 1929]. Already at that time, Berger noticed that brain waves varied with the individual's state of consciousness.

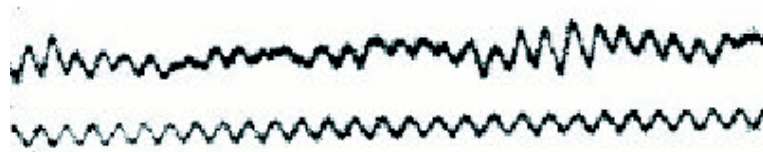


Figure 2.5: The first electroencephalogram.

It was recorded by Berger in ca. 1924 and already at that time he noticed that the brain waves varied with the individual's state of consciousness.

Scalp electrodes are not sensitive enough to pick up the activity of a single neuron. Instead, EEG electrodes pick up the mixed activity of a large population of neurons, which produces a greater voltage than the firing of an individual neuron. A second drawback is the limited anatomical specificity when compared with other functional brain imaging techniques such as functional magnetic resonance imaging (fMRI). Some anatomical specificity can be gained with the use of EEG topography, which uses a large number of electrodes to triangulate the source of the electrical activity. The main advantage of EEG as a tool of exploring the brain activity is its very high time resolution. As other methods for researching brain activity have time resolution between seconds and minutes, the EEG has a resolution down to sub-millisecond. EEG is the only method to directly measure brain's electrical activity. Recent attempts try to combine EEG or MEG with MRI, fMRI or PET to improve spatial resolution [Babiloni et al., 2003].

2.2.1.1 Frequency range

EEG recordings usually present rhythmical patterns. EEG waves can be classified according to different brain functions, but the terminology is imprecise and sometimes abused because traditionally brain waves were classified on the basis of visual inspection and not using precise frequency analysis. Keeping in mind that there is no precise agreement on the frequency ranges for each type, we can define five main types of EEG waves as shown in Figure 2.6.

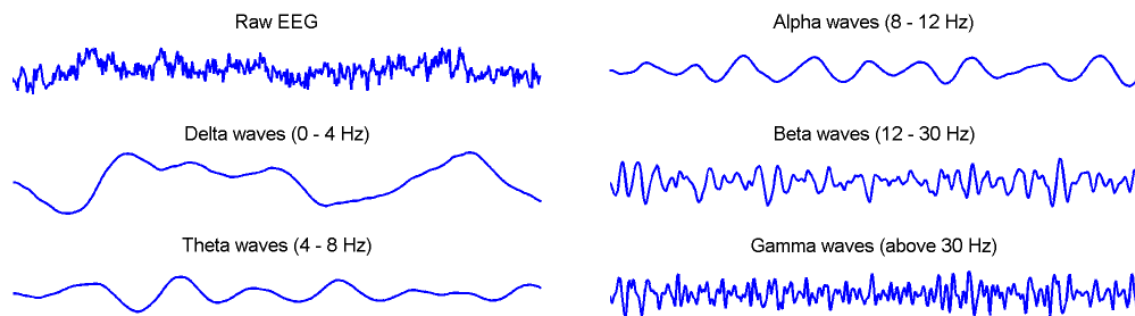


Figure 2.6: Different types of sinusoidal EEG activity

The top left plot shows one second of raw EEG signal and the five other plots illustrate the five described EEG waves.

- **Delta** is the lowest frequency range, below 4 Hz. It is typical of infants and is present in deep sleep and in some organic brain diseases.
- **Theta** is the frequency range from 4 to 8 Hz and is associated with drowsiness, childhood, adolescence and young adulthood. This EEG frequency can sometimes be produced by hyperventilation. Theta waves can be seen during hypnagogic states such as trances, hypnosis, deep day dreams, lucid dreaming and light sleep and the preconscious state just upon waking, and just before falling asleep.
- **Alpha** is the frequency range from 8 to 12 Hz. It is characteristic of a relaxed, alert state of consciousness. For alpha rhythms to arise, usually the eyes need to be closed. Alpha attenuates with drowsiness and open eyes, and typically come from the occipital (visual) cortex. An alpha-like normal variant called **Mu** is sometimes seen over the motor cortex (central scalp) and attenuates with movement, or rather with the intention to move.
- **Beta** is the frequency range from 12 to 30 Hz. Low amplitude beta with multiple and varying frequencies is often associated with active, busy or anxious thinking and active concentration. Rhythmic beta with a dominant set of frequencies is associated with various pathologies and drug effects.
- **Gamma** is the frequency range from approximately 30 to 100 Hz. Gamma rhythms may be involved in higher mental activity, including perception, problem solving, fear, and consciousness.

2.2.1.2 Artifacts

Contamination of EEG by signals of non-cerebral origin is present in most EEG recordings, these contaminating signals are called artifacts. Eye artifacts can be of several different nature. Eyeball artifacts are caused by the potential difference between the cornea and retina, which is quite large compared to cerebral potentials. There are nearly always small or large reflexive eye movements, which generates a potential which is picked up by the fronto-polar and frontal electrodes. Eye movements (whether vertical or horizontal) are caused by ocular muscles, which also generate another kind of artifacts, electro-myographic (EMG) artifacts. Purposeful or reflexive eye blinking also generates EMG artifacts, but more importantly there is reflexive movement of the eyeball during blinking which gives a characteristic artifactual appearance of the EEG. Electrocardiogram (ECG) artifacts are quite common and can be mistaken for spike activity. Glossokinetic artifacts are caused by the potential difference between the base and the tip of the tongue. Minor tongue movements can contaminate the EEG.

2.3 Event-related potentials (ERP)

Event-related potentials (ERP) are signals generated by a population of neurons in response to a perceptual, cognitive or motor event, in opposition to spontaneous activity that reflects the brain activity related to volunteer self-paced tasks. Evoked potentials can be seen as a specific kind of ERP generated directly in response to external stimulus such as visual evoked potentials (VEP) and auditory evoked potentials (AEP). Reactions to stimuli or events lead to variations of the electrical activity of specific brain areas and the resulting EEG traces exhibit modifications called potentials. In the case of external stimuli, for example discrete visual feedback, the precise time of the stimulus is known. It's therefore possible to extract averages of the stimulus-locked response of the brain. Time-locked averages allow elimination of random noise while keeping track of the ERP components. When the precise time of the stimulus is not available, it's much more complicated to extract ERP from the ongoing EEG. In this case, a specific action of the subject related to the nature of the stimulus can be used as trigger. In any case, the main challenge is the detection of ERP no more in averages of many single trials, but directly at the level of the single trial. Several ERP are now described with a specific focus on error-related potentials (ErrP).

- **The P300** shows up as a very prominent positive deflection (usually with a parietal focus) of the ongoing EEG about 300 ms after the occurrence of an infrequent or particularly significant stimulus interspersed with frequent stimuli. This stimulus can be of various nature: visual, auditory or somatosensory.
- **Visual evoked potentials (VEP)** are EEG waveforms generated in response to visual stimuli that can be used to determine the direction of eye gaze. When stimuli are presented in a rapid succession, the evoked potentials overlap in time and the presentation rate is high enough to evoke a steady wave, this is referred to as steady-state visual evoked potentials (SSVEP).
- **Bereitschaftspotential (PB)**, also called readiness potential or pre-motor potential reflects activity in the motor cortex during voluntary muscle movement preparation. BP is a negative deflection which develops in a central area during the last second before limb

movement. The spatial distribution of BP varies according to the used limb, for example a larger amplitude shows up contralateral to the moving finger. BP is also present in the case of movement imagination.

2.3.1 Error-related potentials (ErrP)

Up to 1989, only a few studies mentioned the presence of error-related potentials on error trials and specific error-related EEG components were not recognized in any of them [Coles et al., 1985, Donchin et al., 1988]. Since the early 1990s, many studies have shown the presence of ErrP components such as error-related negativity (ERN or Ne) and error-related positivity (Pe) in many different situations. The *response ErrP*, described exhaustively since the early 1990s [Falkenstein et al., 1990, Gehring et al., 1990, Bernstein et al., 1995, Scheffers et al., 1996], is elicited in speeded reaction tasks ~ 100 ms following an erroneous response [Carter et al., 1998, Falkenstein et al., 2000, Vidal et al., 2000, Gehring and Fencsik, 2001, Coles et al., 2001, Rodriguez-Fornells et al., 2002, Ridderinkhof et al., 2003, Nieuwenhuis et al., 2001]. The *feedback ErrP*, first reported in 1997 [Miltner et al., 1997], is elicited in reinforcement learning tasks ~ 250 ms following presentation of a feedback that indicates incorrect performance [Miltner et al., 1997, Badgaiyan and Posner, 1998, Holroyd and Coles, 2002, Gehring and Willoughby, 2002, Ullsperger and von Cramon, 2003, Holroyd et al., 2003, Luu et al., 2003]. Finally other studies reported the presence of *observation ErrP* following observation of errors made by an operator during choice reaction tasks where the operator needs to respond to stimuli [van Schie et al., 2004]. As in the case of Feedback ErrP, the main component here is a negative potential showing up 250 ms after the incorrect response of the subject performing the task.

2.3.1.1 Choice reaction task

In a typical choice reaction task, a subject is asked to respond as quickly as possible to a stimulus. This stimulus can be of different kind: visual, auditory, etc and the response usually consists in pressing a key or in moving a limb. There are two main components of ErrP in an error trial: a negative potential (error negativity ERN, or Ne) showing up as a sharp negative component peaking at ~ 100 ms after the incorrect key press with a fronto-central maximum, and a positive potential (error positivity Pe) showing up as a broader positive component with a peak between 200 and 500 ms after the incorrect key press with a centro-parietal maximum (for reviews see [Falkenstein et al., 2000, Vidal et al., 2000]). More generally, ErrP are associated with error processing. Studies showed the presence of the Ne in correct trials and propose that the Ne reflects a comparison process, required in both correct and erroneous trials. Pe is a further error-specific component, independent of the Ne, and associated with a later aspect of error processing [Vidal et al., 2000]. Other studies focused on awareness of errors, they showed that irrespective of whether the subject is aware of the error or not, erroneous trials are followed by the Ne. In contrast the Pe is much more pronounced for perceived errors. These findings verify the hypothesis that the Ne reflects an unconscious comparison process and the Pe reflects a later conscious error processing [Nieuwenhuis et al., 2001].

2.3.1.2 Reinforcement learning tasks

In a typical reinforcement learning task, a subject is asked to make a choice among several possibilities knowing that some are correct and some others are incorrect. He is also asked to adapt his strategy in order to try to minimize the occurrence of the negative feedback. After making a choice the system provides a feedback indicating correct or incorrect performance. The negative feedback, indicating incorrect performance, elicits an ERN, whose main component is a negative deflection that occurs ~ 250 ms after the feedback [Miltner et al., 1997]. This negative deflection is also present in correct trials, but its amplitude is smaller [Holroyd et al., 2003, Gehring and Willoughby, 2002]. Since the feedback is delivered some time after the response of the subject, this kind of ERN is not elicited by the response itself. Indeed, several studies show that the error processing neural system is concerned with aspects of error processing not directly linked to erroneous responses [Ullsperger and von Cramon, 2003, Badgaiyan and Posner, 1998, Luu et al., 2003]. ERN are elicited by erroneous response in a wide variety of tasks and the error processing system is sensitive to various sources of error information. Similarly, ERN can be elicited by different negative feedback, visual, auditory and somatosensory. All these aspects give the evidence of a highly flexible error processing system (for a review see [Holroyd and Coles, 2002]).

2.3.1.3 Observation tasks

In observation tasks, the subject is asked to carefully observe an operator executing a choice reaction task. This task is a monitoring task since the subject has no particular action to perform. As in the case of the reinforcement learning tasks, the main component is here a negative deflection of the ongoing EEG 250 ms after an erroneous response of the operator was observed in this kind of tasks [van Schie et al., 2004].

2.3.1.4 Source localization

The pre-supplementary motor area (pre-SMA, Brodmann area 6) and the rostral cingulate zone (RCZ, Brodmann areas 24 & 32) seem to be the main brain areas involved in error processing, although there exists disagreement on the exact role played by these areas in error processing [Holroyd and Coles, 2002, Luu et al., 2003, Fiehler et al., 2004]. Lateral prefrontal cortex (PFC) and orbito-frontal cortex activation during error detection have been extensively commented. Interaction between the PFC on one side and the ACC and the pre-SMA on the other side has been reported during monitoring behavior and guiding compensatory system [Gehring and Knight, 2000]. PFC could be related to monitoring processes whereas ACC and pre-SMA are more involved in error detection. Finally, associative areas (such as somatosensory association cortex, Brodmann areas 5 & 7) activation could be related to the fact that the subject becomes aware of the error. It has been proposed that the positive peak generated in a reaction task was associated with conscious error recognition [Nieuwenhuis et al., 2001].

2.4 Estimation of intracranial activity from EEG

Brain electromagnetic tomography, i.e. the non-invasive three-dimensional reconstruction of the neuronal sources of the brain's electrical activity measured at the scalp, is a very wide and

complex research field. For this work, we only considered these inverse solutions as available tools for BCI systems, a description of the design and development of such models is beyond the scope of our work, but interested readers will find several references throughout this section. However, this section should provide all the necessary elements for a good understanding of the results presented later in this thesis.

2.4.1 The inverse problem: A general approach

Estimating the neuronal sources that generated a given potential map at the scalp surface requires the solution of an inverse problem. As we will see in Section 2.4.2, such inverse problems are always initially undetermined, i.e. there is no unique solution. These problems require therefore supplementary a priori constraints in order to be univocally solved. The ultimate goal is then to unmix the signals measured at the scalp, attributing to each brain area its own estimated temporal activity.

Historically, two different possible directions have been investigated in order to solve this inverse problem and find the generators of a given scalp activity, a global review can be found in [Michel et al., 2004]. On one hand, the so-called *dipole localization models* assume that only a limited number of generators are active over a period of time [Scherg, 1990, Darvas et al., 2004]. These generators are typically built as equivalent current dipoles (ECD). The number of generators that can be active at a given time is limited by the number of electrodes used for EEG measurements. Thus, when in a given problem, the exact number of dipole sources cannot be determined a priori, this family of methods is not very appropriate. In such cases, *distributed models* based on the linear theory in conjunction with mathematical and/or biophysical a priori constraints are more appropriate [Dale and Sereno, 1993, Fuchs et al., 1999, Babiloni et al., 2000a, 2003, Grave de Peralta Menendez et al., 2004]. These distributed models do not need a priori assumptions about the number of source generators, and estimate cortical current density by using sophisticated computational algorithms and detailed geometrical models of the head as volume conductor. With this approach, typically thousands of ECD covering evenly the cortical mantle are used, and their strength is estimated by using linear inverse procedures. In this work, only distributed models are considered.

2.4.2 Distributed linear inverse estimation

For clarity purposes, we adopt the notation used in [Cincotti, 2002] for the formulation of inverse estimation, and we also follow that work to present the general form of a distributed linear inverse estimation. Assuming a measurement noise \mathbf{n} , an estimate of the dipole source configuration that generated a scalp potential \mathbf{b} is obtained by solving the linear system:

$$\mathbf{A}\mathbf{x} + \mathbf{n} = \mathbf{b} \quad (2.1)$$

where \mathbf{A} is a $m \times n$ matrix with m the number of sensors and n the number of considered sources. The matrix \mathbf{A} is called the leadfield matrix: the j^{th} column A_j represents the potential distribution over the m sensors due to each unitary j^{th} dipole, and the collection of A_j describe how each dipole generates the potential distribution over the head model. The estimation of the cortical current density \mathbf{x} is called the solution of the linear inverse problem, or inverse solution. In most cases, the dimension of the vector \mathbf{x} is greater than the number of measurements \mathbf{b} of about one order of magnitude; thus, the linear system is strongly under-determined, and can

have an infinite number of possible solutions. In order to solve this problem for a unique solution, assuming that \mathbf{n} is normally distributed, a regularization scheme utilizing the Lagrange multiplier λ is applied, and the following functional has to be minimized:

$$\hat{\mathbf{x}} = \arg \min_x (\Phi), \quad \Phi = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_M^2 + \lambda^2 \|\mathbf{x}\|_N^2 \quad (2.2)$$

where the matrix \mathbf{N} is the metric of the source space, i.e. the space of the current strength solutions \mathbf{x} , and the matrix \mathbf{M} is the metric of the data space, namely the space in which \mathbf{b} is considered. If no a priori information is added to equation (2.2), \mathbf{M} and \mathbf{N} are set to identity, and the estimation made is called *minimum norm estimation* (MN). Interpreting (2.2), it appears that on one hand, we try to minimize the energy of the error on the sensor data, given by the first term of Φ . On the other hand, a second term involving the energy of the source \mathbf{x} regularizes the ill-posed problem: this term, modulated by λ , tends to minimize the overall intensity of the current distribution. At the end, a unique solution will be found, because only one combination of intracranial sources fit exactly the data, and has at the same time the lowest overall intensity. The problem is that the algorithm favors weak and localized activation patterns, instead of solutions with strong activation of a large number of solution points. Thus, the MN algorithm favors superficial sources, since less activity is required in superficial solution points to provide a certain surface voltage distribution: such models are not satisfying, because it means that deeper sources are incorrectly projected on the surface of the scalp.

In order to cope with this problem, a well-known solution proposes to take into account a compensation factor for each dipole that equalizes the visibility of the dipoles from the sensors point of view. This so-called *column norm normalization* changes the source metric \mathbf{N} as follows:

$$(\mathbf{N}^{-1})_{ii} = \|A_{\cdot i}\|^{-2} \quad (2.3)$$

with $(\mathbf{N}^{-1})_{ii}$ the i^{th} element of the inverse of the diagonal matrix \mathbf{N} and $\|A_{\cdot i}\|$ the L_2 norm of the i^{th} column of the lead matrix \mathbf{A} . The use of this definition of the matrix \mathbf{N} is known as *weighted minimal norm solution* (WMN), and penalizes dipoles close to the sensors in the solution of the inverse problem, since they have a large $\|A_{\cdot i}\|$. Thus, WMN solutions provide better estimates of intracranial activity, especially in the case of deep sources. Equations (2.1), (2.2) and (2.3) set a general framework for distributed linear inverse models. From that point, a lot of free parameters have to be carefully chosen in order to converge to the best unique solution as possible. For example, the choice of additional constraints is crucial in terms of model specificity, and can drastically change the behavior of the inverse solution. Additional constraints come from assumptions about likely current source distribution and statistics, sensor statistics, and information from other imaging techniques. In the next section, three inverse models with different assumptions are presented.

2.4.3 Inverse solutions

2.4.3.1 CCD inverse model

The first presented model, that we will call CCD inverse model for *cortical current density inverse model*, has been developed and made available by a research group working in the IRCCS

Fundazione Santa Lucia, located in Rome². References about this approach can be found in [Babiloni et al., 2000a, 2005]. The model aims at providing an estimation of the activity of the cortical mantle. The procedure follows the reasoning of section 2.4.2 and includes:

- a realistic magnetic resonance-constructed average head model.
- a multi-dipole cortical source model.
- a regularized, weighted, minimum-norm linear inverse source estimate based on boundary element mathematics (WMN).

First, a geometrical reconstruction of the cortical surface is obtained from magnetic resonance imaging (MRI). In this model, the 152 subjects average brain of Montreal Neurological Institute³ was used to have a realistic head model. At that point, an important anatomical constraint is considered: it is assumed that much of the observable EEG is produced by currents flowing in the apical dendrites of cortical pyramidal cells. The columnar organization of the cortex implies that the resulting local dipole moment is assumed to be oriented perpendicularly to the cortical surface. Thus, if the shape of the cortical mantle is known, we can divide it into patches that are sufficiently small so that a dipole in the center of a patch is representative of any dipole distribution within the patch. With the constraint of perpendicular orientation of the dipoles, the inverse problem reduces to estimating scalar distributions of dipole strength over the oriented patch.

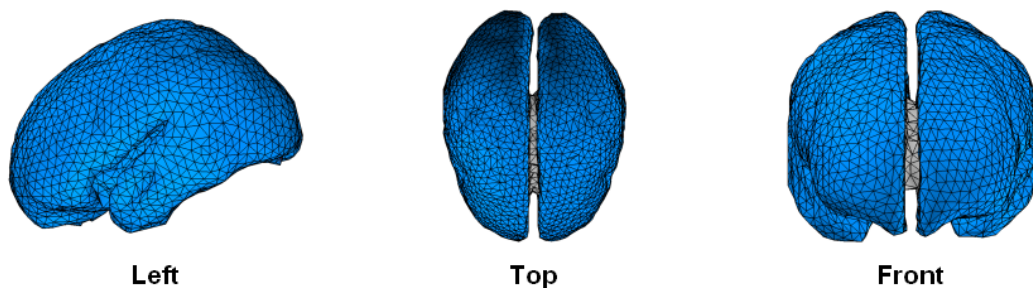


Figure 2.7: CCD inverse model brain views.

Views of the cortical mantle reconstructed with 2949 vertices on a polyhedron with triangular faces.

In the case of the CCD inverse model, the MRI-based reconstruction of the head models the cortical mantle as a polyhedron with triangular faces, preserving the general features of the neocortical envelope. An orthogonal unitary ECD is then placed in each node (or *vertex*) of the triangulated surface. On the whole, 2949 discrete current dipoles are chosen to represent the continuum current source distribution. Figure 2.7 shows a view of the brain provided by the model.

The second constraint of the CCD inverse model is based on WMN estimates, and forces the dipoles to explain the recorded data with a minimum or a low amount of energy, without

²<http://www.hsantalucia.it/>

³<http://www.bic.mni.mcgill.ca/>

penalizing too much deeper sources, as explained in section 2.4.2. During this thesis work, the CCD inverse model has been used extensively, both for localization studies and for BCI-oriented classification procedures, and showed impressive results.

2.4.3.2 sLORETA inverse model

The second inverse model is the *standardized low resolution brain electromagnetic tomography* method (sLORETA). This software, known for its zero localization error, is freely provided by the KEY Institute for Brain-Mind Research⁴ in Zürich. We used this software only as a localization tool throughout the studies, but a description of the method for localizing sources is useful here. The volume conductor model is a three-shell spherical head model registered to the Talairach human brain atlas [Talairach and Tournoux, 1988], available as a digitized MRI from the Montreal Neurological Institute imaging center. The solution points are placed on a 3D regular grid covering the whole brain.

Historically, a first method called low resolution electromagnetic tomography (LORETA) was introduced by Pascual-Marqui in [Pascual-Marqui et al., 1994, 2002]. In this method, an additional constraint called *Laplacian Weighted Minimum Norm* was added to the typical WMN depth weighting. This method selects the solution having the smoothest spatial distribution by minimizing the Laplacian of the weighted sources, a measure of spatial roughness. A physiological assumption is hidden behind this method: the model assumes that neighboring grid points, i.e. neighboring neurons, are more likely to be synchronized (similar orientation and strength) than grid points that are far from each other. Thus, this maximization of smoothness is applied to find a unique distribution of electrical activity in the brain. The characteristic feature of this solution is its low spatial resolution, which is a direct consequence of the smoothness constraint. LORETA provides rather blurred images of a point source, conserving the location of the maximal activity with a certain degree of dispersion. Furthermore, the assumption that two neighboring areas are correlated has to be considered with caution. Indeed, functionally distinct areas can be anatomically very close. However, the localizations made by LORETA are satisfying in most cases.

Recently, a new version of the method, called *standardized low resolution brain electromagnetic tomography* (sLORETA) has been developed, and yields images of standardized current density with zero localization error. The difference with the previous algorithm is that sLORETA employs the current density estimate given by the minimum norm solution, and localization inference is based on standardized values of the current density estimates, as explained in [Pascual-Marqui, 2002]. Only by itself, the solution of the MN inverse solution is incapable of correct localization of deep sources. With this standardization process, sLORETA reaches zero localization error, even if the sources are deep. However, the drawback of this method is that because of this standardization process, sLORETA *is not* an authentic solution to the inverse problem; according to the KEY Institute web site, it seems that a new version of the software, called eLORETA (for *exact low resolution brain electromagnetic tomography*) will soon be released, and will provide a formal solution providing exact localization to test point sources.

⁴<http://www.unizh.ch/keyinst/index.html>

2.4.3.3 ELECTRA inverse model

The third inverse solution presented here is slightly different from the previous models. This is a distributed source model called ELECTRA (for electrical analysis), developed in the Geneva University Hospital (HUG) [Grave de Peralta Menendez et al., 2000, 2004]. In conjunction with this linear distributed model, a regularization strategy called LAURA (local autoregressive averages) is applied on the inverse solution. The difference of ELECTRA-LAURA model is that the source model is changed with respect to the previous models, based on the following considerations. The microscopic current flowing in biological tissue can be decomposed into two terms. A *primary current* (or *active current*) and a *secondary current* (or *volume current*). Primary currents are induced by ionic flow between intra- and extra-cellular space in activated neurons, whereas volume currents are passive currents representing the electrical response of the media to compensate charge accumulation driven by primary currents, according to electrochemical gradient. It has been shown in [Plonsey, 1982] that only volume currents are measured by EEG, and not active currents. This observation is crucial, since the mathematical implication is that the currents measured by EEG are ohmic and can be restricted to *irrotational* currents. Thus, the ELECTRA source model only estimates ohmic currents. It is not an inverse solution, but rather a source model in which the generators of the scalp maps are the intracranial potentials instead of the usual 3D current densities. In order to reach a unique solution, the LAURA regularization strategy incorporates biophysical laws as constraints in the MN algorithm [Grave de Peralta Menendez et al., 2001, 2004]. According to Maxwell equations, the strength of the sources fall off with the inverse of the cubic distance for vector fields, and with the inverse of the squared distance for potential fields. LAURA integrates these laws in terms of a local autoregressive average with coefficients depending on the distances between solution points. The model is composed of a solution space formed by 4024 nodes (referred to as *voxels*) homogeneously distributed within the inner compartment of a realistic head model. Figure 2.8 shows different views of this solution space. Once again, the head model is the average brain of Montreal Neurological Institute was used. The voxels are restricted to the grey matter and form an isotropic grid of 6 mm resolution.



Figure 2.8: ELECTRA solution space

Views of the solution space for the 3D reconstruction of the brain by means of 4024 voxels.

The most interesting point with ELECTRA-LAURA inverse solution is that the model allows an estimation of the 3D distribution of electrical potentials in the brain as if they were recorded with intracranial electrodes. The EEG activity mainly reflects the synchronous activity of a large population of neurons, and more precisely their postsynaptic activity. The

intracranial measure of this postsynaptic activity is called local field potentials (LFPs) [Bear et al., 2001]. Recently, LFPs revealed themselves to be of crucial interest for providing meaningful information about neuronal processes related to motor actions [Donchin et al., 2001, Rickert et al., 2005, Scherberger and Jarvis, 2005], and more generally about brain dynamics [Victor et al., 1994, Galván et al., 2002]. Hence, estimating the LFP activity from the scalp EEG represent a very challenging and exciting issue, since it can provide a non-invasive way to investigate neuronal processes in humans with a highly relevant physiological meaning.

2.5 Brain-computer interfaces (BCIs)

The idea of an EEG-based communication system was first introduced in the 1970s by Vidal [Vidal, 1973, 1977]. But the first BCI prototypes became feasible only recently with the discovery of the mechanisms and spatial location of many brain wave phenomena and their relationship with specific aspects of brain function. Furthermore, the development of powerful computers made possible the creation of systems able to perform the online analysis of multichannel EEG. Up to date the main applications of EEG-based BCIs are the operation of a virtual keyboard for letter selection on a computer screen [Birbaumer et al., 1999, Obermaier et al., 2003, Wolpaw et al., 2003, Millán et al., 2004b], the basic control of a hand prosthesis [Pfurtscheller et al., 2000] and the control of a wheelchair in an indoor-like environment [Millán et al., 2004a, Philips et al., 2007]. As already mentioned in the Introduction of this thesis, a BCI monitors the user's brain activity and translates their intentions into actions without using activity of any muscle or peripheral nerve. The central tenet of a BCI is the capability to distinguish different patterns of brain activity, each being associated to a particular intention or mental task. Such a kind of BCI is a natural way to augment human capabilities by providing a new interaction link with the outside world and is particularly relevant as an aid for paralyzed humans, although it also opens up new possibilities in natural and direct interaction for able-bodied people. Figure 1.1 shows the general architecture of a non-invasive EEG-based BCI. Brain electrical activity is recorded with a portable device. These raw signals are first processed and transformed in order to extract some relevant features that are then passed on to some mathematical models (e.g., statistical classifiers or neural networks). This model computes, after some training process, the appropriate mental commands to control the device. Finally, visual feedback, and maybe other kinds such as tactile stimulation, informs the subject about the performance of the brain-actuated device so that they can learn appropriate mental control strategies and make rapid changes to achieve the task.

Present-day non-invasive BCIs can be classified into 4 groups according to the electrophysiological signals they use. The first group, using visual evoked potentials (VEP), are said to be dependent BCIs because they depend on peripheral nerves or on muscular control. The other three groups, using slow cortical potentials (SCP), P300 evoked potentials, and mu and beta rhythms, are said to be independent BCIs because they do not depend on peripheral nerves or muscles, but this assumption is sometimes controversial. Invasive BCIs using electrodes directly implanted in the motor cortex of monkeys and humans have provided stable neuronal recordings and encouraging results [Chapin et al., 1999, Carmena et al., 2003, Musallam et al., 2004, Kennedy and Bakay, 1998, Kennedy et al., 2000, Hochberg et al., 2006], but non-invasive techniques are necessary for humans. For reviews on BCIs see [Millán, 2002, Nicolelis, 2001, Wickelgren, 2003, Wolpaw et al., 2002].

2.5.1 Visual evoked potentials

The most widely used evoked potentials in BCIs are visual evoked potentials that are EEG waveforms generated in response to visual stimuli that can be used to determine the direction of eye gaze. The BCI developed by Vidal in the 1970s [Vidal, 1973, 1977] using visual evoked potentials satisfy the current definition of a dependent BCI. This system basically use visual evoked potentials recorded from the scalp over visual cortex to recognize the direction of the visual fixation point and so, for instance, can be used to determine the direction in which the subject wishes to move a cursor. When the repetition rate of the visual stimulus is faster than 6 Hz, the new stimulus is presented before the last response of the visual system vanishes, a periodic response called steady state VEP (SSVEP) is generated. SSVEP are usually used in the short-term identification of evoked responses because of its high signal-to-noise ratio (SNR) [Sutter, 1992, Middendorf et al., 2000, Gao et al., 2003].

2.5.2 P300 evoked potentials

P300 is a positive deflection in the EEG occurring about 300 ms after an infrequent or particularly significant stimulus interspersed with frequent stimuli. This stimulus can be of various nature: visual, auditory or somatosensory. Donchin and his colleagues have used this P300 potential to design a BCI [Farwell and Donchin, 1998, Donchin et al., 2000]. P300-based BCIs have the advantage that they require no prior training as P300 is a typical response to a desired choice. Nevertheless, the amplitude of P300 diminishes over time so that the performances of the BCI could decrease. Another drawback of P300-base BCIs and any kind of BCI based on evoked potentials is the dependency on external stimulations. The variety of tasks is limited and the subject only sustains the stimuli instead of being in total control on the interaction.

2.5.3 Slow cortical potentials

Slow cortical potentials (SCP) are slow voltage changes lasting from 500 ms to a few seconds. Negative SCP are typically associated with movements and other functions involving cortical activation whereas positive SCP are usually associated with reduced cortical activation. Different studies have shown that people can learn to control SCP and so control movements of an object on a computer screen [Birbaumer et al., 1999, Elbert et al., 1980]. This demonstration is the basis for a BCI referred to as a “Thought Translation Device” (TTD) [Birbaumer et al., 1999].

2.5.4 Mu and beta rhythms

In awake people, primary sensory or motor cortical areas often display 8-12 Hz EEG activity called mu rhythms. These mu rhythms usually appear in conjunction with 18-26 Hz beta rhythms. These rhythms could be good signal for EEG-based communication. Movements or preparation of movements are associated with a decrease in mu and beta rhythms. This decrease has been named *event-related de-synchronization* (ERD) [Pfurtscheller and da Silva, 1999]. The opposite increase in rhythms (named *event-related synchronization*, ERS) occurs after movements and during relaxation. The most important property of ERD and ERS is that

they occur also with motor imagery and so they could be used in the framework of an independent BCI. Several groups have developed mu/beta rhythms-based BCIs [Wolpaw et al., 2003, Pfurtscheller and Neuper, 2001, Kostov and Polak, 2000, Penny et al., 2000].

2.6 The IDIAP BCI

In the section we review the main components of our BCI system, which is based on the online analysis of spontaneous EEG signals and recognizes three mental tasks. Our approach relies on three principles. The first one is an asynchronous protocol where subjects decide voluntarily when to switch between mental tasks and perform those mental tasks at their own pace. The second principle is mutual learning, where the user and the BCI are coupled together and adapt to each other. In other words, we use machine learning approaches to discover the individual EEG patterns characterizing the mental tasks executed by the user while users learn to modulate their brain waves so as to improve the recognition of the EEG patterns. Finally, the third principle is the combination of the user's intelligence with the design of intelligent devices that facilitate interaction and reduce the user's cognitive workload. This is particularly useful for mental control of robots.

2.6.1 Spontaneous EEG and asynchronous operation

Evoked potentials are, in principle, easy to pick up with scalp electrodes. The necessity of external stimulation does, however, restrict the applicability of evoked potentials to a limited range of tasks. In our view, a more natural and suitable alternative for interaction is to analyze components associated with spontaneous *intentional* mental activity. This is particularly the case when controlling robotics devices. Spontaneous BCIs are based on the analysis of EEG phenomena associated with various aspects of brain function related to mental tasks carried out by the subject at his/her own will. EEG-based BCIs are limited by a low channel capacity. Most of the current systems have a channel capacity below 0.5 bits/s [Wolpaw et al., 2002]. One of the main reasons for such a low bandwidth is that they are based on synchronous protocols where EEG is time-locked to externally paced cues repeated every 4-10 seconds and the response of the BCI is the overall decision over this period [Birbaumer et al., 1999, Pfurtscheller and Neuper, 2001, Wolpaw and McFarland, 2004]. Such synchronous protocols facilitate EEG analysis since the starting time of mental states are precisely known and differences with respect to background EEG activity can be amplified. Unfortunately, they are slow and BCI systems that use them normally recognize only two mental states. On the contrary, we utilize more flexible asynchronous protocols where the subject makes self-paced decisions on when to stop doing a mental task and start immediately the next one [Millán et al., 2004a,b, Birch et al., 2002]. In such asynchronous protocols the subject can voluntarily change the mental task being executed at any moment without waiting for external cues. The time of response of an asynchronous BCI can be below 1 second. For instance, in our approach the system responds every 1/2 second. The rapid responses of our asynchronous BCI gives a theoretical channel capacity between 1 and 1.5 bits/s.

2.6.2 The machine learning way to BCI

A critical issue for the development of a BCI is training, i.e. how users learn to operate the BCI. Some groups have demonstrated that some subjects can learn to control their brain activity through appropriate, but lengthy, training in order to generate fixed EEG patterns that the BCI transforms into external actions [Birbaumer et al., 1999, Wolpaw and McFarland, 2004]. In this case the subject is trained over several months to modify the amplitude of their EEG signals. We follow a mutual learning process to facilitate and accelerate the user's training period. Indeed, our approach allows subjects to achieve good performances in just a few hours of training in the presence of feedback [Millán et al., 2004b]. Most BCI systems deal with the recognition of just two mental tasks [Babiloni et al., 2000b, Pfurtscheller and Neuper, 2001, Blankertz et al., 2006, Birch et al., 2002]. Our approach achieves error rates below 5% for three mental tasks, but correct recognition is 70%. In the remaining cases (around 20-25%), the classifier doesn't respond, since it considers the EEG samples as uncertain. The incorporation of rejection criteria (see below) to avoid making risky decisions is an important concern in BCI. From a practical point of view, a low classification error is a critical performance criterion for a BCI; otherwise users can become frustrated and stop utilizing it. As explained in the next sections, we use machine learning techniques at two levels, namely feature selection and training the classifier embedded into the BCI. The approach aims at discovering subject-specific spatio-frequency patterns embedded in the continuous EEG signal, i.e. EEG rhythms over local cortical areas that differentiate the mental tasks.

2.6.3 Shared control

BCI systems are being used to operate a number of brain-actuated applications that augment people's communication capabilities, provide new forms of entertainment, and also enable the operation of physical devices. Until recently, EEG-based BCIs have been considered too slow for controlling rapid and complex sequences of movements. But we have shown for the first time that asynchronous analysis of EEG signals is sufficient for humans to continuously control a mobile robot along non-trivial trajectories requiring fast and frequent switches between mental tasks [Millán et al., 2004a,b]. Two human subjects learned to mentally drive the robot between rooms in a house-like environment visiting 3 or 4 rooms in the desired order. Furthermore, mental control was only marginally worse than manual control on the same task. A key element of this brain-actuated robot is shared control between two intelligent agents—the human user and the robot—so that the user only gives high-level mental commands that the robot performs autonomously. In particular, the user's mental states are associated with high-level commands and that the robot executes these commands autonomously using the readings of its on-board sensors. Another critical feature is that a subject can issue high-level commands at any moment. This is possible because the operation of the BCI is asynchronous and, unlike synchronous approaches, does not require waiting for external cues. The robot relies on a behavior-based controller to implement the high-level commands to guarantee obstacle avoidance and smooth turns. In this kind of controller, on-board sensors are read constantly and determine the next action to take. As explained in Section 2.8, we have recently extended this work to the mental control of both a simulated and a real wheelchair.

2.7 Data acquisition and processing

2.7.1 Hardware

For the studies carried out in the BCI Group of the IDIAP Research Institute⁵, we acquire EEG potentials with a portable BioSemi⁶ system using a cap with either 32 or 64 integrated electrodes arranged in the modified 10/20 International System as shown in Figure 2.9. This ActiveTwo EEG acquisition system requires no prior skin preparation and is battery powered. The subject is protected for leakage currents from the main supply by the isolation barrier between the amplifier and the PC: the optical fiber data-link combined with battery power supply provides complete safety. The digital resolution is 31 nV and the input range is 524 mVpp. With the ActiveTwo, all filtering is digital. High-pass filtering (if any) is completely performed in software, the hardware is completely DC coupled. Low pass filtering is performed in the ADC's decimation filter (hardware bandwidth limit), which has a 5th order sinc response with a -3 dB point at 1/5th of the selected sample rate. Further variable bandwidth limiting (high and low pass) can be applied at will in software. Two extra electrodes CMS (Common Mode Sense) and DRL (Driven Right Leg) replace the ground electrodes which are used in conventional EEG systems. The CMS electrode measures the potential of the patient, the signals on file are the voltages between each electrode and CMS. The DRL electrode closes the loop between the patient and the A/D converter. The DRL electrode is directly connected to a Driven Right Leg circuit in order to reduce the common-mode voltage and protect the patient by limiting the output current [Metting Van Rijn et al., 1990].

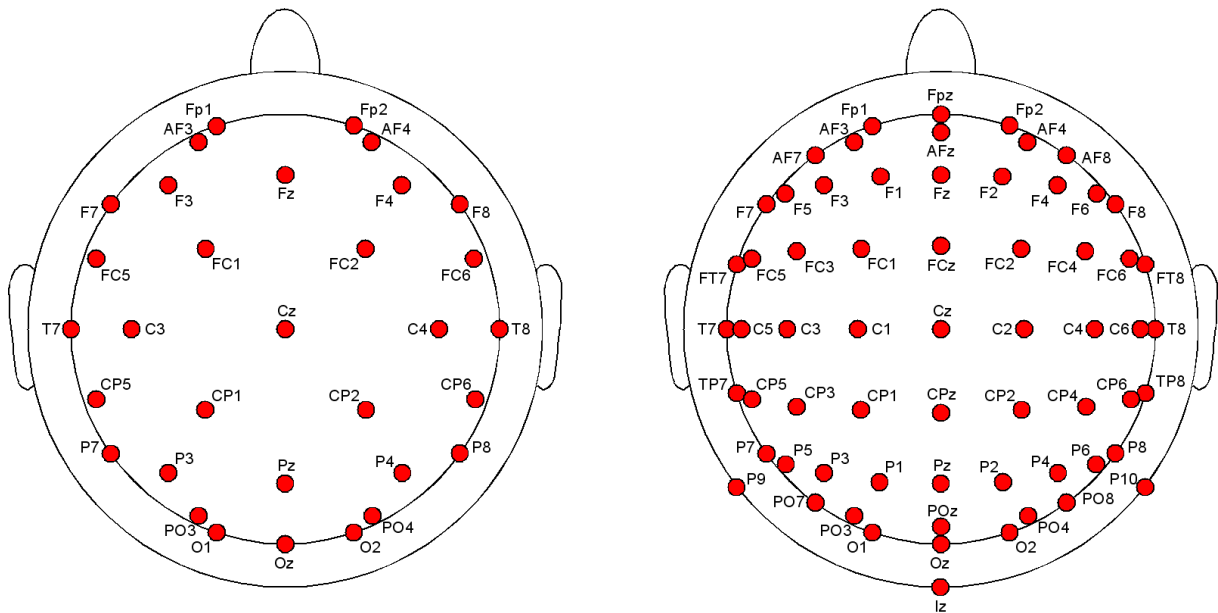


Figure 2.9: EEG electrodes positions.

Top view of the positions of the 32 (left) and the 64 (right) electrodes according to the modified 10/20 International System (frontal on top).

⁵<http://www.idiap.ch>

⁶<http://www.biosemi.com>

2.7.2 Signal processing

EEG potentials are acquired at 512 Hz. EEG signals are characterized by a poor signal-to-noise ratio (SNR) and spatial resolution. Their quality is greatly improved by means of spatial filtering techniques. We use the common average reference (CAR) procedure, where at each time step the average potential over all the channels is subtracted from each channel. This re-referencing procedure removes the background activity, leaving activity from local sources beneath the electrodes. Alternatively, raw EEG potentials can be transformed by means of a Surface Laplacian (SL) derivation. The SL estimate yields new potentials that represent better the cortical activity originated in radial sources immediately below the electrodes. The superiority of SL and/or CAR transformed signals over raw potentials for the operation of a BCI has been demonstrated in different studies [Babiloni et al., 2000a, Mouriño, 2003]. Then, for ErrP analysis we apply a 1 to 10 Hz bandpass filter, as ErrP are known to be relatively slow cortical potentials [Gehring et al., 1993]. Finally, EEG signals are sub-sampled from 512 Hz to 64 Hz (i.e., we took 1 point out of 8) before classification, which for ErrP is entirely based on temporal features.

2.7.3 Feature selection

As already mentioned, error-related potentials are most probably generated in the anterior cingulate cortex that is a deep fronto-central cortical area. EEG potentials are acquired by means of 32 or 64 electrodes uniformly arranged on the scalp. Therefore, it's natural to think that only the fronto-central electrodes (like FCz and its neighbors) will be relevant for successful ErrP classification. This point is even more evident when working with inverse solutions. From the scalp EEG we will estimate the activity of several thousands of solution points. However, it's most probable that only a small number of points are relevant to discriminate erroneous and correct responses of the interface. In this respect, it is a normal practice in machine learning to apply feature selection techniques to reduce the dimensionality of the input vector to the classifier, especially when the input space has a huge dimensionality and many of the features are not relevant for discriminating among the classes. This is particularly the case for the analysis of EEG signals and we have previously shown the advantages of such a feature selection for BCI [Millán et al., 2002a]. The goal of feature selection is to extract those relevant features for which the performance of the trained classifier is the best and to discard irrelevant features that may act as noise and hinder the classification. Therefore, feature selection and classifier training are closely related. There are two different approaches:

- Filter methods, where feature selection precedes the training of the classifier.
- Wrapper methods, where the feature selection algorithm uses the classifier training algorithm.

The main drawback of filter methods is that feature selection is completely independent of the classifier training algorithm, and the former cannot be guided by the classifier error rate. Wrapper methods use this feedback from the classifier to make their selection. The main advantage of filter methods is that their computational cost is usually lower since all features are treated independently.

In this section, we present a very simple but efficient feature selection method (filter method) that was called *discriminant power function* (DP) [Gonzalez Andino et al., 2006]. More precisely, we implemented a modified version of the DP function that can deal with noisy data. The basic method will be first introduced, and then the modified method implemented during this work will be described. It is to note that this simple algorithm has been compared to other more sophisticated filter methods⁷ and that it couldn't be outperformed when tested on EEG data [Uldry et al., 2007]. An interesting result reported in [Uldry et al., 2007] is that the best classification accuracies are already obtained with a small number of features, typically less than 50. Thus the dimension of the input space for the classification is rather small, allowing real-time implementation. For example, Figure 2.10 shows the classification rates of error and correct single trials using different feature selection methods and different numbers of selected features. As already mentioned, the DP algorithm performs as well as other more complicated ones, and the best accuracies are reached using a small number of features. Also, Figure 2.10 shows how performance is rather stable, or slightly decreases with an increasing number of features. Note that the analysis was performed using the CCD inverse model (see Chapter 8) so that there are ~ 3000 possible features for classification.

2.7.3.1 Discriminant power (DP) algorithm

The basic DP function estimates the quality of a given feature following a very simple principle: if the distribution of the feature, namely its *probability density function* (pdf), is different for each class, then the feature is a good candidate to discriminate between these classes. More precisely, let us take the example of a two class problem. If the distribution of the feature f for class 1, $pdf_1(f)$, is well separated from the distribution of the same feature for class 2, $pdf_2(f)$, then the feature f has a high discriminant power. Otherwise if $pdf_1(f)$ and $pdf_2(f)$ are strongly overlapping, the discriminant power of feature f is low.

Actually, the basic version of DP function doesn't make an estimation of the pdf of each class for a given feature f , but simply looks for the maximum and minimum sample values of feature f for each class over all trials. With these boundaries $max(s_{fk})$ and $min(s_{fk})$ for the k^{th} class and for feature f , the algorithm can then calculate the proportion of samples of feature f lying in the non-overlapping zones between boundaries of each class. For a two class problem, the formula of the discriminant power of feature f would be:

$$ND_{f1} = \sum_{i=1}^{N_{t1}} \left(l(s_{f1}(i) > max(s_{f2})) + l(s_{f1}(i) < min(s_{f2})) \right) \quad (2.4)$$

$$ND_{f2} = \sum_{i=1}^{N_{t2}} \left(l(s_{f2}(i) > max(s_{f1})) + l(s_{f2}(i) < min(s_{f1})) \right) \quad (2.5)$$

$$DP(f) = \frac{ND_{f1} + ND_{f2}}{N_{t1} + N_{t2}} \quad (2.6)$$

where N_{t1} and N_{t2} are the respective number of samples (or trials) for each class, s_{f1} and s_{f2} are vectors containing the samples of class 1 and 2 for feature f , ND_{f1} and ND_{f2} are the

⁷one of the filter methods is Relief [Kira and Rendell, 1992, Robnik-Sikonja and Kononenko, 2003], considered as one of the most efficient in the machine learning community.

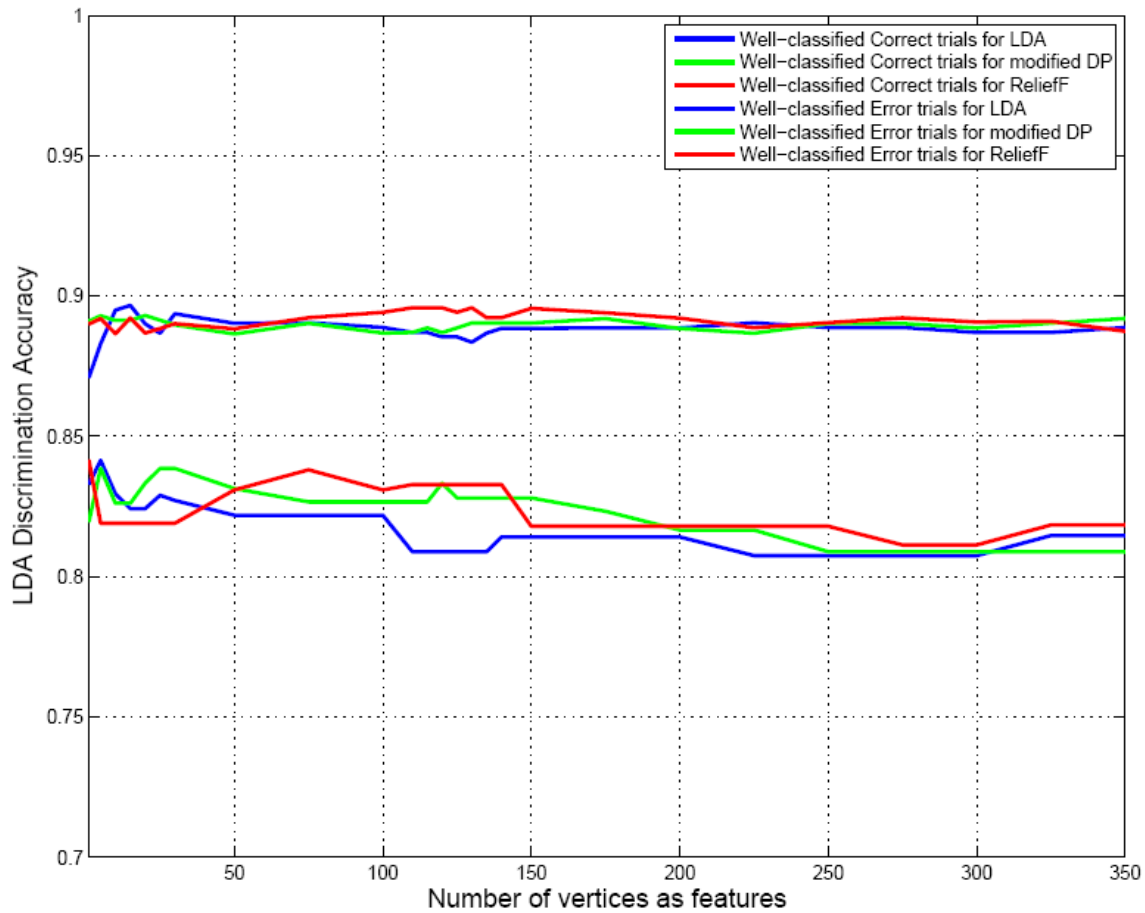


Figure 2.10: Comparison of different feature selection methods.

Classification rates of error and correct single trials using different feature selection methods and different numbers of selected features. The DP algorithm performs as well as other more complicated ones, and the best accuracies are reached using a small number of features.

number of discriminant samples of each class located in non-overlapping zones, and $l(x)$ is a function defined by:

$$l(x) = \begin{cases} 1 & \text{if } x \text{ is true} \\ 0 & \text{otherwise} \end{cases} \quad (2.7)$$

The basic DP algorithm returns a value $DP(f)$ between 0 and 1 for each feature. This value can be thought of as the discriminant power of the feature, since it is the proportion of discriminant samples over all trials. A graphical representation of the process of the basic DP function can be found in Figure 2.11. The score returned by the DP algorithm is the number of samples lying out of the grey shaded area divided by the total number of samples of both classes.

The advantage of this basic method is that the mathematic operations involved in the process are very simple, and thus quickly computed. This feature selection method is indeed very fast and practical for online applications of a learning process. However, the basic DP function has a major drawback, it is highly sensitive to noisy data. An example of this weakness is shown in

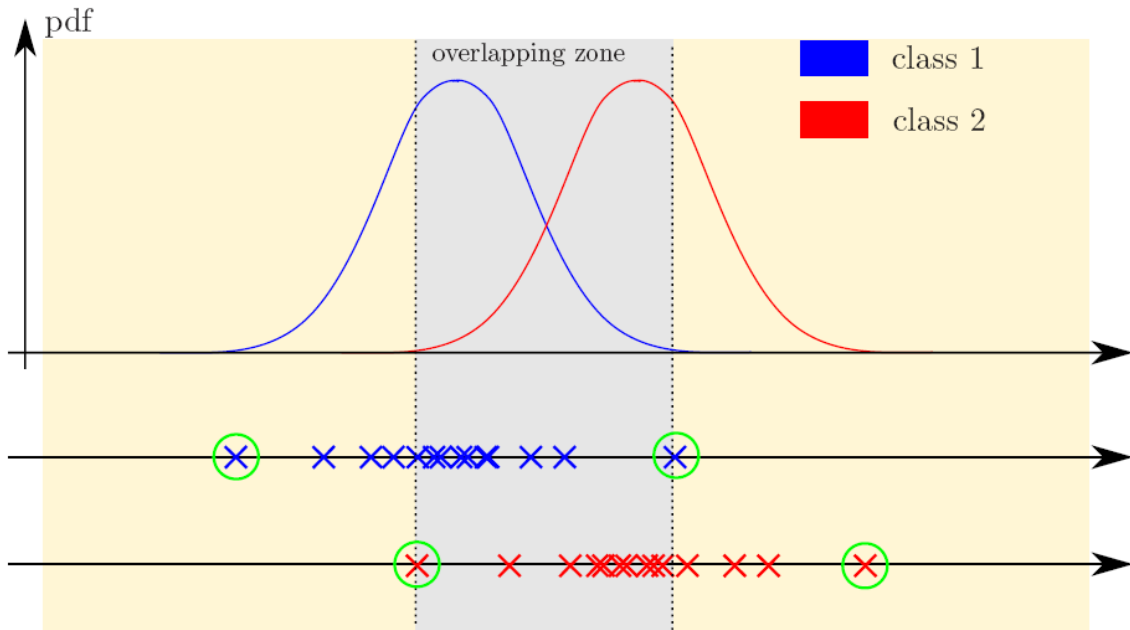


Figure 2.11: The basic DP algorithm.

The basic DP algorithm tries to estimate the separability power of a given feature by comparing the population distributions of this feature for the different classes. In this example, the score returned by the basic DP algorithm is the number of samples lying out of the grey shaded area, the overlapping area, divided by the total number of samples of both classes.

Figure 2.12, the distributions shown are relatively well separated, and should give a good result in terms of discriminant power. The only difference with Figure 2.11 is that one of the samples of class 1 is corrupted by noise, and can be considered as an outlier. The basic DP function will assume this outlier sample to be the maximum of class 1 distribution, and the resulting overlapping zone will entirely encompass the distribution of class 2, since the grey shaded area of Figure 2.12 is only defined by the extrema of class 2. Comparing this situation with Figure 2.11, it is straightforward to conclude that in this case, the resulting score of DP algorithm will not be representative of the non-overlapping property of the observed classes, even though the distributions of class 1 and 2 are not totally overlapping.

2.7.3.2 Modified DP algorithm

In order to cope with noisy data, a little preprocessing step has been added to the original DP algorithm: the distributions of both classes are first truncated, in order to keep only a given percentage of the data around the mean of each distribution. This pruning step will discard outlier data if the percentage of truncated data is well chosen. It is known that under the assumption of a normal distribution, and for a given integer k , a certain percentage of the values are within k standard deviations from the mean μ . Table 2.1 gives several different values of confidence intervals and the corresponding proportion of data within the interval, for normal distributions.

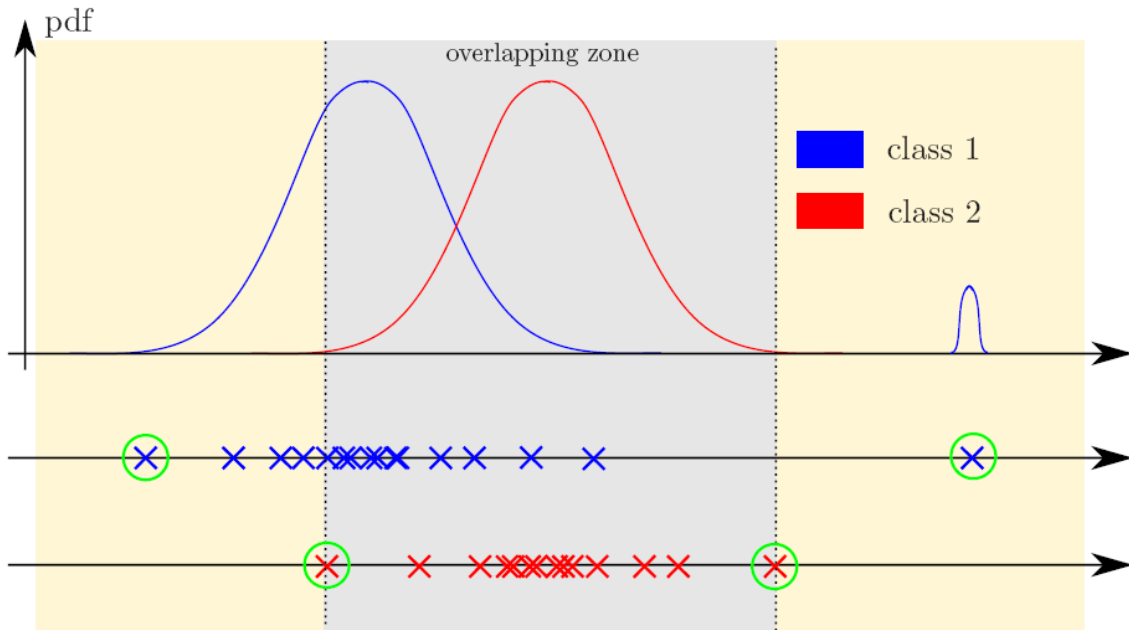


Figure 2.12: DP algorithm sensitivity to noise

Outliers strongly influence the results provided by the DP algorithm making this algorithm highly sensitive to noise.

Table 2.1: Confidence intervals.

Proportion of data within different confidence intervals for normal distributions.

Confidence Interval	% of data in the interval
$[\mu - \sigma; \mu + \sigma]$	68%
$[\mu - 1.177 \cdot \sigma; \mu + 1.177 \cdot \sigma]$	76%
$[\mu - 2 \cdot \sigma; \mu + 2 \cdot \sigma]$	95%
$[\mu - 3 \cdot \sigma; \mu + 3 \cdot \sigma]$	99%

If we are not sure of the normality of the distribution, a more general formula is provided by *Chebyshev's inequality*:

$$\Pr(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2} \quad (2.8)$$

Note that only the case $k > 1$ gives useful information. Thus, even if the distributions are not normal, at least $100 \cdot (1 - \frac{1}{k^2})\%$ of the values are within k standard deviations from the mean μ .

Keeping a too high percentage of the original data could maintain some outlier data in the pruned data set, and the DP score would remain meaningless. The issue is thus to assess which is the percentage of noisy data in a given application. In this sense, the *full width at half maximum* (FWHM) value is a good choice for an upper limit in the pruning process. The FWHM value is an expression of the extent of a function, given by the difference between the

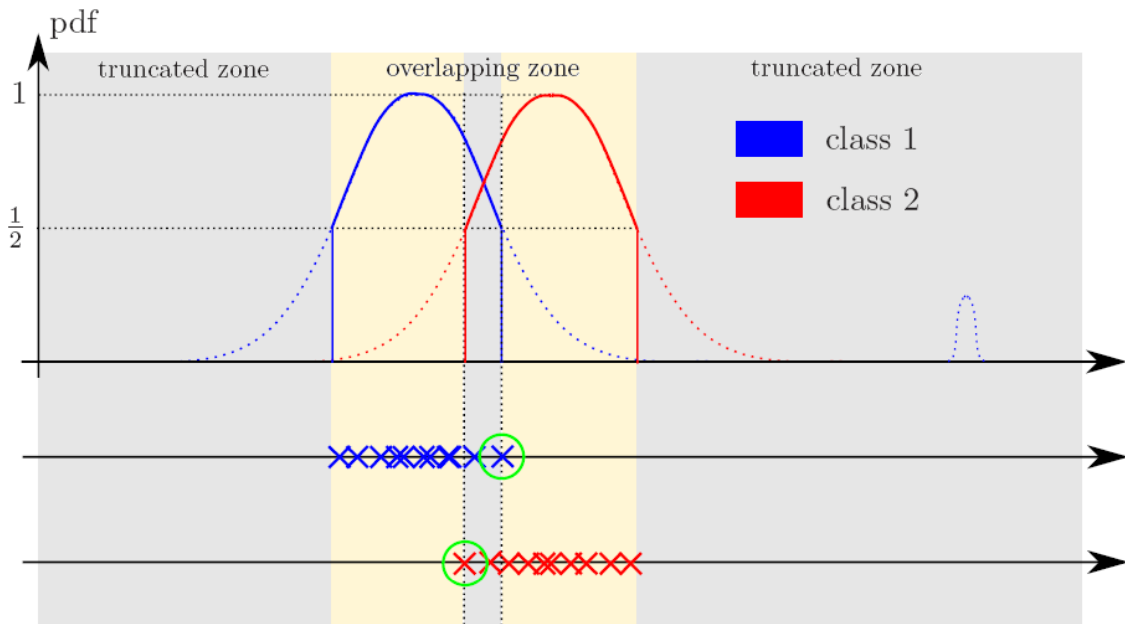


Figure 2.13: The modified DP algorithm.

Truncating the distributions to discard outliers reduces the sensitivity to noise and leads to better estimates of the discriminant power values.

two extreme values of the independent variable at which the dependent variable is equal to half of its maximum value. For a normal distribution, the relationship between FWHM and the standard deviation is:

$$\text{FWHM} = 2 \cdot \sqrt{2 \cdot \ln(2)} \cdot \sigma \approx 2.354 \cdot \sigma \quad (2.9)$$

and FWHM is the interval $\mu \pm \sqrt{2 \cdot \ln(2)} \cdot \sigma$, which means that we keep 76% of the samples after having truncated the distributions, according to normal distribution knowledge. In the implementation of the modified version of DP algorithm, we decided to approximate the distributions by normal distributions, and simply keep data in the interval $\mu \pm \sqrt{2 \cdot \ln(2)} \cdot \sigma = \mu \pm 1.177 \cdot \sigma$, assuming that the extrema are close to the mean μ . Of course, taking the interval $[\mu - 1.177 \cdot \sigma; \mu + 1.177 \cdot \sigma]$ as the threshold of our algorithm is arbitrary and gives a comfortable margin. In any specific application, if the percentage of noisy data is precisely known, the pruning threshold can be tuned so that only outlier data are discarded. Nevertheless, it is crucial to keep in mind that if the signal-to-noise ratio (SNR) is too low, the distributions will not be strictly normal anymore. In this case, the percentages given in table 2.1 are not correct any longer, and the pruning threshold will have to be chosen with care. This approximation of normal distributions is however acceptable for most applications, and the choice of FWHM as truncated intervals provides a good trade-off to remove noise without losing relevant information.

With this modified DP algorithm, the problematic situation of Figure 2.12 can be solved, as shown in Figure 2.13. If the percentage of noisy data is not too big, the outliers of class 1 will be discarded by the preprocessing step, and the remaining truncated distributions will reflect the real non-overlapping property of the classes. Indeed, we can see that the grey shaded

overlapping zone is defined by extreme values of both classes, and the resulting score given by the modified DP algorithm is thus meaningful in terms of discriminant power.

2.7.4 Statistical classifier

The different classes are recognized by a Gaussian classifier trained to classify single trials as one of the classes. The output of this statistical classifier is an estimation of the posterior class probability distribution for a single trial; i.e., the probability that a given single trial belongs to one of the classes. The present classifier is quite similar to that proposed by Millán *et al.* [Millán *et al.*, 2002b, 2004a] in the framework of EEG-based BCIs, the main difference being the update rules.

In this statistical classifier, every Gaussian unit represents a prototype of one of the classes to be recognized. We use several prototypes per mental task. We assume that the class-conditional probability density function of class C_k is a superposition of N_k Gaussians (or prototypes) and that classes have equal prior probabilities. In our case, all the classes have the same number of prototypes N_p . In addition, we assume that all prototypes have an equal weight of $1/N_p$. Then, dropping constant terms, the activity a_k^i of the i^{th} prototype of the class C_k for the input vector, or sample, x derived from a trial as described above is

$$a_k^i(x) = |\Sigma_k|^{-1/2} \exp \left(-1/2 (x - \mu_k^i)^T \Sigma_k^{-1} (x - \mu_k^i) \right) \quad (2.10)$$

where μ_k^i is the center of the i^{th} prototype of the class C_k , Σ_k is the covariance matrix of the class C_k , and $|\Sigma_k|$ is the determinant of that matrix. Usually, each prototype has its own covariance matrix Σ_k^i . In order to reduce the number of parameters, we restrict our model to a diagonal covariance matrix Σ_k that is common to all the prototypes of the class C_k . Now, the posterior probability y_k of the class C_k is

$$y_k(x) = p(x|C_k) = \frac{a_k(x)}{A(x)} = \frac{\sum_{i=1}^{N_p} a_k^i(x)}{\sum_{k=1}^{N_c} \sum_{i=1}^{N_p} a_k^i(x)} \quad (2.11)$$

where a_k is the activity of class C_k , A is the total activity of the network and N_c is the number of classes. The response of the classifier for the input vector x is simply the class with the highest probability.

To initialize the center of the prototypes, μ_k^i , of the class C_k we run a clustering algorithm—typically, self-organizing maps [Kohonen, 1997]. We then initialize the diagonal covariance matrix Σ_k of the class C_k by setting

$$\Sigma_k = \frac{1}{|S_k|} \sum_{x \in S_k} (x - \mu_k^{i*}) (x - \mu_k^{i*})^T \quad (2.12)$$

where S_k is the set of the training samples belonging to the class C_k , $|S_k|$ is the cardinality of this set, and i^* is the nearest prototype of this class to the sample x .

During learning we improve these initial estimations iteratively by stochastic gradient descent so as to minimize the mean square error $E = 1/2 \sum_k (y_k - t_k)^2$, where t_k is the k^{th}

component of the target vector in the form 1-of-c. Taking the gradient of the error function yields

$$\mu_k^i(x)_{new} = \mu_k^i(x) + \alpha \frac{\partial E}{\partial \mu_k^i}(x) = \mu_k^i(x) + \alpha \frac{a_k^i(x)}{A(x)} \frac{(x - \mu_k^i)}{\Sigma_k^i} e_k(x) \quad (2.13)$$

and

$$\Sigma_k^i(x)_{new} = \Sigma_k^i(x) \exp\left(\beta \frac{\partial E}{\partial \Sigma_k^i}(x)\right) = \Sigma_k^i(x) \exp\left(\beta \frac{a_k^i(x)}{A(x)} \frac{\left((x - \mu_k^i)^2 - \Sigma_k^i\right)}{(\Sigma_k^i)^2} e_k(x)\right) \quad (2.14)$$

where α and β are the learning rates and

$$e_k(x) = (t_k(x) - y_k(x)) - \sum_j y_j(x) (t_j(x) - y_j(x)) \quad (2.15)$$

An exponential rule is preferable for the covariance in order to avoid negative values. After updating μ_k^i and Σ_k^i for each training sample, the covariance matrices of all the prototypes of the same class are averaged to obtain the common class covariance matrix Σ_k . This simple operation leads to better performance than if separate covariance matrices are kept for each individual prototype. It can also be sometimes preferable to have a single covariance matrix common to all classes, which is obtained by averaging all the individual Σ_k , but this is case dependant.

More practically, the different hyper-parameters—that is, the learning rates of the center and diagonal covariance matrices, number of prototypes, and common/single covariance matrices for each class—were chosen by model selection in the training sets. Regarding the learning rates, usual values for the centers were 10-4 to 10-6 but no larger than 10-3 and no smaller than 10-8. For the covariances usual values were 10-6 to 10-8 but no larger than 10-4 and no smaller than 10-10. Finally, the usual number of prototypes was between 2 and 4 but never exceeded 6 and usually, using a single common covariance matrix for all classes leads to better classification rates.

2.7.5 Measuring performance: The bit rate

A traditional measure of the performance of a system is the bit rate, the amount of information communicated per unit time. The bit rate is usually expressed in bit per trial (bit per selection) or in bit per minute. The bit rate depends on both speed and accuracy.

If a single trial has N_c possible selection, if the probability p that the desired selection is actually the one selected (accuracy of the BCI) and if finally each of the other selections has the same probability of selection (i.e. $(1 - p)/(N_c - 1)$), then the information transfer rate in bits per trial is determined as following [Shannon, 1948]:

$$Bits/Trial = \log_2(N_c) + p \cdot \log_2(p) + (1 - p) \cdot \log_2\left(\frac{1 - p}{N_c - 1}\right) \quad (2.16)$$

This formula makes the assumption that BCI errors and ErrP detection errors are independent, which might not always be the case in particular situations like lack of concentration,

longer lasting artifacts, or fatigue. If t is the mean time in seconds between two consecutive selections, the information transfer rate in bits per minute becomes:

$$\text{Bits/Minutes} = \text{Bits/Trials} \cdot \left(\frac{60}{t}\right) \quad (2.17)$$

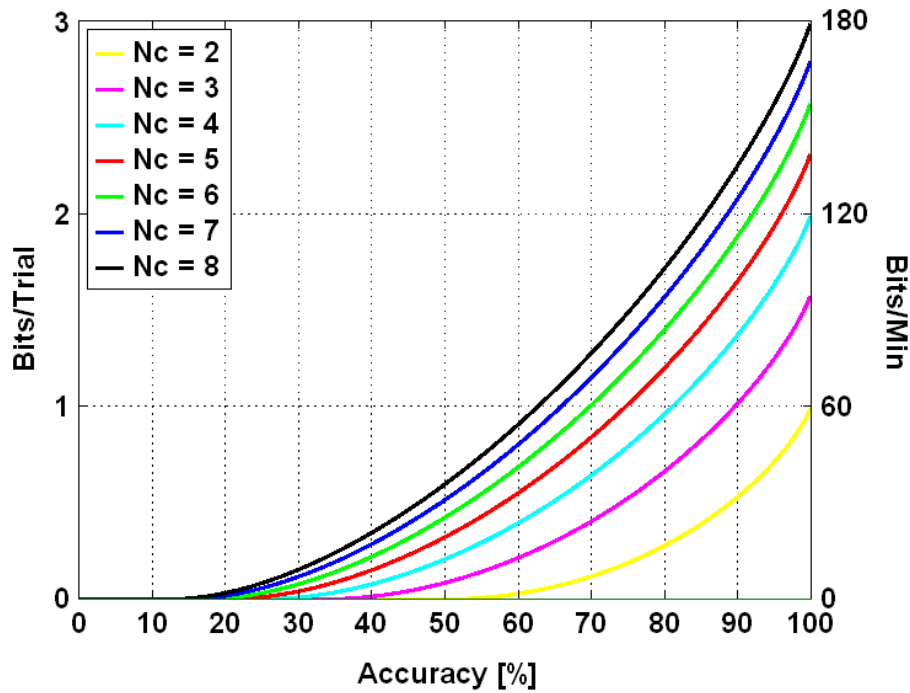


Figure 2.14: Information transfer rate.

The bit rate is given in bits per trial and in bits per minute (for a mean time between 2 consecutive selections of 1 second) for different number of selection N_c .

Figure 2.14 shows the bit rate in bits/trial and in bits/minutes (for a response time of 1 second) for different numbers of possible selections N_c . Let's consider now how the performance of the BCI changes after introducing ErrP and that the system detects a proportion e of erroneous selections and a proportion c of correct selections. In the general case, after detecting an erroneous trial the outcome of the interface is simply stopped and not sent to the brain-actuated device. The proportion of effectively transmitted p_t selections becomes:

$$p_t = p \cdot c + (1 - p) \cdot (1 - e) \quad (2.18)$$

The new accuracy p' of the BCI becomes:

$$p' = \frac{p \cdot c}{p_t} \quad (2.19)$$

and taking into account the proportion of stopped selections, the new information transfer rate in bits per trial becomes:

$$Bits/Trial = p_t \cdot \left(\log_2(N_c) + p' \cdot \log_2(p') + (1 - p') \cdot \log_2\left(\frac{1 - p'}{N_c - 1}\right) \right) \quad (2.20)$$

In the case of a two-class BCI ($N_c = 2$), after detecting an erroneous trial, it could be possible to replace the *wrong* outcome by the opposite one, what yields a new accuracy p'' :

$$p'' = p \cdot c + (1 - p) \cdot e \quad (2.21)$$

The information transfer rate in this case is calculated by replacing p by p'' in Equation 2.16, because now there is no stopped outcome.

2.8 The MAIA project

The goal of the EU project called MAIA⁸ (Mental Augmentation through determination of Intended Action) is to develop non-invasive brain-computer interfaces that recognize the subject's voluntary imagination of motor actions and transmits this intention to a device that performs the necessary low-level steps to achieve complex tasks. This project is founded under the Information Society Technologies (IST) programme of the Sixth Framework Programme (FP6), it started in September 2004 and will end in December 2007. The partners are the IDIAP Research Institute (coordinator, Switzerland), the Katholieke Universiteit Leuven (Belgium), The University Hospital of Geneva (Switzerland), the Fondazione Santa Lucia in Rome (Italy) and the Helsinki University of Technology (Finland). The main innovative principles are the use of estimates of intracranial activity from scalp EEG for the recognition of the subject's motor intent, the adaptive shared autonomy between the human and the robot, the use of vibro-tactile feedback to speed up training of the subject, the integration of high-level cognitive states such as errors to increase the reliability of the interface and the online adaptation of the interface to the subject to constantly track the changes of brain activity. The main achievement of the MAIA project so far was the demonstration in 2006 of the first brain-actuated wheelchair (shown in Figure 2.15) controlled by a subject sitting on it in the laboratories of the Katholieke Universiteit Leuven. In this case, we have incorporated shared control principles into the BCI [Philips et al., 2007, Vanacker et al., 2007]. In shared control, the intelligent controller relieves the human from low level tasks without sacrificing the cognitive superiority and adaptability of human beings that are capable of acting in unforeseen situations. Although our first brain-actuated robot had already some form of cooperative control, shared autonomy is a more principled and flexible framework.

⁸<http://www.maia-project.org>



Figure 2.15: Brain-actuated wheelchair.

Subject driving the wheelchair in a natural environment from non-invasive EEG. Note the laser scanner in front of the wheelchair, in between the subject's legs.

Chapter 3

Preliminary results

This chapter summarizes preliminary results on error-related potentials obtained prior to the investigation of the use of error potentials to improve the performance of the BCI. The goal of these first experiments is to validate our system, both hardware and software (general data processing), by repeating previously reported protocols to see if the results we obtain match those available in the literature. The two main kind of error potentials are response ErrP (choice reaction tasks) and feedback ErrP (reinforcement learning tasks). Therefore we decide to replicate one protocol of each kind. However, most studies about error potentials describe grand averages and try to elaborate theories about the mechanisms underlying ErrP generation. The following experiments were conducted in a BCI framework, i.e., apart from validating our system, the main goal is to test the feasibility of recognizing ErrP at the level of single trials, and no more in grand averages.

3.1 Choice reaction task

The goal of the first set of experiments was to explore the feasibility of detecting traditional ErrP within the framework of a typical choice reaction task. A healthy volunteer subject was asked to respond as quickly as possible to a stimulus delivered by the system. The stimulus consisted in one of the words: *BLUE*, *RED* and *GREEN*, and each word could be written either in blue, red or green. The subject was asked to press a key with his left hand if the word was written in blue, another key with his right hand if the word was written in red and both keys with the two hands if the word was written in green. This means that the word itself had no importance, only its color counts. In this first protocol, ErrP arose following an erroneous response made by the subject. The subject *reacts* to his own errors. This protocol is very similar to that used by Vidal et al. in their experiment I [Vidal et al., 2000].

At each trial, the selected word was picked up randomly. Then, the probabilities that a word was written in the matching color or in one of the 2 non-matching colors are equal (50%-25%-25%). No feedback such as reaction time or error signals were provided. The subject performed 10 series of 60 trials, the delay between 2 consecutive trials was of 1.5-2 seconds (random delay to prevent habituation). This protocol ensured a sufficient number of erroneous responses, the mean error rate was of $\sim 15\%$ of the trials.

3.1.1 Grand averages

Figure 3.1 shows the averages of error and correct trials plus the difference error-minus-correct for channels Fz. The Ne and Pe can clearly be seen in this figure. Ne shows up about 40 ms after the response (response occurs at time = 0) as a sharp negative peak and Pe shows up about 210 ms after the response as a broader positive peak, as expected. Indeed, many studies report these two peaks (Ne and Pe) as the main response ErrP components [Falkenstein et al., 2000]. In the top right corner of Figure 3.1, we show the potentials reported by Falkenstein et al. (2000) in a similar protocol. Our results match relatively well, only Pe occurs earlier in our case (200 ms vs. 300 ms). Figure 3.1 also shows the scalp potentials topographies 40 ms (occurrence of Ne) and 210 ms (occurrence of Pe) after the response. Both peaks are mainly located in a fronto-central area, this seems to be a first confirmation that the ACC plays a role in ErrP generation.

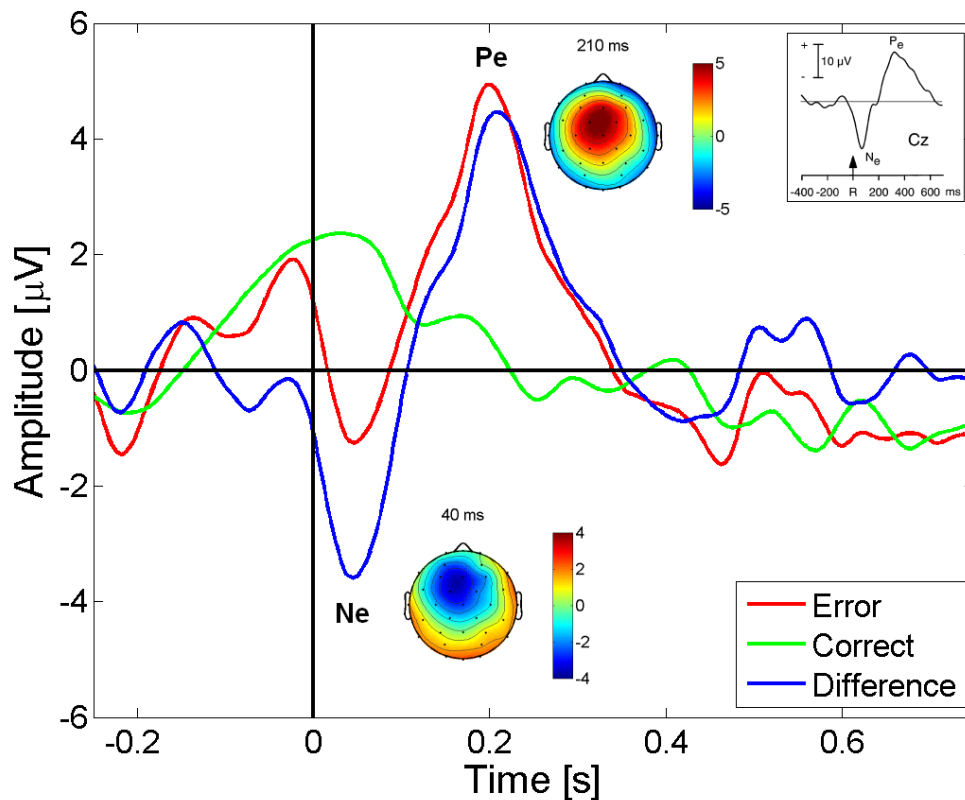


Figure 3.1: Response ErrP.

Averages of error and correct trials plus the difference error-minus-correct for channels Fz (one subject). Ne and Pe show up about 40 ms and 210 ms after the response, respectively and both show fronto-central foci. The reported potentials are very similar to those reported by Falkenstein et al. in 2000 (top right corner).

3.1.2 Source localization

As already mentioned, ErrP are most probably generated in a deep frontal cortical area called anterior cingulate cortex (ACC). To test this hypothesis, the localization software sLORETA was used to isolate the foci of activity at the occurrence of the two peaks. Figure 3.2 shows Talairach slices of localized activity for the grand average at the occurrence of the two described peaks. For the negative peak, the focus of activity is located at $X=-5\text{mm}$, $Y=-10\text{mm}$, $Z=30\text{mm}$ (MNI coordinates). The best match is Brodmann area 24 (ventral anterior cingulate cortex). For the positive peak, the focus of activity is located at $X=5\text{mm}$, $Y=25\text{mm}$, $Z=15\text{mm}$. The best match here is also Brodmann area 24, and the second best match is Brodmann area 32 (dorsal anterior cingulate cortex). These localization for ErrP sources are conform to expectations [Holroyd and Coles, 2002, Fiehler et al., 2004].

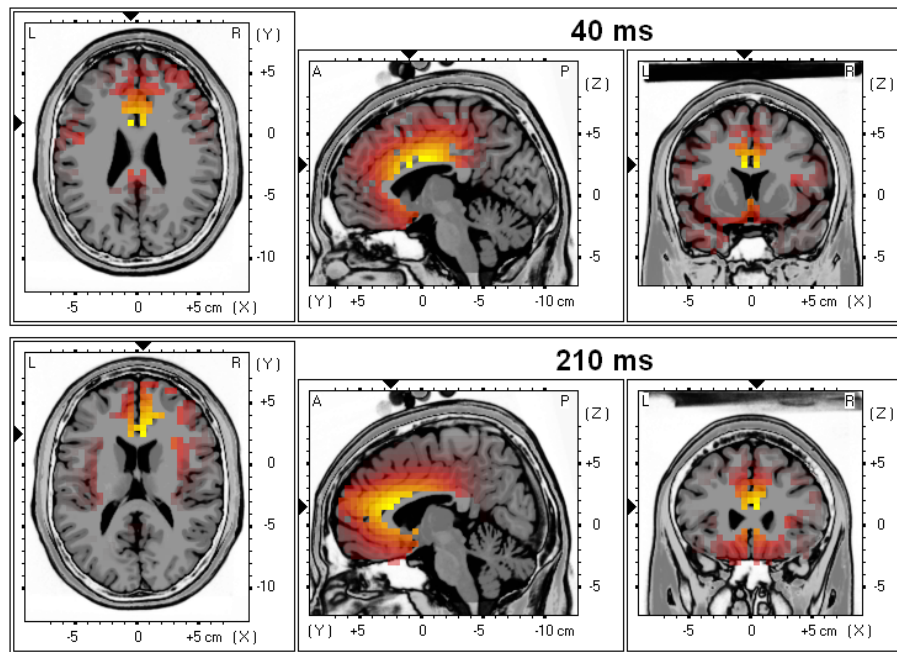


Figure 3.2: Localization of response ErrP.

Talairach slices of localized activity for the grand average at the occurrence of the 2 described peaks. The anterior cingulate cortex, both ventral and dorsal parts, is systematically activated.

3.1.3 Classification

To explore whether it is possible to recognize single-trial erroneous responses, we have done a 10-fold cross-validation study where the testing set consists of one of the recorded sessions using the statistical classifier described in Section 2.7.4. In this way, testing is always done on a different recording session to those used for training the model. This yields a better estimation of the actual generalization capabilities because brain activity naturally changes over time.

EEG potentials were acquired using our BioSemi system (for details see Section 2.7.1) using 32 electrodes and raw EEG signals were processed according to the procedure explained

in Section 2.7.2. For each single trial, the input to the classifier was a vector with the potentials in a one second window starting 250 ms before the response and ending 750 ms after the response for channels Cz and Fz. No systematic feature selection was performed for this preliminary study. The selection of channels Cz and Fz is based on the fact that the ACC has a median fronto-central localization and more importantly on the fact that, according to the scalp topographies of Figure 3.1, channels Cz and Fz are among the electrodes showing the highest ErrP amplitudes. The percentages of correctly recognized single trials were the following:

$$\begin{aligned} 77.8\% \pm 12.6 & \quad \text{for error trials} \\ 88.4\% \pm 5.4 & \quad \text{for correct trials} \end{aligned}$$

Even if it is based on a single subject, this first experiment not only clearly confirms the presence of specific ErrP following erroneous response but also demonstrates the feasibility of differentiating erroneous and correct trials. Furthermore, it seems to confirm that the ACC plays an important role in the generation of ErrP.

3.2 Reinforcement learning task

For the validation using a reinforcement learning task, the same healthy volunteer subject was asked to imagine that the four circles positioned on a row (O O O O) appearing on each trial were balloons, and that one of the balloons contained 1\$. The subject had to chose a balloon and had to try to collect as much money as possible.

The selection of a balloon by the subject was followed either by a positive feedback (four dollar signs in a row \$ \$ \$ \$) or a negative feedback (four Xs in a row X X X X). With this protocol, ErrP are also elicited by errors committed by the subject himself, the main difference is that the subject is not automatically aware of his error, the feedback of the system is the element triggering ErrP. This protocol is the one used by Holroyd et al. in their experiment [Holroyd et al., 2003]. At each trial, there was a 25% probability that the balloon contains no money, so that the average error rate was 25%. The subject performed 10 series of 60 trials, the delay between 2 consecutive trials was of 1.5-2 seconds (random delay to prevent habituation).

3.2.1 Grand averages

Figure 3.3 shows the averages of error and correct trials plus the difference error-minus-correct for channels FCz. As expected, a negative deflection can be seen 250 ms after the feedback for error trials [Holroyd and Coles, 2002]. A second positive component can be seen about 330 ms after the feedback. In the top right corner of Figure 3.3 we show the potentials reported by Holroyd and Coles (2002). Our potentials are very similar in timing and amplitudes, note the negative potentials for both conditions 100 ms after the feedback. Scalp potentials topographies at 250 and 330 ms are also shown on Figure 3.3. The negative peak seems to cover the whole right hemisphere whereas the positive peak covers a rather centro-parietal area. This centro-parietal localization could indicate that the 330 ms positive peak has a P300 component. Indeed, the peak occurs at ~ 300 ms and P300 are known to usually have a parietal focus.

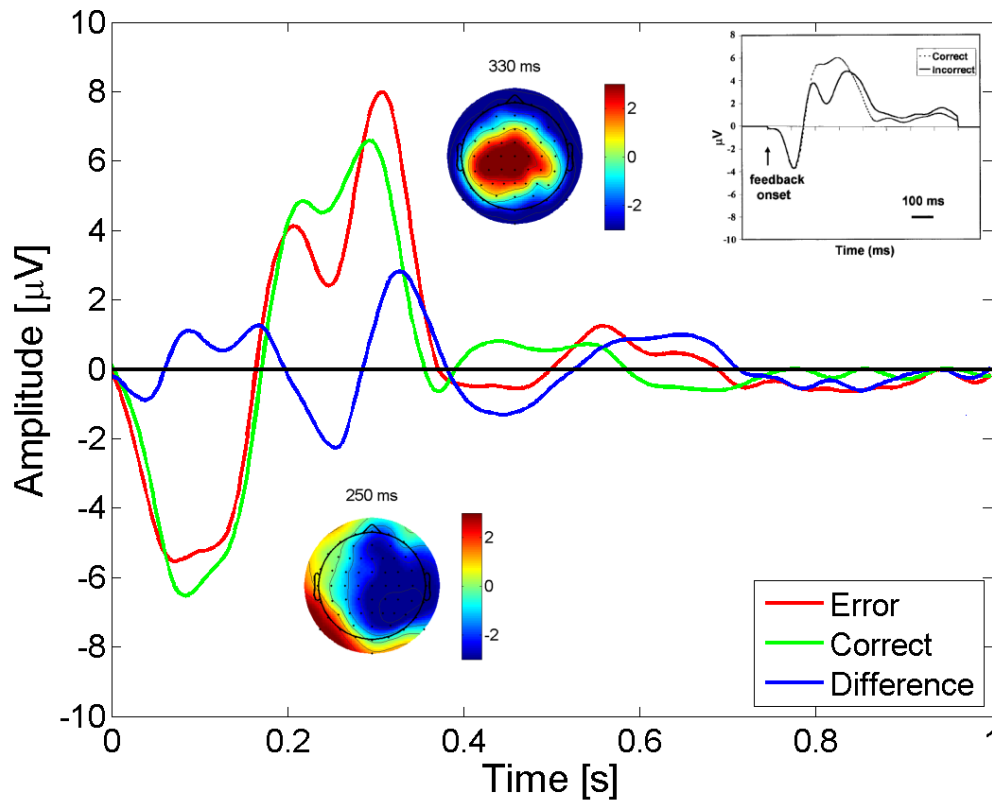


Figure 3.3: Feedback ErrP.

Averages of error and correct trials plus the difference error-minus-correct for channels FCz (one subject). An expected negative deflection can be seen 250ms after the feedback and a later positive peak shows up 330 ms after the feedback. The negative peak seems to cover the whole right hemisphere whereas the positive peak covers a centro-parietal area. The reported potentials are very similar to those reported by Holroyd and Coles in 2002 (top right corner).

3.2.2 Source localization

The localization software sLORETA was again used to isolate the foci of activity at the occurrence of the two peaks. Figure 3.4 shows Talairach slices of localized activity for the grand average at the occurrence of the two described peaks. For the negative deflection, the focus of activity is located at X=10mm, Y=-75mm, Z=-5mm (MNI coordinates). The best match is Brodmann area 18 (visual association cortex), the visual activity is paramount in this case. For the positive peak, the focus of activity is located at X=0mm, Y=-10mm, Z=35mm. The best match is Brodmann area 24, the ventral anterior cingulate cortex. However, parietal areas are also activated for the positive peak, confirming the hypothesis of a P300 component.

3.2.3 Classification

As for the choice reaction task experiment, we have done a 10-fold cross-validation study where the testing set consists of one of the recorded sessions using the statistical classifier introduced in Section 2.7.4. EEG potentials were this time acquired using 64 electrodes and

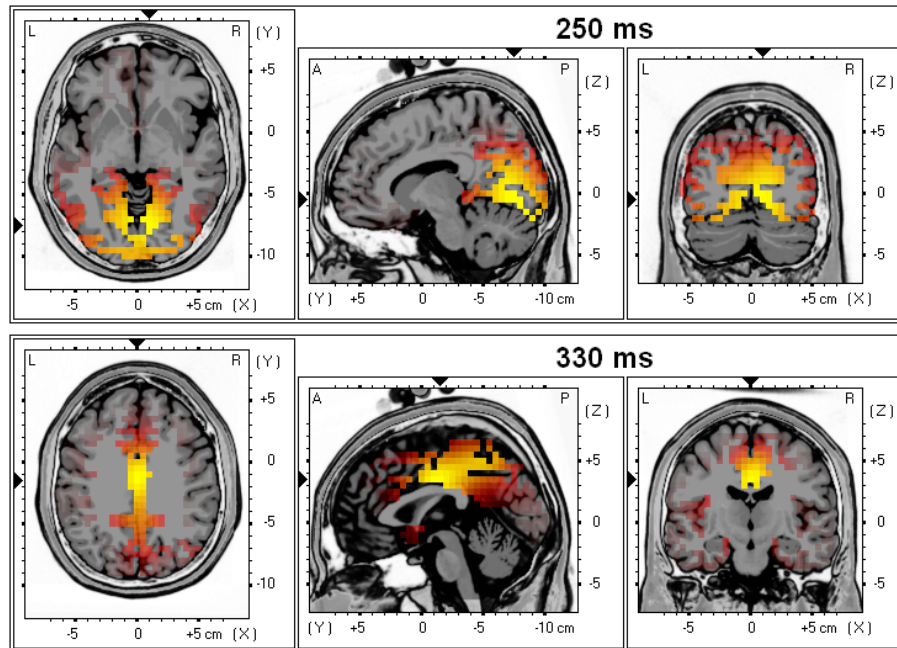


Figure 3.4: Localization of feedback ErrP.

Talairach slices of localized activity for the grand average at the occurrence of the 2 described peaks. For the 250ms negative deflection, visual activity seems to be predominant, whereas for the positive peak (330ms), activity is focused in the anterior cingulate cortex.

the data pre-processing remained the same (cf. Section 2.7.2). For each single trial, the input to the classifier was a vector with the potentials in a 300 ms window starting 200 ms after the feedback for channels Cz and Fz. Again no feature selection was performed for this preliminary analysis, electrodes Cz and Fz were chosen for comparison purposes with the results obtained in the choice reaction task experiment. The percentages of correctly recognized single trials were the following:

$$\begin{aligned} 64.3\% \pm 11.2 & \quad \text{for error trials} \\ 67.2\% \pm 12.3 & \quad \text{for correct trials} \end{aligned}$$

The results obtained with this second experiment are not as encouraging as the results obtained with a choice reaction task. First the potentials are not so clear, they only show low amplitude, but more importantly the classification results are not high above random level. Finally only the second component (positive peak 330 ms after the feedback) has a focus in the ACC. However, since these results were obtained with a single subject, they have to be taken with care.

Nevertheless, the goal of these first small experiments was mainly the validation of our system. For both protocols, the results in terms of average potentials are very similar to those available in the literature. Our validation goal can therefore be considered as achieved.

Chapter 4

First attempt: Progress bars

In this chapter, we report the first attempt of recording error-related potentials elicited no more by an error made by the subject himself, but rather following an error of the interface when executing a command given by the subject. Part of the results reported in this chapter were presented at the 19th International Joint Conference on Artificial Intelligence (IJCAI) in Edinburgh (UK) in 2005 [Ferrez and Millán, 2005]. A throughout description can be found in the *Toward Brain-Computer Interfacing* MIT book [Ferrez and Millán, 2007]. Finally a brief summary of these results can be found in [Buttfield et al., 2006].

4.1 Experimental setup

To test the presence of ErrP after a feedback indicating errors made by the interface in the recognition of the subjects intent, we have simulated a real interaction with a robot where the subject wishes to bring the robot to one side of a room (left or right) by delivering repetitive commands until the robot reaches the target. This virtual interaction is implemented by means of two horizontal progress bars made of ten steps each. One of the bars goes from the center to the left side (left bar), and the other bar progresses to the right side (right bar). Figure 4.1 shows the left and right horizontal progress bars used as feedback.



Figure 4.1: First graphical interface: Progress bars.

The subject delivers repetitive commands until the bar he selected is full. The system fills the bars with an error rate of 20%. This simulates a *noisy* displacement of a robot to one side of the room (left or right).

To isolate the issue of the recognition of ErrP from the more difficult and general problem of a whole BCI where erroneous feedback can be due to non-optimal performance of both the interface (i.e., the classifier embedded into the interface) and the users themselves, in the following experiments the subjects deliver commands manually and not mentally. That is, they simply press a left or right key with the left or right hand. In this way, any error feedback

is due only to a wrong recognition by the interface of the subjects intention. Four healthy volunteer subjects participated in these experiments. The subjects press a key after a stimulus delivered by the system (the word **GO** appears on the screen). After a one second delay to avoid movement-related artifacts, the system filled the bars with an error rate of 20%; that is, at each step, there was a 20% probability that the incorrect progress bar was filled. Subjects performed ten series of five progress bars, the delay between two consecutive steps (two consecutive GOs from the system) was between 3 and 4 seconds (random delay to prevent habituation). Duration of each interaction experiment (i.e., filling a progress bar) was about 40 seconds, with breaks of a few minutes between two series but no break between interaction experiments of the same series.

EEG potentials were acquired using 32 electrodes and raw EEG data were pre-processed according to the procedure described in Section 2.7.2. The input vector for the statistical classifier (cf. Section 2.7.4) is a half a second window starting 150 ms after the feedback and ending 650 ms after the feedback for channels Cz and Fz. The choice of these channels follows the fact that ErrP are characterized by a median fronto-central distribution. Thus, the dimensionality of the input vector is 64 that is, concatenation of two windows of 32 points (EEG potentials) each. No feature selection procedure was performed for the selection of channel Cz and Fz. However, and this will be discussed later in this chapter, in next chapter we introduce a new protocol that is more realistic and that avoids several drawbacks. This new protocol has been used for all the other analysis presented later in this thesis, in particular for the real-time validation of ErrP detection. In next chapter we present detailed feature selection results that determine the best choice of electrodes for ErrP classification.

No artifact rejection algorithm (for removing or filtering out eye or muscular movements) was applied and all trials were kept for analysis. It is worth noting, however, that after a visual a posteriori check of the trials, we found no evidence of muscular artifacts that could have contaminated one condition differently from the other.

4.2 Grand averages

With this protocol, it is first necessary to investigate whether or not ErrP are present no more in reaction to errors made by the subjects themselves, but in reaction to erroneous responses made by the interface as indicated by the feedback visualizing the recognized subjects intentions.

Figure 4.2 shows the average of error and correct trials plus the difference error-minus-correct for channel Cz for the four subjects and Figure 4.3 shows the same plots for the average of the four subjects. Feedback is delivered at time 0 seconds. A first sharp negative peak can clearly be seen 270 ms after the feedback. A later positive peak appears about 400 ms after the feedback. Finally, an additional negative peak occurs about 550 ms after the feedback. All four subjects show slightly different ErrP time courses and amplitudes, but all subjects exhibit the main ErrP components previously described, only subject 2 shows no negative peak around 270 ms after the feedback.

Figure 4.3 also shows the scalp potentials topographies for the average of the four subjects at the occurrence of the three described peaks. A first fronto-central negativity appears 270 ms after the feedback. A large fronto-centro-parietal positivity can be seen 400 ms after the feedback, followed by a fronto-central negativity at 550 ms.

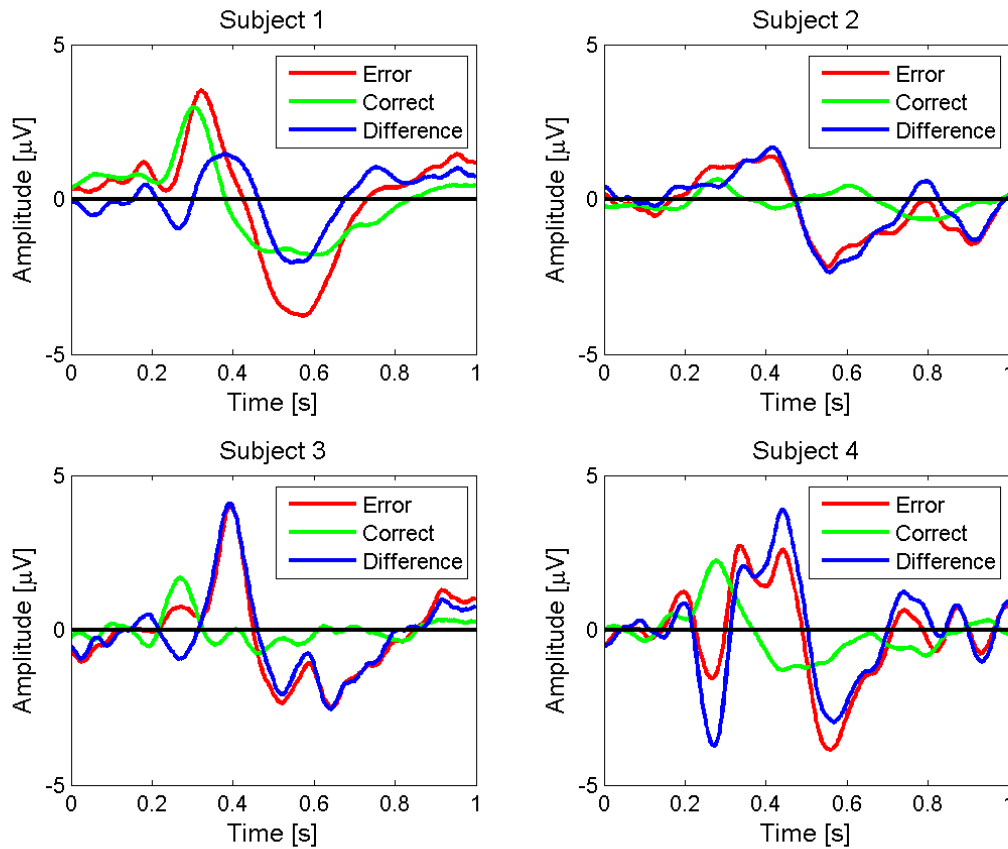


Figure 4.2: Interaction ErrP, 4 subjects.

Average of error and correct trials plus the difference error-minus-correct for channel Cz for the four subjects. Feedback is delivered at time 0 seconds. All four subjects show slightly different ErrP time courses and amplitudes, but all subjects exhibit the main ErrP components, only subject 2 shows no negative peak around 270 ms after the feedback.

These experiments seem to reveal a new kind of error-related potentials that, for convenience, we call *interaction ErrP*. The general shape of these ErrP is quite similar to the shape of the response ErrP in a choice reaction task, whereas the timing is similar to the feedback ErrP of reinforcement learning tasks and to the observation ErrP. As in the case of response ErrP, interaction ErrP exhibit a first sharp negative peak followed by a broader positive peak. However, interaction ErrP are also characterized by a second negative peak that does not appear in response ErrP. This is quite different from the shape of feedback ErrP and observation ErrP that are only characterized by a small negative deflection. On the other hand, the time course of the interaction ErrP bears some similarities to that of the feedback ErrP and observation ErrP: in both cases a distinctive feature (negative peak and negative deflection, respectively) appears ~ 250 ms after feedback. This delay represents the time required by the subject to process the visual feedback. The time course of response ErrP is definitely different. The peaks show up much faster because the subject's proprioception makes him/her aware of his/her erroneous action before he/she presses the wrong key.

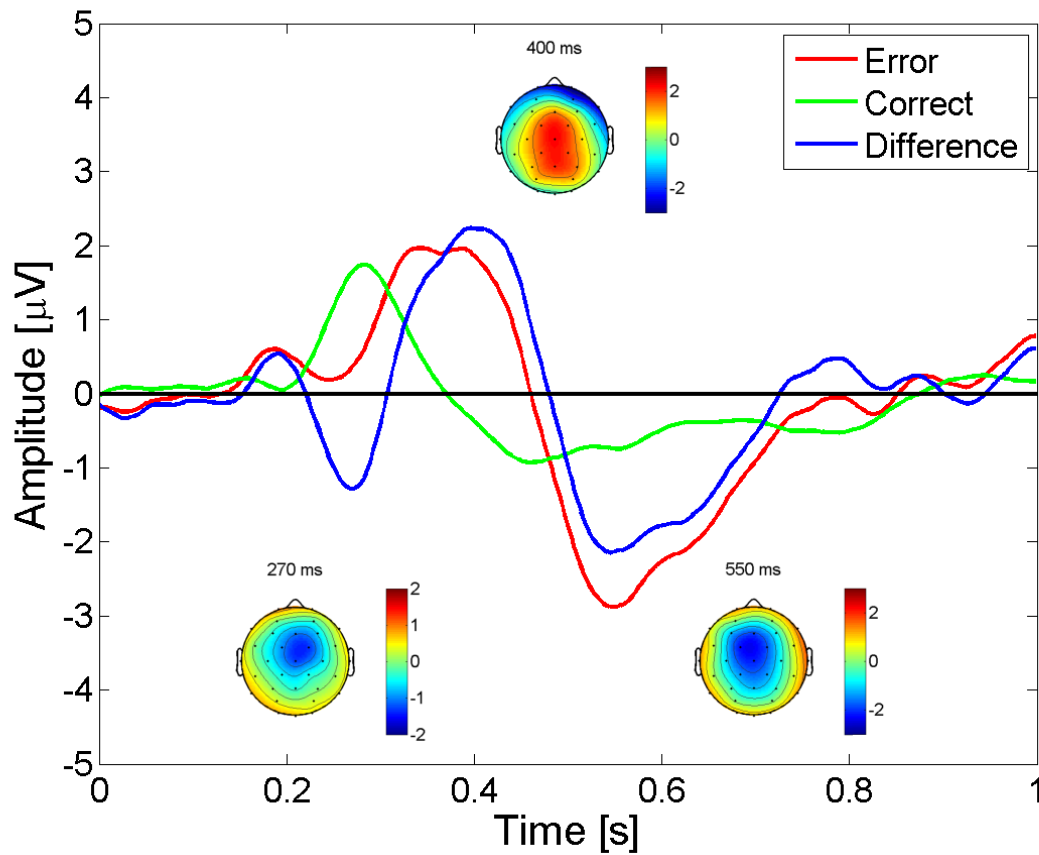


Figure 4.3: Interaction ErrP, average.

Average of error and correct trials plus the difference error-minus-correct for channel Cz and scalp potentials topographies at the occurrence of the three main peaks for the average of the four subjects. Feedback is delivered at time 0 seconds. Small filled circles indicate positions of the electrodes (frontal on top), Cz being in the middle of the scalp.

4.3 Estimation of intracranial activity

The sLORETA software described in Section 2.4.3.2 was used as a localization tool to estimate the focus of intracranial activity at the occurrence of the three ErrP peaks described in the previous section. Figure 4.4 shows Talairach slices of localized activity for the grand average of the four subjects at the occurrence of the three peaks. For the first negative peak (270 ms after the feedback), the focus of activity is located at X=0mm, Y=45mm, Z=0mm (MNI coordinates). The best match is Brodmann area 32 (dorsal anterior cingulate cortex). For the positive peak (400 ms after the feedback), the focus of activity is located at X=10mm, Y=45mm, Z=10mm. Again the best match is Brodmann area 32. Finally, for the second negative peak (550 ms after the feedback), the focus of activity is located at X=5mm, Y=5mm, Z=35mm. In this case the best match is Brodmann area 24, the ventral anterior cingulate cortex. As expected, the anterior cingulate cortex (ACC) is systematically activated [Holroyd and Coles, 2002].

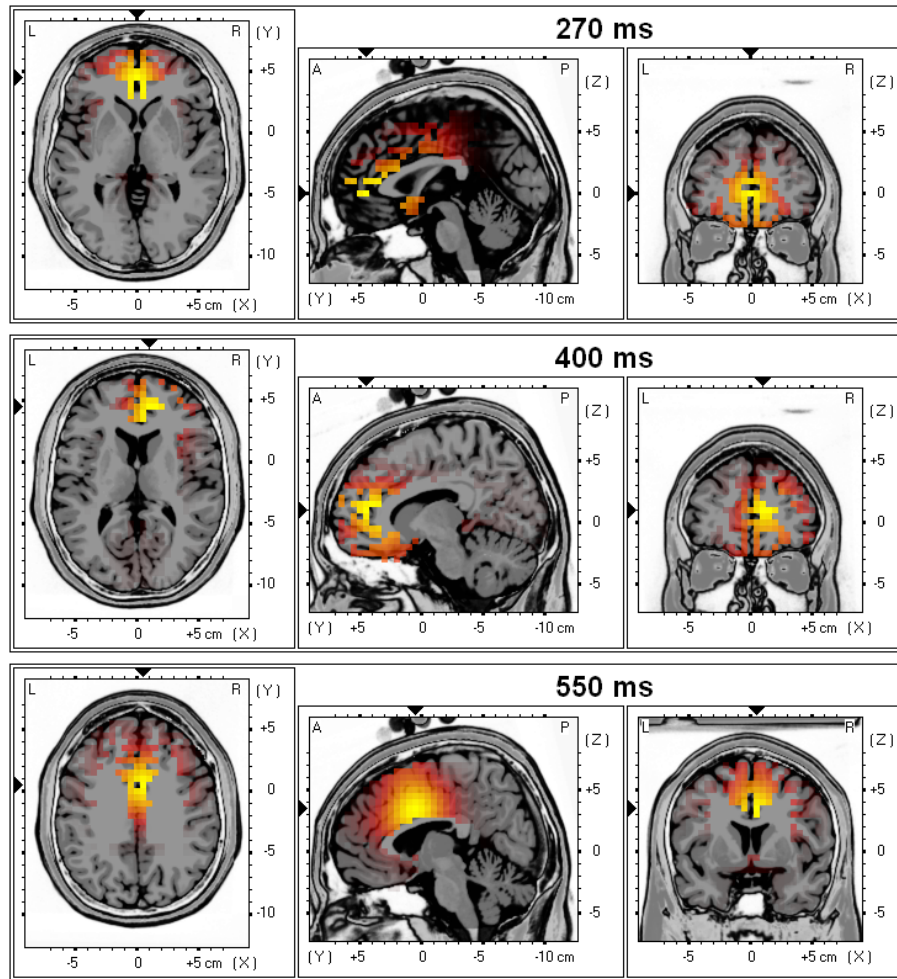


Figure 4.4: Localization of interaction ErrP.

Talairach slices of localized activity for the grand average of the four subjects at the occurrence of the three described peaks. As expected, the anterior cingulate cortex (ACC, Brodmann areas 24 & 32) is systematically activated.

4.4 Single-trial classification

To explore the feasibility of detecting single-trial erroneous responses, we have done a 10-fold cross-validation study where the testing set consists of one of the recorded sessions. In this way, testing is always done on a different recording session to those used for training the model. As explained above, classification is based on temporal features of channels Cz and Fz. The input vector to the statistical classifier is a half a second window starting 150 ms after the feedback. This time window encompasses the main ErrP components described above.

Table 4.1 reports the recognition rates (mean and standard deviations) for the four subjects plus the average of them. These results show that single-trial recognition of erroneous responses is 80% on average, while the recognition rate of correct responses is slightly better (83.7%). Quite importantly, even for the subject with the worse detection rates, they are around 75%. Beside the crucial importance to integrate ErrP in the BCI in a way that the subject still feels comfortable, for example, by reducing as much as possible the rejection of actually correct

commands, a key point for the exploitation of the automatic recognition of interaction errors is that they translate into an actual improvement of the performance of the BCI, which we can measure in terms of the bit rate.

Table 4.1: Recognition rates with the first protocol.
Percentages of correctly recognized error trials and correct trials for the four subjects and the average of them.

	Error [%]	Correct [%]	Average [%]
Subject 1	87.3±11.3	82.8±7.2	85.1±3.2
Subject 2	74.4±12.4	75.3±10.0	74.9±0.6
Subject 3	78.1±14.8	89.2±4.9	83.7±7.8
Subject 4	80.9±11.3	87.3±5.2	84.1±4.5
Average	80.2±5.4	83.7±6.2	81.9±4.7

4.5 Bit rate improvement

Table 4.2 reports the theoretical performances of a BCI that integrates ErrP for the four subjects and the average of them, where we have assumed an accuracy of 80% for the recognition of the subjects intent. These theoretical performances were calculated using the tools described in Section 2.7.5. These figures are to be compared to the performance of a standard BCI (i.e., without integrating ErrP). We have also reported the performances in the case $N_c = 3$, as the mind-controlled robot described by Millán et al. [Millán et al., 2004a]. In the case of standard two-class and three-class BCI, the performances are 0.28 and 0.66 bits per trial, respectively. Results indicate that there is a significant improvement in performance in the case of stopping outcomes, which is above 75% on average and higher than 90% for one of the subjects. Surprisingly, replacing the wrong outcome leads to smaller improvements and, in the case of subject 2, even to a significant degradation.

Table 4.2: Gain in performance with the first protocol.
Performances of the BCI integrating ErrP for the four subjects and the average of them.

	$N_c = 3$			$N_c = 2$				
	Initial BpT	Stop BpT	Gain	Initial BpT	Stop BpT	Gain	Replace BpT	Gain
Subject 1	0.66	0.91	37%	0.28	0.53	91%	0.36	29%
Subject 2	0.66	0.73	10%	0.28	0.40	42%	0.19	-32%
Subject 3	0.66	0.92	38%	0.28	0.52	86%	0.44	59%
Subject 4	0.66	0.91	37%	0.28	0.52	86%	0.42	50%
Average	0.66	0.86	30%	0.28	0.49	76%	0.34	23%

4.6 ErrP and Oddball N200 and P300

Since our protocol is quite similar to an oddball paradigm, the question arises of whether the potentials we describe are simply oddball N200 and P300. An oddball paradigm is characterized by an infrequent or especially significant stimulus interspersed with frequent stimuli. The subject is accustomed to a certain stimulus and the occurrence of an infrequent stimulus generates a negative deflection (N200) about 200 ms after the stimulus, followed by a positive peak (P300) about 300 ms after the stimulus. Our protocol is very close to an oddball paradigm in the sense that the subject is accustomed to seeing the increase in stages of the correct progress bars, and the increase in stages of the wrong progress bar is the infrequent stimulus. To clarify this issue, we have run new series of experiments for the ErrP study. In the new series of experiments, the interface executed the subjects command with an error rate of 50% and, so, error trials are no longer less frequent than correct trials. Analysis of the ErrP for different subjects using error rates of 20% and 50% show no difference between them except that the amplitude of the potentials are smaller in the case of an error rate of 50%, but the time course remains the same. This is in agreement with all previous findings on ErrP that show that the amplitude is inversely proportional to the error rate [Falkenstein et al., 2000]. It is worthwhile to note that the average classification rate with an error rate of 50% was about 75%. We can conclude then that, while we cannot exclude the possibility that N200 and P300 contribute to the potentials in the case of an error rate of 20%, the oddball N200 and P300 are not sufficient to explain the reported potentials.

4.7 Ocular artifacts

In the reported experiments, subjects look in the middle of the two progress bars, awaiting the central GO to press the key corresponding to the desired bar. After the feedback, the subjects become aware of the correct or erroneous response and they will shift their gaze to the side of the progress bar that has just been filled, so that there is a gaze shift in every single trial. Nevertheless, it is possible that the subjects concentrate upon the side of the progress bar they want to complete. After an erroneous trial, they will shift their gaze to the other side, so that the gaze shift could be present in erroneous trials only. The statistical classifier could therefore pick those gaze shifts since several prototypes per class were used.

To demonstrate that there is no systematical influence of gaze shifts on the presented ErrP as well as on classification results, we have calculated the different averages of the single trials with respect to the side of the progress bar that was intended to be completed: left error, right error, left correct, right correct. Figure 4.5 shows these four averages at channel Cz. The top left graph shows the average of erroneous single trials when the left progress bar was selected for the four subjects and the average of them. The top right graph shows the average of erroneous single trials with respect to the right bar. The bottom left and right graph show the average of correct trials with respect to the left and right progress bar, respectively. The left and right erroneous averages as well as the left and right correct averages are very similar whereas the left erroneous and correct as well as the right erroneous and correct are very different. So it appears that there is no systematical influence of gaze shifts on the reported potentials. Eye blinks are another potential source of artifacts. Indeed, it is conceivable that subjects may blink more frequently after one of the two conditions, and so the classifier could partly rely on

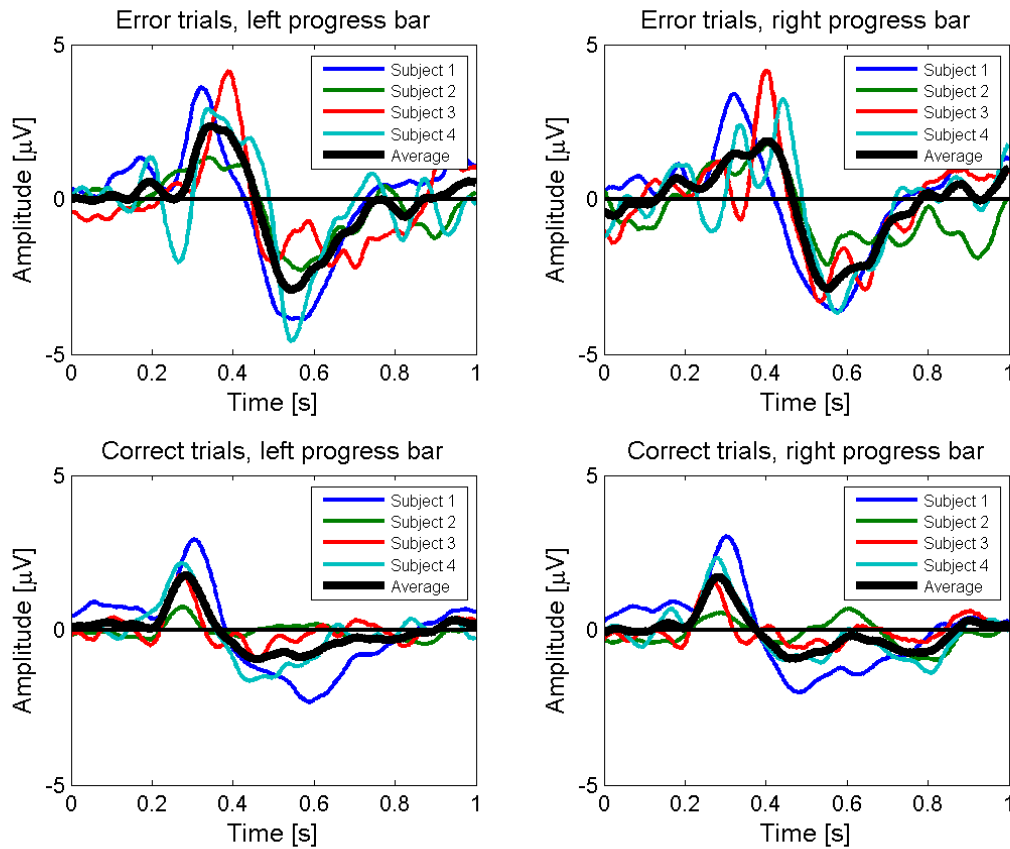


Figure 4.5: Ocular artifacts.

Averages of the single trials at channel Cz with respect to the side of the progress bar that was intended to be completed for the four subjects and the average of them. The left and right erroneous averages as well as the left and right correct averages are very similar whereas the left erroneous and correct as well as the right erroneous and correct are very different. This probably excludes any artifacts due to gaze shifts.

eye blinks to discriminate error and correct trials. However, the scalp potentials topographies of Figure 4.3 show that the three ErrP components do not have a front focus, which would be expected in blink-related potentials. So, as for the gaze shifts, it appears that there is no systematical influence of eye blinks on the reported results.

4.8 Conclusion

In this study we have reported first results on the detection of the neural correlate of error awareness for improving the performance and reliability of BCI. In particular, we have found what seems to be a new kind of error-related potentials elicited in reaction to an erroneous recognition of the subjects intention. An important difference between response ErrP, feedback ErrP, and observation ErrP on one side and the reported interaction ErrP on the other side is that the former involve a stimulus from the system for every single trial whereas the latter involve a choice of a long-term goal made by the subjects themselves (choice of the progress bar). More

importantly, we have shown the feasibility of detecting single-trial erroneous responses of the interface that lead to significant improvements of the information transfer rate of a BCI even though these improvements are theoretical. Indeed, the introduction of an automatic response rejection strongly interferes with the BCI. The user needs to process additional information that induces higher workload and may considerably slow down the interaction. These issues will be investigated when running online BCI experiments integrating automatic error detection. Furthermore, this study seems to confirm the major role played by the ACC in error potentials generation.

As already discussed, the major drawback of the protocol described in this chapter is that it exhibits the main characteristics of an Oddball paradigm, so that the contribution of Oddball N200 and P300 to the reported potentials remains unclear. Therefore, before working on the actual integration of ErrP detection into the BCI, we decided to repeat these experiments using a more realistic graphical interface that prevents habituation of the subject to one of the stimuli and that is also more engaging. This new protocol will be the starting point for ErrP integration in an online system, i.e. the system will simultaneously detect ErrP and classify mental tasks for control. It will also be used to investigate the potential benefit of integrating an inverse model to improve the single trial classification.

Chapter 5

Second protocol: Moving cursor

In this chapter, we report a second attempt of recording error-related potentials elicited no more by an error made by the subject himself, but rather following an error of the interface when executing a command given by the subject. The main difference with the results presented in Chapter 4 is that the graphical interface is more realistic and engaging for the subject and that the protocol prevents habituation to one of the stimuli. The results reported in this chapter were presented in 2006 in Rome (Italy) at the MAIA Project's Workshop *Challenging Brain Computer Interfaces: Neural Engineering Meets Clinical Needs in Neurorehabilitation*. They were also recently accepted for publication in the *IEEE Transactions on Biomedical Engineering* journal.

5.1 Experimental setup

To test the presence of ErrP after a feedback indicating errors made by the interface in the recognition of the subject's intent in a more realistic and engaging protocol, we have simulated a human-robot interaction task where the subject has to bring the robot to targets 2 or 3 steps either to the left or to the right. This virtual interaction is implemented by means of a green square cursor that can appear on any of 20 positions along an horizontal line. The goal with this protocol is to bring the cursor to a target that randomly appears either on the left (blue square) or on the right (red square) of the cursor. The target is no further away than 3 positions from the cursor (symbolizing the current position of the robot). Contrarily to the protocol used in Chapter 4, this new protocol prevents the subject from habituation to one of the stimuli since the cursor reaches the target within a small number of steps. Figure 5.1 illustrates the protocol with the target (blue) initially positioned 2 steps away on the left side of the cursor (green). An error occurred at step 3) so that the cursor reaches the target in 5 steps.

As for the protocol using the progress bars, to isolate the issue of the recognition of ErrP out of the more difficult and general problem of a whole BCI where erroneous feedback can be due to non-optimal performance of both the interface (i.e., the classifier embedded into the interface) and the user himself, in the following experiments the subject delivers commands manually and not mentally. That is, he/she simply presses a left or right key with the left or right hand. In this way, any error feedback is only due to a wrong recognition by the interface of which is the subject's intention. Five volunteer healthy subjects participated in these experiments. After the presentation of the target, the subject pressed the corresponding key until the cursor reached the target. The system moved the cursor with an error rate of 20%; i.e., at

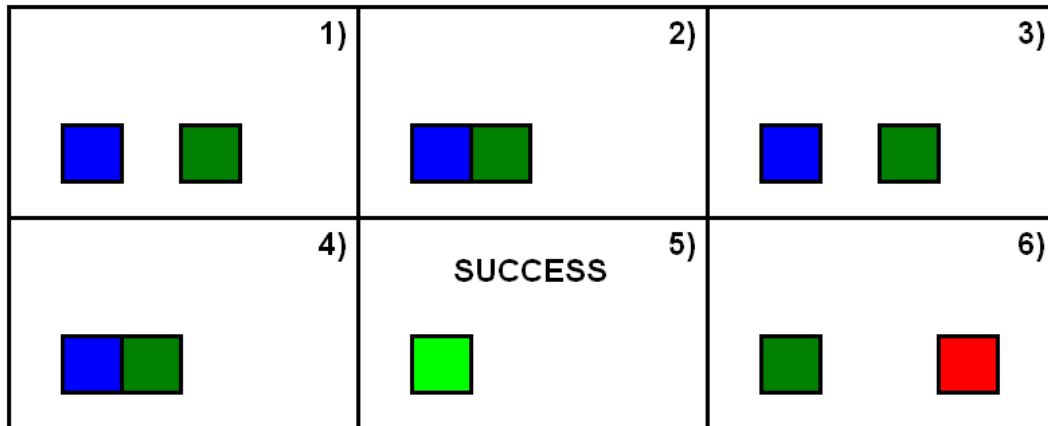


Figure 5.1: New graphical interface: Moving cursor.

Illustration of the protocol. (1) The target (blue) appears 2 steps on the left side of the cursor (green). (2) The subject pressed the left key and the cursor moves 1 step to the left. (3) The subject still pressed the left key, but the system moves the cursor in the wrong direction. (4) Correct move to the left, compensating the error. (5) The cursor reaches the target. (6) A new target (red) appears 3 steps on the right side of the cursor. The system moved the cursor with an error rate of 20%; i.e., at each step, there was a 20% probability that the robot made a movement in the wrong direction.

each step, there was a 20% probability that the cursor moved in the opposite direction. After the cursor reached the target, the word **SUCCESS** appeared above the target, and a new target was randomly selected by the system. If the cursor didn't reach the target after 10 steps, the word **FAILED** appeared under the target and a new target was selected. Subjects performed 10 sessions of 3 minutes on 2 different days, corresponding to ~ 75 single trials per session. The delay between the two days of measurements was about 3 months.

EEG potentials were acquired using 64 electrodes and raw EEG data were pre-processed according to the procedure described in Section 2.7.2. As in the previous chapter, no artifact rejection algorithm was applied and all trials were kept for classification. Again, after a visual a posteriori check, we found no evidence of muscular artifacts that could have contaminated one condition differently from the other.

5.2 Grand averages

With this new protocol, it is also first necessary to check whether or not ErrP are present no more in reaction to an error made by the subject himself, but in reaction to an erroneous response made by the interface as indicated by the feedback visualizing the recognized subject's intention.

Figure 5.2 shows the difference error-minus-correct for channel FCz for the five subjects plus the grand average of the five subjects for the two days of recordings. A first positive peak shows up 200 ms after the feedback. A negative peak and a positive peak can be seen 250 ms and 320 ms after the feedback, respectively. Finally, a second broader negative peak occurs about 450 ms after the feedback. All five subjects show very similar ErrP time courses whose amplitudes slightly differ from one subject to the other. These potentials are very similar to those described in Chapter 4, the only difference is that the different components appear faster.

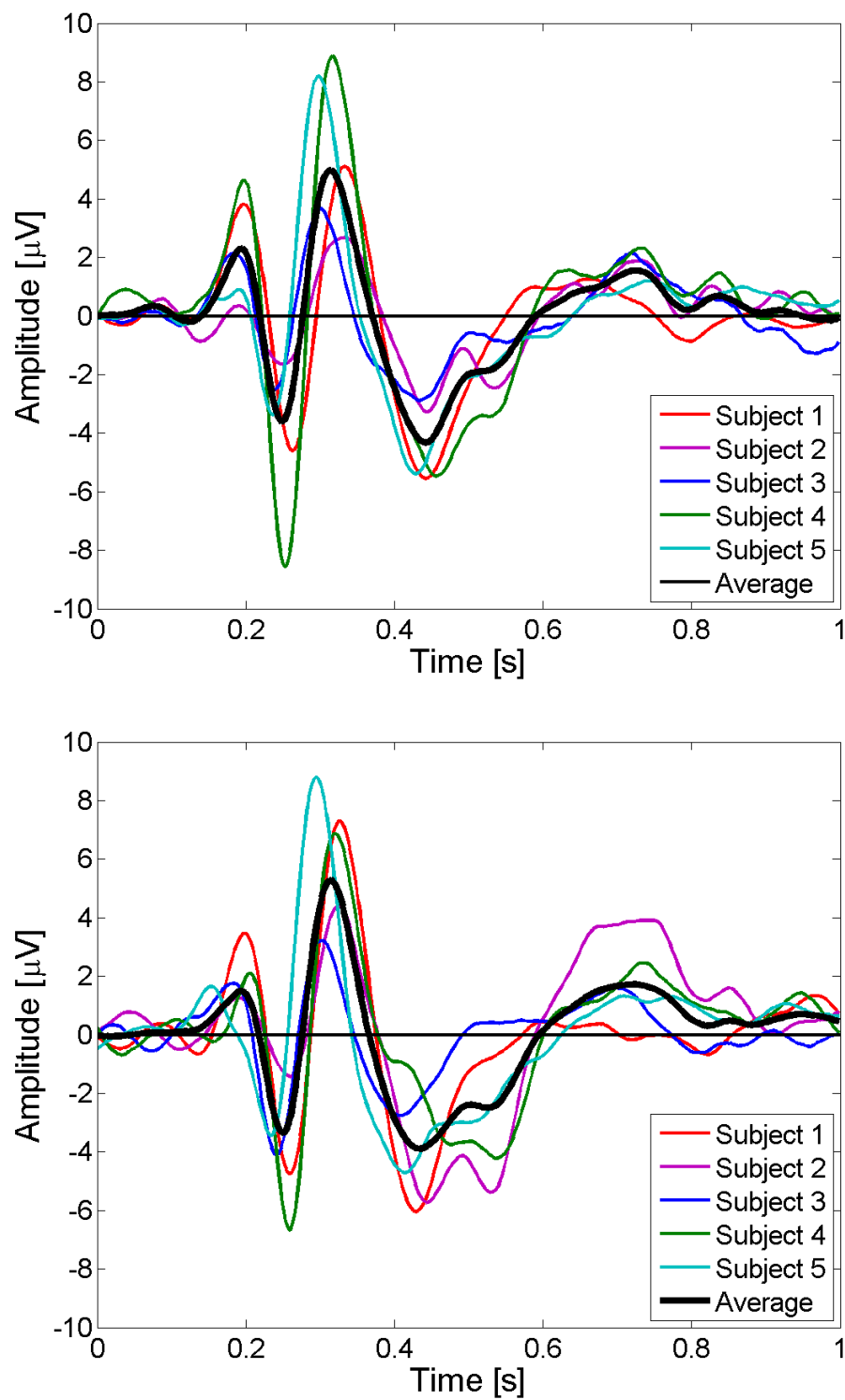


Figure 5.2: Interaction ErrP, 5 subjects.

Average EEG for the difference error-minus-correct at channel FCz for the five subjects plus the grand average of them for the first day (top) and for the second day (bottom) with an error rate of 20%.

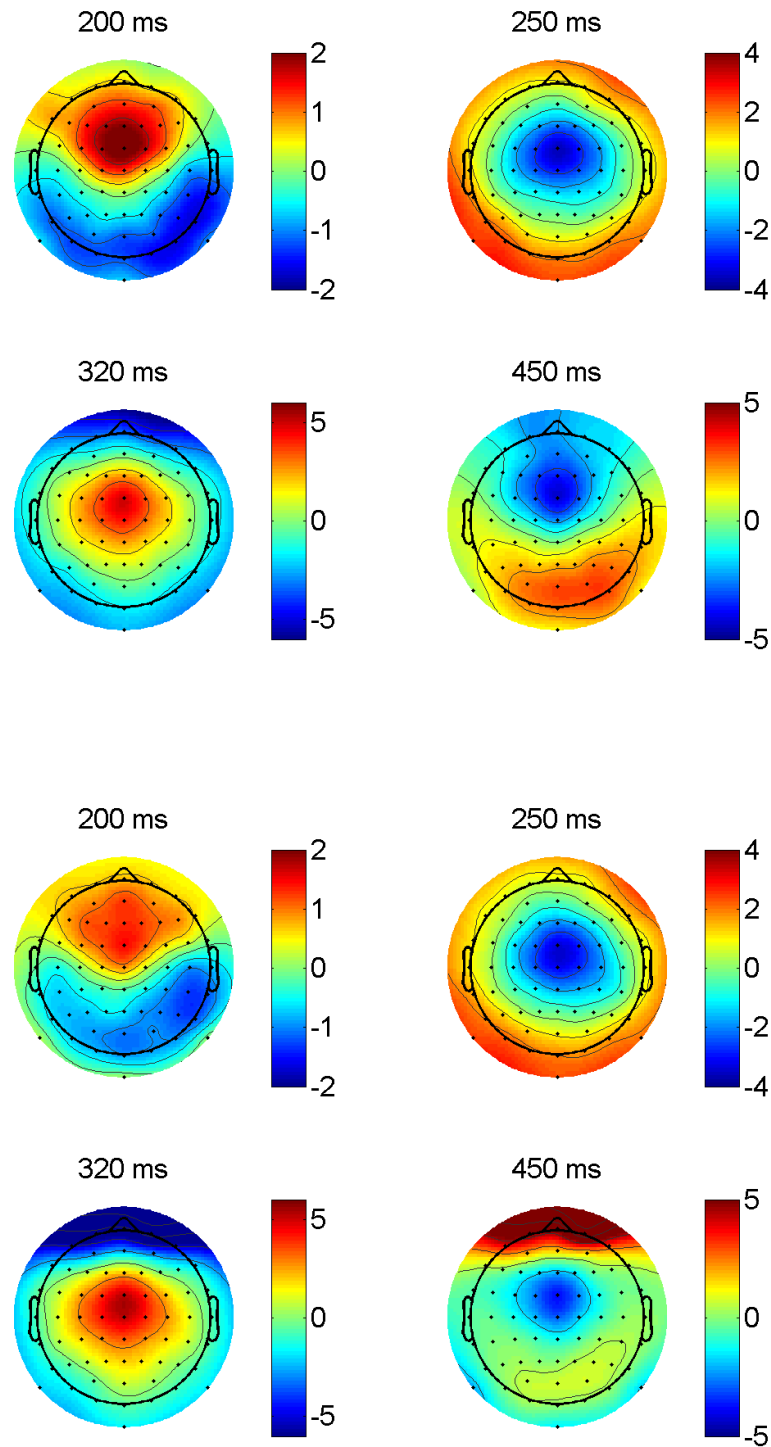


Figure 5.3: Scalp potentials topographies.

Scalp topographies for the average of the five subjects for the first day (top) and the second day (bottom). A first fronto-central positivity appears after 200 ms, followed by a fronto-central negativity at 250 ms, a fronto-central positivity at 320 ms and a fronto-central negativity at 450 ms.

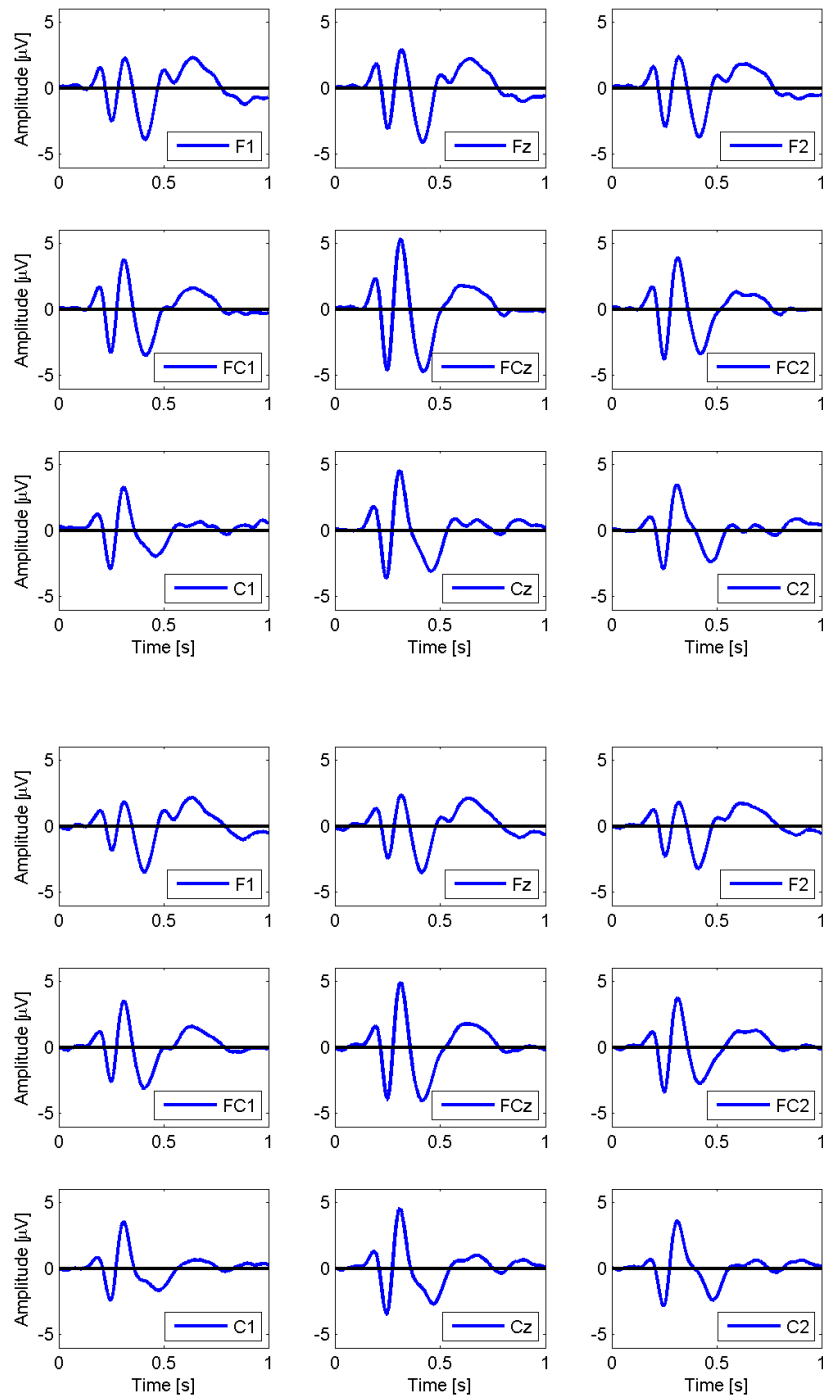


Figure 5.4: FCz and its 8 neighbors.

Average EEG for the difference error-minus-correct at channel FCz and its 8 neighbors for the average of the five subjects for the first day (top) and for the second day (bottom). ErrP are very stable over time (about 3 months between both days of recordings) and channel FCz shows the largest ErrP amplitudes.

Indeed, if for both protocols there's a negative peak about 250 ms after the feedback, the main positive peak occurs 80 ms earlier with this new protocol (320 ms v.s. 400 ms). In the same way, the second negative peak occurs 100 ms earlier (450 ms v.s. 550 ms). It is also to note that amplitudes are on average double with this new protocol. This could reflect the fact that the simulation is more realistic and engaging and that the errors are more disturbing. Indeed, with the progress bars, adding a step to the wrong bar don't go against the goal of filling the chosen bar whereas with this moving cursor, an error means move away from the goal.

Figure 5.3 shows the scalp potentials topographies, for the grand average EEG of the five subjects, at the occurrence of the four previously described peaks: a first fronto-central positivity appears after 200 ms, followed by a fronto-central negativity at 250 ms, a fronto-central positivity at 320 ms and a fronto-central negativity at 450 ms.

Finally, Figure 5.4 shows the difference error-minus-correct for channel FCz and its 8 neighbors (namely, F1, Fz, F2, FC1, FC2, C1, Cz and C2) for the average of the five subjects for both days of recordings. Channel FCz shows the largest ErrP amplitudes, and as it can already be seen in Figure 5.2 and Figure 5.3, even if the delay between the two days of recordings was about three months, potentials as well as scalp topographies are very similar for both days of recordings. These experiments seem to confirm the existence of a new kind of error-related potentials, the so-called *interaction ErrP*. As explained in Chapter 4, the general shape of these ErrP is quite similar to the shape of the response ErrP in choice reaction tasks, whereas the timing is similar to the feedback ErrP in reinforcement learning tasks and to the observation ErrP.

5.3 Electrodes selection

Even if ErrP are detected in a scalp area covered by electrode FCz and neighbors, it is important to run a systematic feature selection procedure to determine the best electrodes for ErrP classification. The most relevant EEG electrodes were selected using the simple feature selection algorithm described in Section 2.7.3.

Figure 5.5 shows the discriminant power (DP) of the 64 electrodes for the 5 subjects and for the average of them. For each electrode, the DP is the average of the DP of all time points in the half a second window used for classification (starting 150 ms after the feedback). This window was selected for classification because it encompasses all ErrP components described above. For all five subjects, a relatively small fronto-central area shows the best discriminant power, and the best electrode is always either FCz or Cz. The average map confirms that FCz and Cz are clearly the 2 best electrodes. Therefore we selected those 2 electrodes for classification for all subjects since all five subjects show very similar discriminant maps. The selection of electrodes FCz and Cz is also motivated by the fact that the anterior cingulate cortex has a median fronto-central localization. Choosing different electrodes for each subject could maybe lead to higher ErrP recognition rates. However, channels FCz and Cz are very relevant for all subjects, and choosing the same electrodes for all subjects will simplify the future online procedures since no feature selection will be needed anymore. More generally, if choosing the same electrodes for all subjects is possible for ErrP, this is not feasible with spontaneous activity, as we will see in Chapter 6.

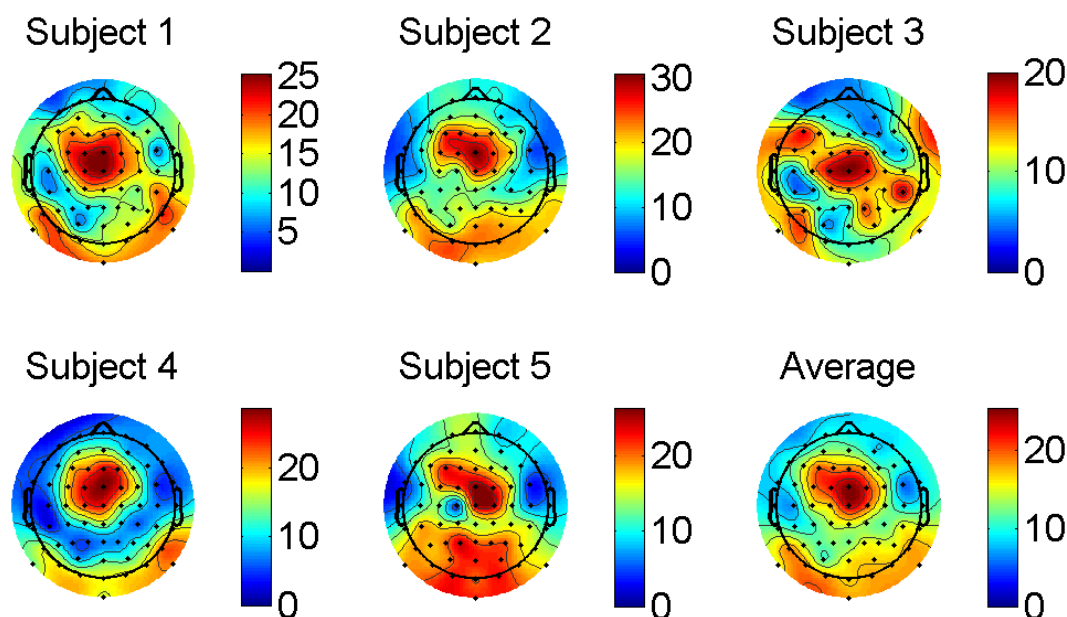


Figure 5.5: Discriminant power of electrodes for ErrP classification. Discriminant power (DP) of the 64 electrodes for the 5 subjects and for the average of them. The best 2 electrodes are FC and Cz and classification will be based on the time series of these 2 channels.

5.4 Estimation of intracranial activity

As for the progress bars protocol, the sLORETA software was used as a localization tool to estimate the focus of intracranial activity at the occurrence of the four ErrP peaks previously described. Figure 5.6 shows Talairach slices of localized activity for the grand average of the five subjects for the first day of recording at the occurrence of the four peaks. As expected, the areas involved in error processing, namely the pre-supplementary motor area (pre-SMA, Brodmann area 6) and the rostral cingulate zone (RCZ, Brodmann areas 24 & 32) are systematically activated [Holroyd and Coles, 2002, Fiehler et al., 2004]. Indeed, for the first positive peak (200 ms after the feedback), the focus of activity is located at $X=0\text{mm}$, $Y=-5\text{mm}$, $Z=50\text{mm}$ (MNI coordinates). The best match is Brodmann area 24 (ventral anterior cingulate cortex). For the first negative peak, (250 ms after the feedback), the focus is at $X=0\text{mm}$, $Y=0\text{mm}$, $Z=40\text{mm}$ and the best match is again Brodmann area 24. For the second positive peak (320 ms after the feedback), the focus is at $X=0\text{mm}$, $Y=5\text{mm}$, $Z=50\text{mm}$ and the best match is in this case Brodmann area 32 (dorsal anterior cingulate cortex). Finally for the second negative peak (450 ms after the feedback), the focus is on Brodmann area 6 (pre-supplementary motor area) at $X=0\text{mm}$, $Y=-10\text{mm}$, $Z=55\text{mm}$. For the second positive peak (320 ms) and the second negative peak (450 ms), parietal areas starts to be also activated. These associative areas (somatosensory association cortex, Brodmann areas 5 & 7) could be related to the fact that the subject becomes aware of the error. It has been proposed that the positive peak generated in a reaction task was associated with conscious error recognition [Nieuwenhuis et al., 2001]. In our case, activation of parietal areas about 300ms after the feedback agrees with this hypothesis. On average, the localization with this new protocol appear to be slightly more posterior than with the progress bars protocol used in Chapter 4. Indeed, even if the best match is on the ACC or on the pre-

SMA, the posterior cingulate cortex (PCC, Brodmann areas 23 & 31, cf. Appendix A) is often in the 5 best matches. The posterior cingulate cortex is involved with the medial pre-frontal cortex in anticipatory allocation of spatial attention [Kim et al., 1999, Mesulam et al., 2001]. With this protocol subjects are anticipating the future position of the cursor. Therefore the PCC could have a different activity when the feedback correspond to the subject's anticipation (correct trials) and when the feedback doesn't correspond to the subject's anticipation (erroneous trials). These considerations could explain the fact that the foci of activity are located at the limits of the ACC, the PCC and the pre-SMA.

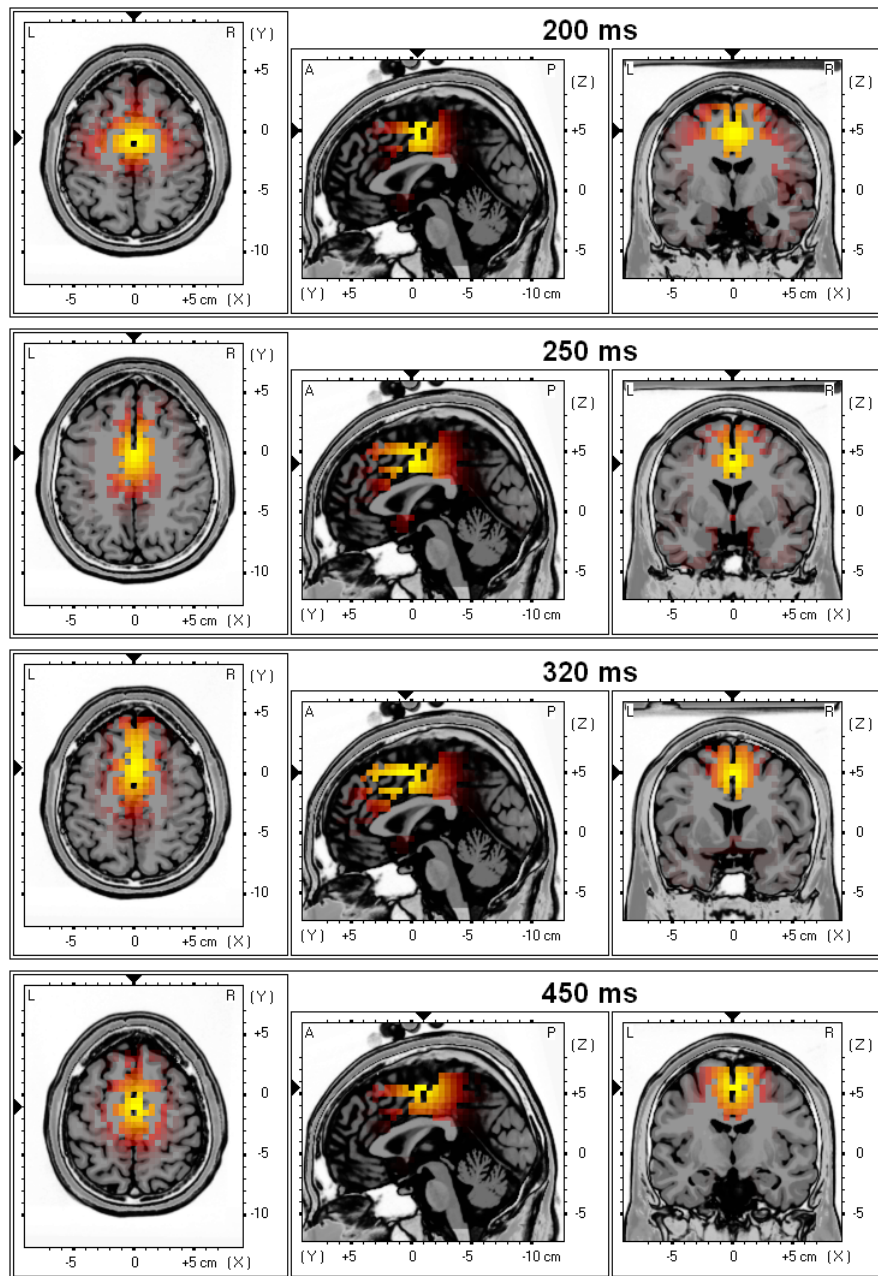


Figure 5.6: Localization of interaction ErrP.

Talairach slices of localized activity for the grand average of the five subjects for the first day of recordings at the occurrence of the four described peaks.

5.5 Single trial classification

As for the progress bars protocol, to explore the feasibility of detecting single-trial erroneous responses, we have done a 10-fold cross-validation study for both days of recordings where the testing set consists of one of the recorded sessions. In this way, testing is always done on a different recording session to those used for training the model. Furthermore, since we had two different days of recording, we have also tried to classify data of the second day using all data of the first day to build a classifier. Table 5.1 reports the recognition rates (mean and standard deviations) for the five subjects plus the average of them. According to the feature selection results, time series of electrodes FCz and Cz were concatenated for classification. However, it is to note here that other couples of electrodes were tested, namely FCz-Fz, FCz-FC1, and FCz-FC2, but the couple FCz-Cz is on average the best couple.

Table 5.1: Recognition rates with the second protocol.

Percentages (mean and standard deviations) of correctly recognized error trials and correct trials for the five subjects and the average of them for both days performing a 10-fold cross validation and for the second day, using a classifier built with data of the first day.

10-fold cross-validation Day I			
	Error [%]	Correct [%]	Average [%]
Subject 1	80.5±11.6	83.4±4.3	82.0±2.1
Subject 2	76.9±13.7	69.3±8.6	73.1±5.4
Subject 3	74.3±17.5	78.8±8.0	76.6±3.2
Subject 4	89.5±9.6	93.1±4.6	91.3±2.5
Subject 5	89.0±9.4	87.3±5.2	88.2±1.2
Average	82.0±6.9	82.4±9.0	82.2±7.6

10-fold cross-validation Day II			
	Error [%]	Correct [%]	Average [%]
Subject 1	82.4±10.4	85.4±6.7	83.9±2.1
Subject 2	81.2±13.4	72.6±5.4	76.9±6.1
Subject 3	83.8±13.1	85.3±3.5	84.6±1.1
Subject 4	85.0±12.6	86.4±6.4	85.7±1.0
Subject 5	86.7±13.8	88.6±4.9	87.7±1.3
Average	83.8±2.2	83.7±6.3	83.7±4.1

Day II classified with Day I			
	Error [%]	Correct [%]	Average [%]
Subject 1	83.4±10.7	85.8±4.8	84.6±1.7
Subject 2	69.5±18.8	76.2±6.9	72.9±4.7
Subject 3	73.5±9.4	80.2±7.3	76.9±4.7
Subject 4	86.0±8.7	85.5±6.6	85.8±0.4
Subject 5	84.0±11.8	89.7±4.5	86.9±4.0
Average	79.3±7.3	83.5±5.3	81.4±6.2

The results of the cross-validations show that single-trial recognition of erroneous and correct responses is above 82% for both days. More importantly, classification using a classifier built with data recorded up to three months earlier show similar results (79.3% and 83.5% for error and correct trials, respectively). These classification rates are similar with the results obtained in the previous chapter. This stability is in accordance with the stability of the potentials described above. But again, beside the crucial importance to integrate ErrP in the BCI in a way that the subject still feels comfortable, for example by reducing as much as possible the rejection of actually correct commands, a key point for the exploitation of the automatic recognition of interaction errors is that they translate into an actual improvement of the performance of the BCI. Indeed, as previously shown in Chapter 4, the integration of error potentials detection theoretically leads to an increase of more than 75% of performance expressed in terms of bit rate.

5.6 ErrP and Oddball N200 and P300

Even if this new protocol is more realistic and engaging and even if it prevents habituation of the subject to one of the stimuli, it still shows frequent (correct movements) and infrequent (erroneous movements) stimuli so that this new protocol still has similarities with an oddball paradigm and the question arises of whether the potentials we describe are simply oddball N200 and P300. As in Chapter 4, to check this issue we have run a series of experiments where the interface executed the subject's command with an error rate of 50% and, so, error trials are no longer less frequent than correct trials. Analysis of the ErrP for the five subjects at channel FCz using error rates of 20% and 50% show no difference between them except that the amplitude of the potentials are smaller in the case of an error rate of 50%, but the time course remains the same as shown in Figure 5.7. This is in agreement with all previous findings on ErrP that show that the amplitude is inversely proportional to the error rate [Falkenstein et al., 2000]. Figure 5.7 shows scalp topographies at the occurrence of the four previously described peaks. All peaks show a fronto-central focus, only the first positive peak (200 ms after the feedback) has a too low amplitude to show a clear focus. To summarize, we can conclude that, while we cannot exclude the possibility that N200 and P300 contributes to the potentials in the case of an error rate of 20%, the oddball N200 and P300 are not sufficient to explain the reported potentials.

5.7 Ocular artifacts

As for the progress bars protocol, the problem of eventual ocular artifacts arise. With this new protocol, subjects look at the cursor, awaiting its movement after he/she has pressed the key corresponding to the target. After the feedback, the subject gets aware of the correct or erroneous response and he/she will shift gaze to the new position of the cursor, so that there is a gaze shift in every single trial. Nevertheless, it is possible that the subject looks at the target rather than the cursor. Consequently, there will be a gaze shift only after erroneous trials or, in any case, a larger gaze shift in erroneous trials. The statistical classifier could therefore pick those gaze shifts since several prototypes per class were used. To demonstrate that there is no systematical influence of gaze shifts on the presented ErrP as well as on classification results, we have again calculated the different averages of the single trials with respect to the side of the

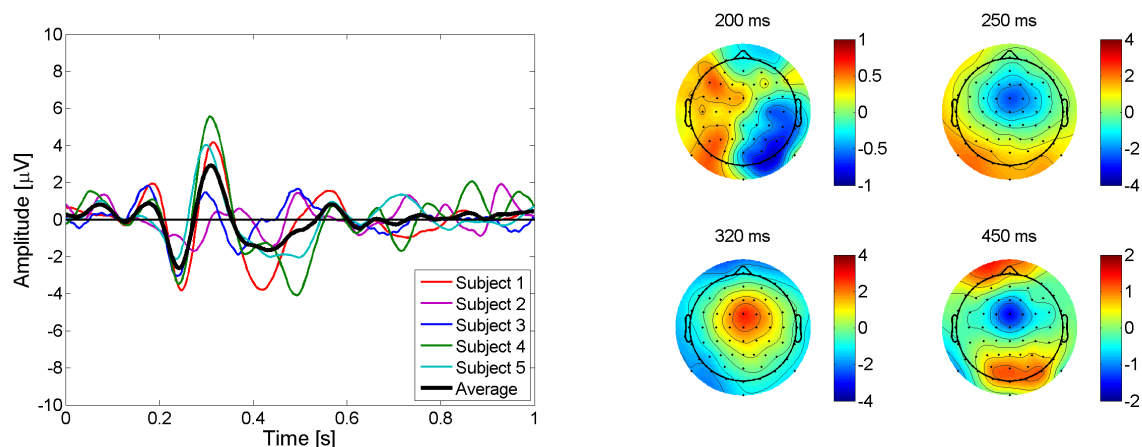


Figure 5.7: ErrP with an error rate of 50%.

Left. Average EEG for the difference error-minus-correct at channel FCz for the five subjects plus the grand average of them with an error rate of 50%. Feedback is delivered at time 0 seconds. The same potentials as in Figure 5.2 can be clearly seen, only the amplitudes are slightly smaller. *Right.* Scalp potentials topographies, for the grand average EEG of the five subjects, at the occurrence of the four described peaks.

target: left correct, right correct, left error, right error. Figure 5.8 shows these four averages at channel FCz. The top left graph shows the average of error single trials when the target appeared on the left for the five subjects and the average of them. The top right graph shows the average of error single trials with respect to the right side. The bottom left and right graph show the average of correct trials with respect to the left and right side, respectively. The left and right correct averages as well as the left and right erroneous averages are very similar whereas the left correct and erroneous as well as the right correct and erroneous are very different. So it appears that there is no systematical influence of gaze shifts on the reported potentials. Eye blinks are another potential source of artifacts. Indeed, it's conceivable that subjects may blink more frequently after one of the two conditions, and so the classifier could partly rely on eye blinks to discriminate error and correct trials. However, the scalp topographies of Figure 5.3 show that the four ErrP components do not have a front focus, which would be expected in blink related potentials. So, as for the gaze shifts, it appears that there is no systematical influence of eye blinks on the reported results.

5.8 Conclusion

In this chapter we have reported results on the detection of the neural correlate of error awareness for improving the performance and reliability of BCI. In particular, we have confirmed the existence of a new kind of error-related potentials elicited in reaction to an erroneous recognition of the subject's intention. More importantly, we have shown the feasibility of detecting single-trial erroneous responses of the interface and we have shown the stability of these potentials over time. In this chapter we have also shown that, as expected, typical cortical areas involved in error processing such as pre-supplementary motor area and anterior cingulate cortex are systematically activated at the occurrence of the different peaks.

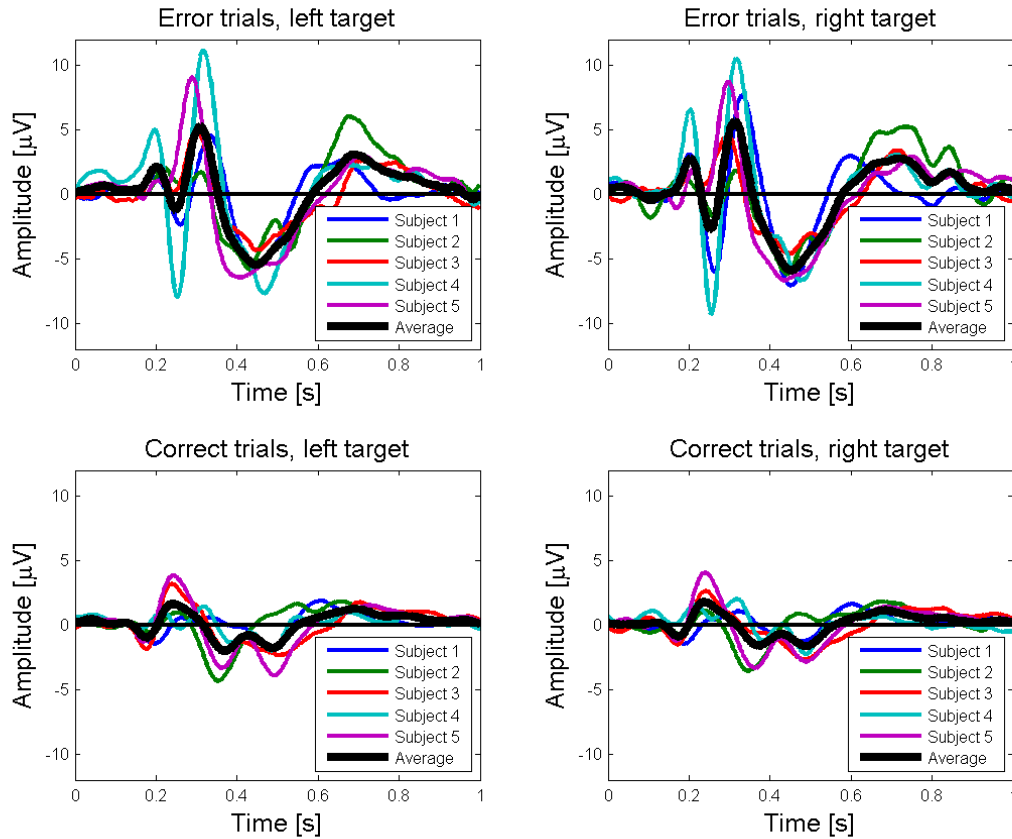


Figure 5.8: Ocular artifacts.

Averages of the single trials at channel FCz with respect to the side of the target for the five subjects and the average of them. The left and right erroneous averages as well as the left and right correct averages are very similar whereas the left erroneous and correct as well as the right erroneous and correct are very different. This probably excludes any artifacts due to gaze shifts.

However, the introduction of an automatic verification procedure strongly interferes with the BCI. The user needs to process additional information which induces higher workload and may considerably slow down the interaction. These issues will be investigated when running online BCI experiments integrating automatic error detection using this more realistic and engaging protocol. The key point becomes to simultaneously classify mental tasks for control and detect ErrP presence to improve the BCI accuracy. As already mentioned, another explored issue will be the potential benefit of using an inverse solution to increase the classification rates of error and correct single trials.

Chapter 6

Mental commands and ErrP detection

In this chapter, we report the feasibility of simultaneously and satisfactorily classify motor imagination for the mental control of a brain-actuated device and detecting erroneous responses of the interface to improve the BCI accuracy. Subjects were using the same protocol as in the previous chapter, but used motor imagination instead of manual commands. The results reported in this chapter were recently submitted to the *Twenty-First Annual Conference on Neural Information Processing Systems (NIPS)* that will be held Vancouver (Canada) in December 2007.

6.1 Experimental setup

The first step to integrate ErrP detection in a BCI is to design a protocol where the subject is focussing on a mental task for device control and on the feedback delivered by the BCI for ErrP detection. To test the ability of BCI users to concentrate simultaneously on a mental task and to be aware of the BCI feedback at each single trial, we have used the same protocol as in Chapter 5. As already explained in Section 5.1, in this simulation of human-robot interaction, the subject has to bring the robot to targets 2 or 3 steps either to the left or to the right. Each target corresponds to a specific mental task. The subjects were asked to imagine a movement of their left hand for the left target and to imagine a movement of their right foot for the right target (note that subject n°1 selected left foot for the left target and right hand for the right target). However, since the subjects had no prior BCI experience, the system was not moving the cursor following the mental commands of the subject, but with an error rate of 20%, to avoid random or totally biased behavior of the cursor.

Six healthy volunteer subjects with no prior BCI experience participated in these experiments, four of them participated in the study presented in the previous chapter. After the presentation of the target, the subject focused on the corresponding mental task until the cursor reached the target. The system moved the cursor with an error rate of 20%; i.e., at each step, there was a 20% probability that the cursor moved in the opposite direction. When the cursor reached a target, it briefly turned from green to light green and then a new target was randomly selected by the system. If the cursor didn't reach the target after 10 steps, a new target was selected. As shown in figure 6.1, while the subject focuses on a specific mental task, the system delivers a feedback about every 2 seconds. This provides a window just before the feedback for BCI classification and a window just after the feedback for ErrP detection for every single trial. Subjects performed 10 sessions of 3 minutes on 2 different days (the delay between the

two days of measurements varied from 1 week to 1 month), corresponding to ~ 75 single trials per session. The 20 sessions were split into 4 groups of 5, so that classifiers were built using a group and tested on the following group. The classification rates presented in this chapter are therefore the average of 3 prediction performances: classification of group $n + 1$ using group n to build a classifier. This rule applies for both mental tasks classification and ErrP detection.

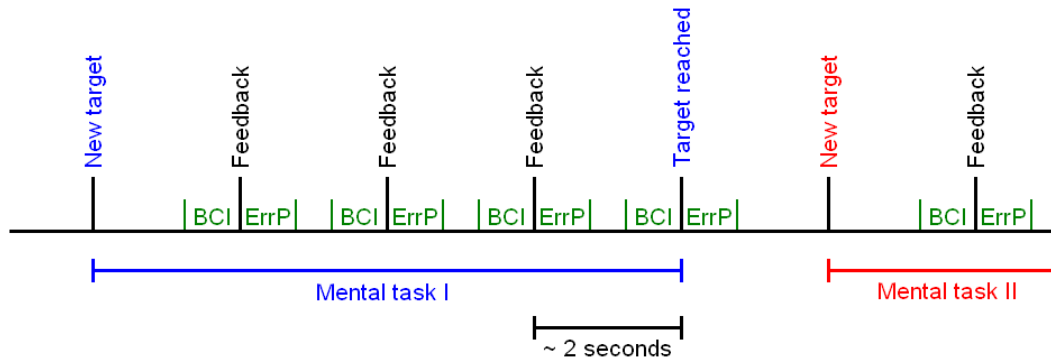


Figure 6.1: Timing of the protocol.

The system delivers a feedback about every 2 seconds, this provides a window just before the feedback for BCI classification and a window just after the feedback for ErrP detection for every single trial. As a new target is presented, the subject focuses on the corresponding mental task until the target is reached.

EEG potentials were acquired using 64 electrodes (cf. Section 2.7.1) and raw EEG data were pre-processed according to the procedure described in Section 2.7.2 but the bandpass filter and the decimation to 64 Hz was only applied for ErrP detection. For off-line mental tasks classification, the power spectrum density (PSD) of EEG channels was estimated over a window of one second just before the feedback. PSD was estimated using the Welch method resulting in spectra with a 2 Hz resolution from 6 to 44 Hz. The most relevant EEG channels and frequencies were selected using the feature selection algorithm described in Section 2.7.3. For off-line ErrP detection, the actual input vector for the statistical classifier is a 150 ms window starting 250 ms after the feedback for channels FCz and Cz. The choice of these 2 electrodes rely on the feature selection performed in the previous chapter. The time window starts later and is shorter than in the previous analysis. Indeed, as explained later in this chapter, the different ErrP components occur a little later (~ 30 ms) than in the previous studies. The selection of a shorter window was mainly decided for practical online purposes, a shorter window leads to faster ErrP detection.

6.2 Mental tasks classification

Subjects were asked to imagine a movement of their left hand when the left target was proposed and to imagine a movement of their right foot when the right target was proposed (note that subject n°1 was imagining left foot for the left target and right hand for the right target). The most relevant EEG channels and frequencies were selected by a simple feature selection algorithm based on the overlap of the distributions of the different classes as described in Section 2.7.3.

Figure 6.2 shows the discriminant power (DP) of frequencies and Figure 6.3 shows the DP of electrodes for the 6 subject. For frequencies, the DP is based on the best electrode, and for

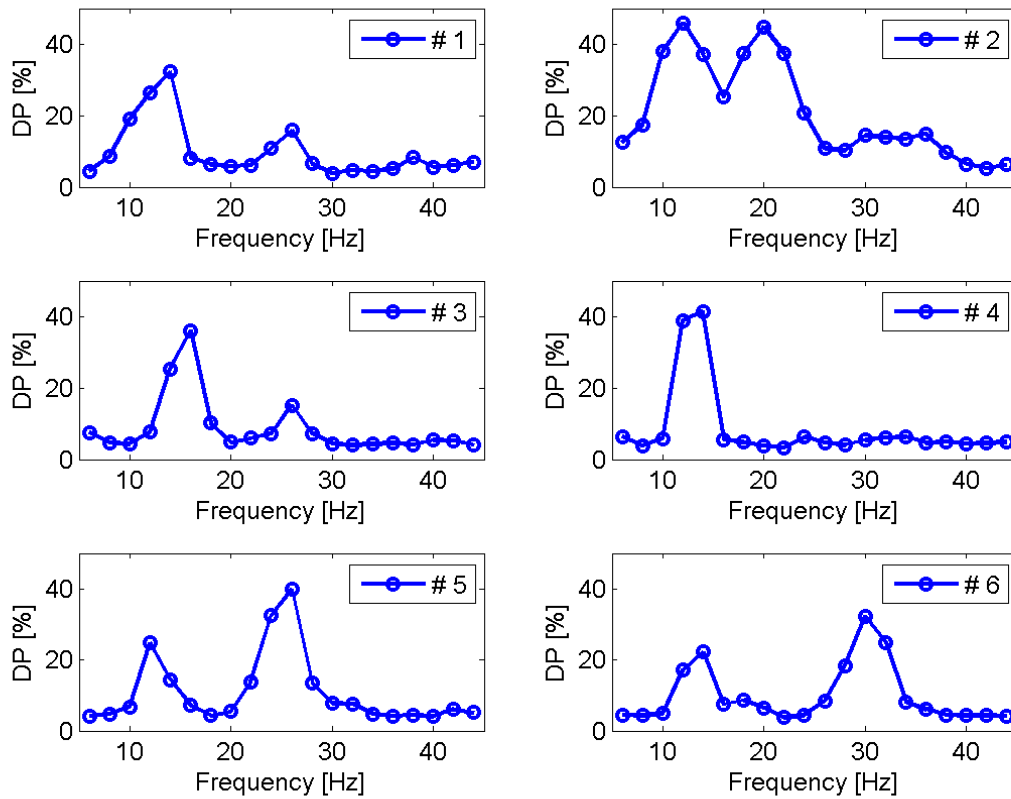


Figure 6.2: Discriminant power of frequencies.
Alpha rhythm and some beta components are discriminant for all six subjects.

electrodes it is based on the best frequency. Table 6.1 shows the classification rates for the two mental tasks and the general BCI accuracy for the 6 subjects and the average of them, it also shows the features (electrodes and frequencies) used for classification.

For all 6 subjects, the 12-16 Hz band appears to be relevant for classification. Subject 1, 3 and 5 show a peak in DP for frequencies around 25 Hz (beta band). For subject 2 this peak in the beta band is centered at 20 Hz and for subject 6 it is centered at 30 Hz. Finally subject 4 shows no particular discriminant power in the beta band. Previous studies confirm these results. Indeed, alpha and beta rhythm over left and/or right sensorimotor cortex have been successfully used for BCI control [McFarland and Wolpaw, 2005]. Event-related de-synchronization (ERD) and synchronization (ERS) refer to large-scale changes in neural processing. During periods of inactivity, brain areas are in a kind of idling state with large populations of neurons firing in synchrony resulting in an increase of amplitude of specific alpha (8-12 Hz) and beta (12-26 Hz) bands. During activity, populations of neurons work at their own pace and the power of this idling state is reduced, the cortex has become de-synchronized [Pfurtscheller and da Silva, 1999]. In our case, the most relevant electrodes for all subjects are in the C3, C4 or Cz area. These locations confirm previous studies since C3 and C4 areas usually show ERD/ERS during hands movement or imagination whereas foot movement or imagination are focused in the Cz area [Pfurtscheller and da Silva, 1999].

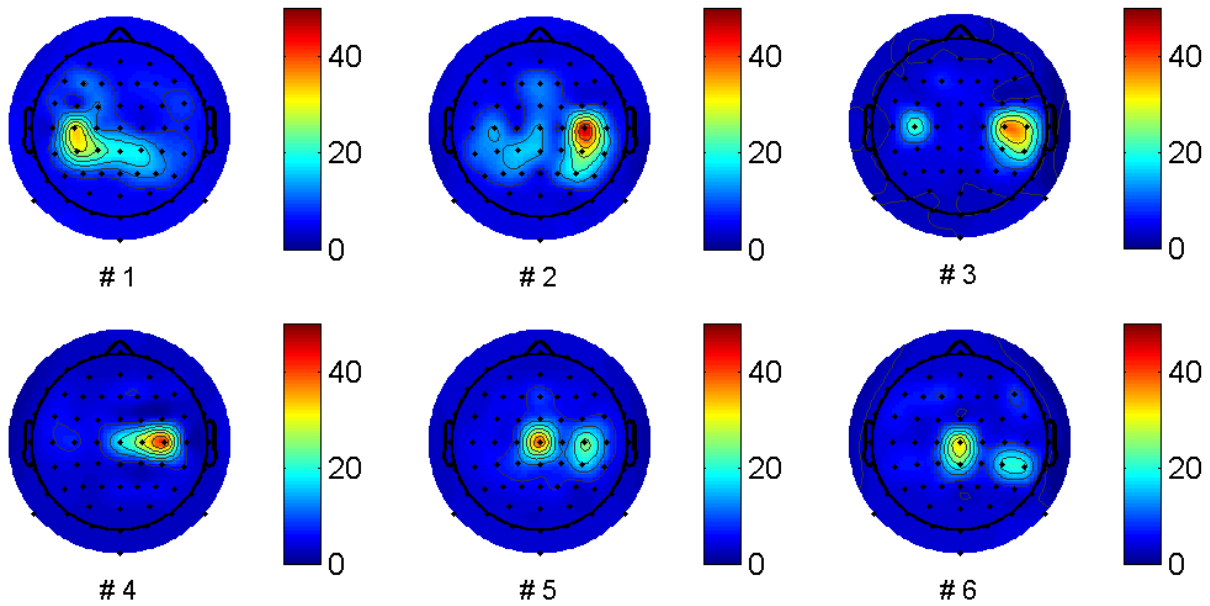


Figure 6.3: Discriminant power of electrodes.

The most relevant electrodes are in the central area (C3, C4 and Cz) according to the ERD/ERD location for hand and foot movement or imagination.

All 6 subjects show classification rates of about 70-75% for motor imagination. These figures were achieved with a relatively low number of features (up to 5 electrodes and up to 6 frequencies) and the general BCI accuracy is 73%. This level of performance can appear relatively low for a 2-class BCI. However, keeping in mind that first all subjects had no prior BCI experience and second that these figures were obtained exclusively in prediction (i.e. classifiers were always tested on new data), the performance is satisfactory.

Table 6.1: Motor imagery classification.

Percentages (mean and standard deviations) of correctly recognized single trials for the 2 motor imagination tasks for the 6 subjects and the average of them. All subjects show classification rates of about 70-75% for motor imagination and the general BCI accuracy is 73%. Features used for classification are also shown.

	Electrodes	Frequencies [Hz]	Left hand [%]	Right foot [%]	Accuracy [%]
# 1*	C3 CP3 CP1 CPz	10 12 14 26	77.2 ± 3.7	70.4 ± 3.2	73.8 ± 4.8
# 2	C4 CP4 P4	10 12 14 18 20 22	71.8 ± 9.0	80.9 ± 7.1	76.4 ± 6.4
# 3	C3 C4 C6 CP6 CP4	14 16 26	76.4 ± 5.8	62.6 ± 6.7	69.5 ± 9.8
# 4	Cz C2 C4	12 14	79.6 ± 1.6	66.3 ± 10.1	73.0 ± 9.4
# 5	Cz C4 CP4	12 24 26	73.5 ± 16.1	71.9 ± 13.3	72.7 ± 1.1
# 6	CPz Cz CP6 CP4	12 14 28 30 32	77.9 ± 7.4	69.0 ± 13.7	73.5 ± 6.3
Avg			76.1 ± 2.9	70.2 ± 6.2	73.1 ± 4.2

* Left foot and Right hand

6.3 Error-related potentials

Figure 6.4 shows the averages of error trials (red curve), of correct trials (green curve) and the difference error-minus-correct (blue curve) for channel FCz for the six subjects. A first small positive peak shows up about ~ 230 ms after the feedback ($t=0$). A negative peak clearly appears ~ 290 ms after the feedback for 5 subjects. This negative peak is followed by a positive peak ~ 350 ms after the feedback. Finally a second broader negative peak occurs about ~ 470 ms after the feedback. The ErrP components appear between 20 and 40 ms later than in the study presented in the previous chapter. This could be explained by the complexity of the task executed by the subjects. Indeed, in Chapter 5, subjects only had to press a key and then they were awaiting the feedback executing no particular task. In our case, subjects focus on a mental task until the feedback appears, so that they are involved in a task that requires a high concentration level. The delay of ~ 30 ms could reflect the time needed by the brain to stop focusing on the mental task and to start processing the visual feedback.

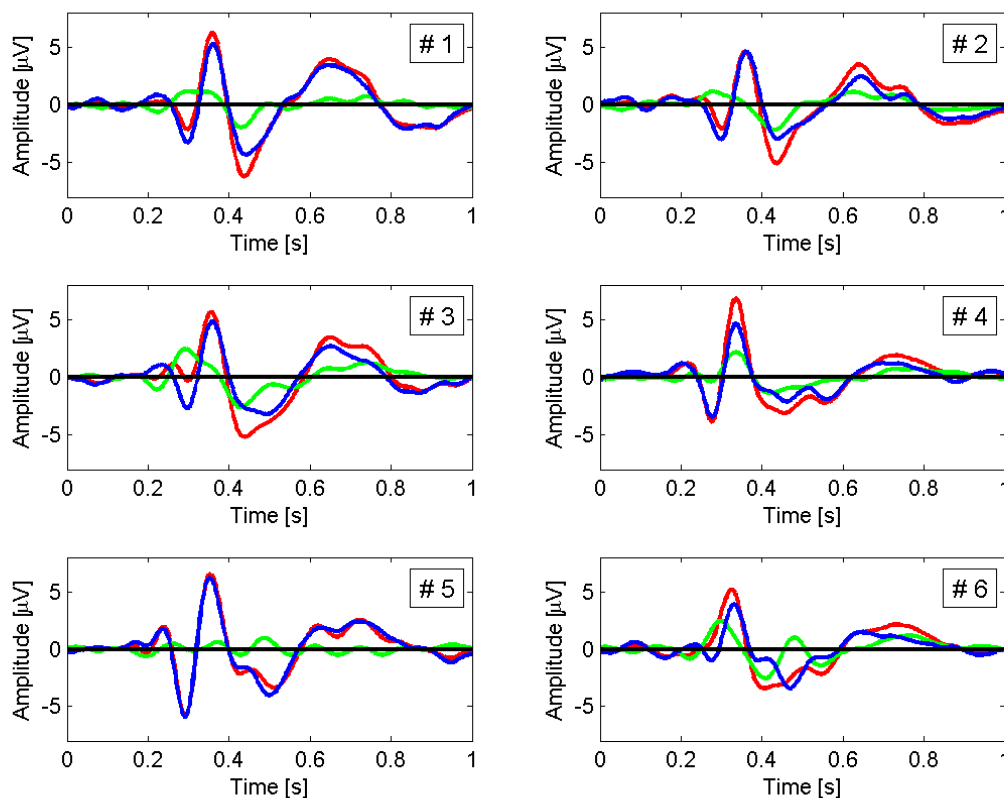


Figure 6.4: Interaction ErrP, 6 subjects.

Averages of error trials (red curve), of correct trials (green curve) and the difference error-minus-correct (blue curve) for channel FCz for the six subjects. All six subjects show similar ErrP time courses whose amplitudes slightly differ from one subject to another.

Figure 6.5 shows the difference error-minus-correct for channel FCz for the six subjects and the average of them (left). It also shows the scalp potentials topographies (right) for the average of the six subjects, at the occurrence of the four previously described peaks: a first

fronto-central positivity appears after ~ 230 ms, followed by a fronto-central negativity at ~ 290 ms, a fronto-central positivity at ~ 350 ms and a fronto-central negativity at ~ 470 ms. All six subjects show very similar ErrP time courses whose amplitudes slightly differ from one subject to the other. The average curves and scalp topographies are very similar to those presented in in Chapter 5, the only difference being the ~ 30 ms delay described above. These experiments confirm the presence of a new kind of error-related potentials generated by an erroneous recognition of the subject's intent. Furthermore, the fronto-central focus at the occurrence of the different peaks again tends to confirm the hypothesis that ErrP are generated in a deep brain region called anterior cingulate cortex [Holroyd and Coles, 2002, Fiehler et al., 2004] (see also Section 6.4).

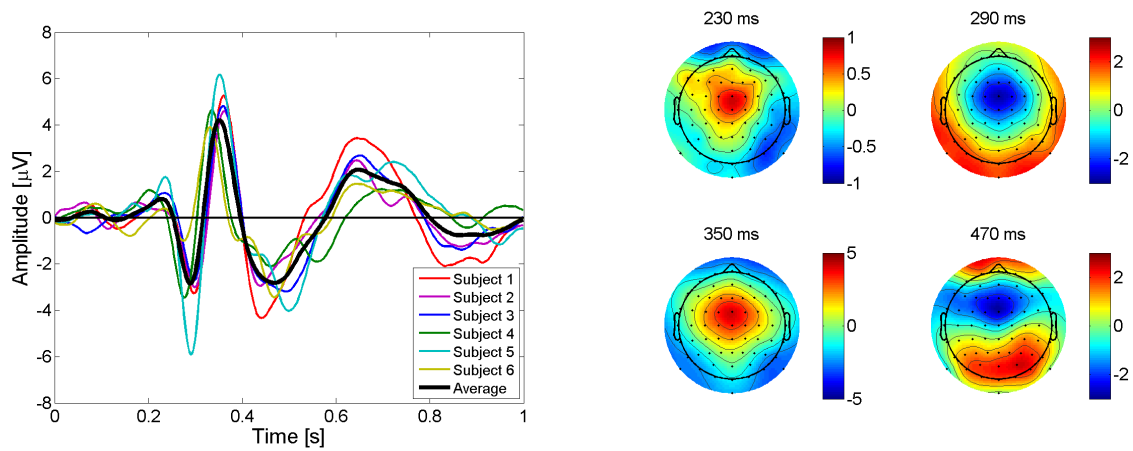


Figure 6.5: Interaction ErrP, average.

(Left) Difference error-minus-correct for channel FCz for the six subjects and the average of them. (Right) Scalp potentials topographies for the average of the six subjects, at the occurrence of the four previously described peaks.

Table 6.2 reports the recognition rates (mean and standard deviations) for the six subjects plus the average of them. These results show that single-trial recognition of erroneous and correct responses are above 75% and 80%, respectively. Beside the crucial importance to integrate ErrP in the BCI in a way that the subject still feels comfortable, for example by reducing as much as possible the rejection of actually correct commands, a key point for the exploitation of the automatic recognition of interaction errors is that they translate into an actual improvement of the performance of the BCI. Table 6.2 also shows the performance of the BCI in terms of bit rate (bits per trial) when detection of ErrP is used or not and the induced increase of performance. The benefit of integrating ErrP detection is obvious since it at least doubles the bit rate for five of the six subjects and the average increase is 124%.

6.4 Estimation of intracranial activity

As in the previous chapters, the sLORETA software was used as a localization tool to estimate the focus of intracranial activity at the occurrence of the four ErrP peaks previously described. Figure 6.6 shows Talairach slices of localized activity for the grand average of the six subjects at the occurrence of the four described peaks and at the occurrence of a late positive component showing up 650 ms after the feedback. As expected, the areas involved in error processing,

Table 6.2: ErrP classification and performance increase.

Percentages (mean and standard deviations) of correctly recognized error trials and correct trials for the six subjects and the average of them. Table also show the BCI performance in terms of bit rate and its increase using ErrP detection. Classification rates are above 75% and 80% for error trials and correct trials, respectively. The benefit of integrating ErrP detection is obvious since it at least doubles the bit rate for five of the six subjects.

	Error [%]	Correct [%]	BCI accuracy [%] (from Table 6.1)	Bit rate [bits/trial] (no ErrP) (ErrP)		Increase [%]
# 1	77.7 ± 13.9	76.8 ± 5.4	73.8 ± 4.8	0.170	0.345	103
# 2	75.4 ± 5.5	80.1 ± 7.9	76.4 ± 6.4	0.212	0.385	82
# 3	74.0 ± 12.9	85.9 ± 1.6	69.5 ± 9.8	0.113	0.324	187
# 4	84.3 ± 7.7	80.1 ± 5.5	73.0 ± 9.4	0.159	0.403	154
# 5	75.3 ± 6.0	85.6 ± 5.2	72.7 ± 1.1	0.154	0.371	141
# 6	70.7 ± 11.4	82.2 ± 5.1	73.5 ± 6.3	0.166	0.333	101
Avg	76.2 ± 4.6	81.8 ± 3.5	73.1 ± 4.2	0.160	0.359	124

namely the pre-supplementary motor area (pre-SMA, Brodmann area 6) and the rostral cingulate zone (RCZ, Brodmann areas 24 & 32) are systematically activated [Holroyd and Coles, 2002, Fiehler et al., 2004]. Indeed, for the first positive peak (230 ms after the feedback), the focus of activity is located at X=15mm, Y=-10mm, Z=60mm (MNI coordinates). The best match is Brodmann area 6 (pre-supplementary motor area). For the first negative peak (290 ms after the feedback), the focus is at X=0mm, Y=-10mm, Z=35mm and the best match is Brodmann area 24 (ventral anterior cingulate cortex). For the second positive peak (350 ms after the feedback), the focus is at X=0mm, Y=-5mm, Z=40mm and the best match is again Brodmann area 24. Finally, for the second negative peak (470 ms after the feedback), the focus is at X=5mm, Y=-10mm, Z=55mm and the best match is Brodmann area 6. For the late positive component (650 ms after the feedback), parietal areas are activated. The focus is at X=0mm, Y=-65mm, Z=45mm and the best match is Brodmann area 7 (somatosensory association cortex). This associative area could be related to the fact that the subject becomes aware of the error. Indeed, it has been proposed that the positive peak was associated with conscious error recognition in case of error potentials elicited in reaction task paradigm [Nieuwenhuis et al., 2001]. These localization are very similar to those obtained in the previous chapter, the best match is always either in the ACC or in the pre-SMA. However, in the present study, Brodmann area 32 (dorsal anterior cingulate cortex) is never the best match. This could confirm the fact that with this moving cursor protocol, the localization are a little more posterior. As already described in Chapter 5, the posterior cingulate cortex (PCC) is involved in anticipatory allocation of spatial attention [Kim et al., 1999, Mesulam et al., 2001], so that the PCC could have a different activity during correct trials and erroneous trials. However, since Brodmann areas 23 & 31 are never the best match, it seems that the activity of the PCC is not as relevant as the activity of the ACC and the pre-SMA. The involvement of the ACC and the pre-SMA in error processing and the involvement of the PCC in anticipation could explain the slightly more posterior foci, at the limits of these 3 cortical areas.

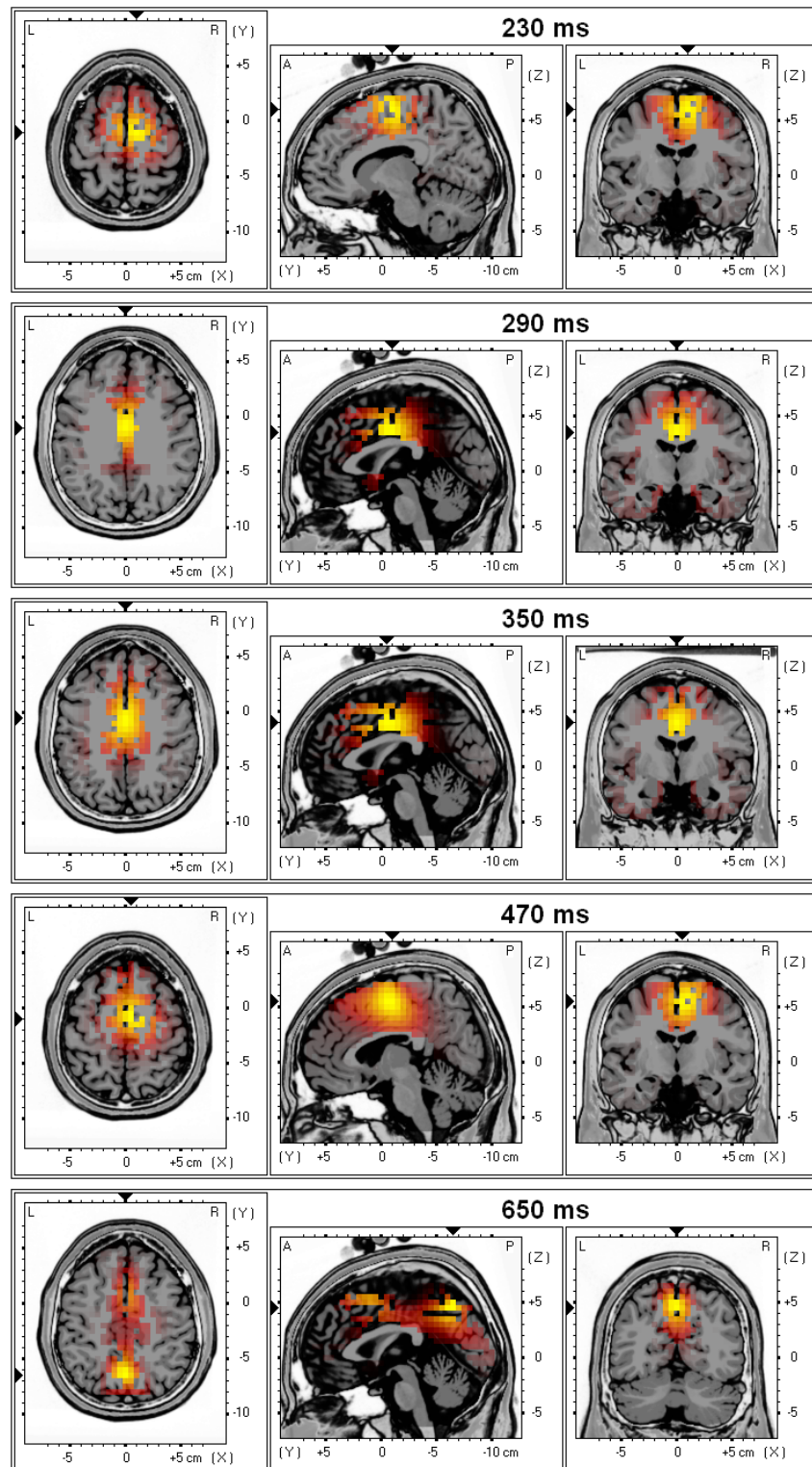


Figure 6.6: Localization of interaction ErrP.

Talairach slices of localized activity for the grand average of the six subjects at the occurrence of the four described peaks and at the occurrence of a late positive component showing up 650 ms after the feedback.

6.5 Conclusion

In this study we have reported results on the detection of the neural correlate of error awareness for improving the performance and reliability of BCI. In particular, we have confirmed the existence of a new kind of error-related potential elicited in reaction to an erroneous recognition of the subject's intention. More importantly, we have shown the feasibility of simultaneously and satisfactorily detecting erroneous responses of the interface and classifying motor imagination for device control at the level of single trials. In this study we have also shown that, as expected, typical cortical areas involved in error processing such as pre-supplementary motor area and anterior cingulate cortex are systematically activated at the occurrence of the different peaks.

This will be confirmed in next chapter, but the timing used in the proposed protocol allows us to perform motor imagery and ErrP detection and to perform a movement of the cursor every 2 seconds. Furthermore, the window (1 second) used for motor imagery could probably be shortened and therefore the speed of the interaction could even be increased. In the next chapter, we report online results with two subjects who successfully controlled the cursor with motor imagery and ErrP were detected after each step.

Chapter 7

Online validation

In this chapter we close the loop. In Chapter 6 we have shown the feasibility of simultaneously and satisfactorily detecting erroneous responses of the interface and classifying motor imagination for device control at the level of single trials. But the study presented in Chapter 6 was an off-line analysis, the final validation presented in this chapter consists in an online BCI integrating ErrP detection to filter out wrong decisions and thus increase the BCI performances.

7.1 Experimental setup

Two among the six healthy volunteer subjects that took part to the experiments described in the previous chapter (subjects 4 & 5) participated in these online experiments. The task is the same as in Chapter 6, beside the fact that the system is moving the cursor following the mental commands of the subject and is cancelling this movement if it is eliciting ErrP. In this human-computer interaction, the subject has to bring the cursor to targets 3 steps either to the left or to the right, according to the position of the target. Each target corresponds to a specific mental task, namely imagination of a movement of the left hand for the left target and imagination of a movement of the right foot for the right target.

After the presentation of the target, the subject focuses on the corresponding mental task until the cursor moves. The system uses a window of 1 second to determine the subject's intent. Then the system uses a 400 ms window to detect the presence of ErrP just after the presentation of the feedback (movement of the cursor). If no ErrP are detected, nothing happens and about 600 ms later, the system starts to accumulate data for the next classification of motor imagery. If ErrP are detected, the movement is cancelled, and again after about 600 ms the system starts accumulating data for next step. Figure 7.1 illustrates this timing. At $t=0$, the target (red) is 3 steps on the right of the cursor (green). The subject is therefore imagining a movement of his right foot, and the system waits until 1 second of data is available for motor imagery classification. At $t=1$ second, the system moves the cursor following the classification of motor imagery. In this example, the classifier recognizes left when the subject was focusing on the right task, i.e. the system makes a mistake when detecting the subject's intent. Right after this feedback, the system waits 400 ms to perform the ErrP detection. At $t=1.4$ second, the system cancels the movement if ErrP were detected. In this example, the system detects ErrP and cancels the wrong movement. It is to note that the system is only cancelling last movement, not replacing the wrong command (right) by the opposite one (left). After a delay of about 600 ms, the system starts accumulating data for the next motor imagery classification.

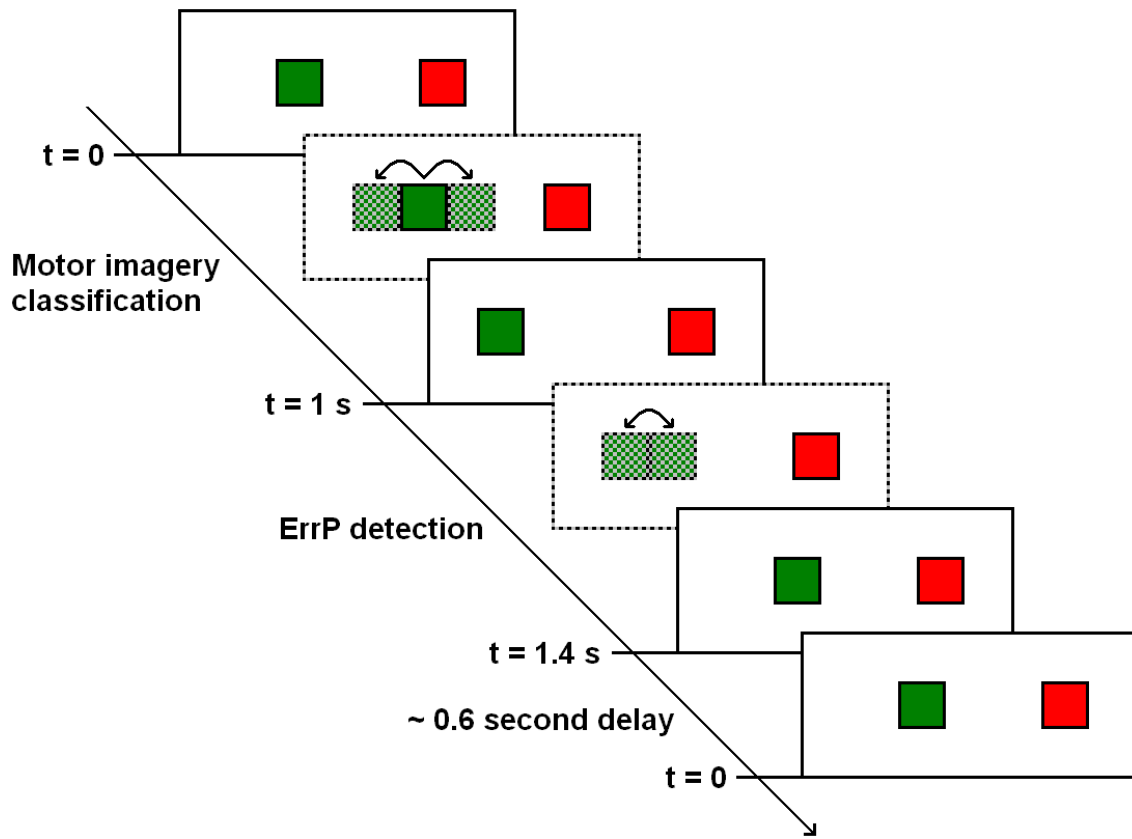


Figure 7.1: Timing of the online protocol.

At $t=0$, the target (red) is 3 steps on the right of the cursor (green). At $t=1$ second, the system moves the cursor following the classification of motor imagery. In this example the system makes a mistake (left movement instead of right movement). At $t=1.4$ second, the system cancels the movement if ErrP were detected. In this example, the system detects ErrP and cancels the wrong movement. After a delay of about 600 ms, the system starts accumulating data for the next motor imagery classification.

In any case, the cursor is moving on average every 2 seconds, and some movements are cancelled if ErrP are detected. When the cursor reached a target, it briefly turned from green to light green and then a new target was randomly selected by the system. If the cursor didn't reach the target after 10 steps, a new target was selected. Subjects performed 10 sessions of 15 targets on 2 different days, the delay between the two days of measurements was about 2 weeks. The 20 sessions were split into 4 groups of 5. For the first day (groups I & II) we used classifiers build with data recorded for the analysis presented in previous chapter, and for the second day (groups III & IV) we used the data of the first day to build classifiers. This rule applies for both motor imagery classification and for ErrP detection. The data acquisition and processing as well as the classification procedures are the same as in Section 6.1. For both subjects we used a 150 ms window starting 250 ms after the feedback for channels FCz and Cz for ErrP detection. For motor imagery classification, we used the features reported in Chapter 6, namely Cz, C2, C4 and 12 Hz, 14 Hz for the first subject and Cz, C4, CP4 and 12 Hz, 24 Hz, 26 Hz for the second subject.

7.2 Performances

For both subjects, Table 7.1 shows the classification rates for ErrP detection (error and correct) for the four groups of recordings and for the average of them. It also shows the error rates and the rejection rates for motor imagery, with and without ErrP detection. Finally the increase in performance expressed in bits per rials (BpT) is also shown.

Table 7.1: Classification rates and performance increase.

For both subjects, this table presents the classification rates for ErrP detection (error and correct) for the four groups of recordings and for the average of them. It also shows the error rates and the rejection rates for motor imagery, with and without ErrP detection. Finally the increase in performance is also shown. The ErrP detection rate is once again around 80% and the error rate of the standard BCI is around 30%. When integrating ErrP detection, this error rate is below 10% with an acceptable rejection rate of 30-35%. Finally, for both subjects the bit rate is multiplied by more than 3 when using ErrP detection.

Subject 1 (Cz, C2, C4 and 12 Hz, 14 Hz)		I	II	III	IV	Average	S.D.
ErrP detection	Error [%]	74.8	83.7	76.1	67.5	75.5	6.6
	Correct [%]	88.3	82.1	91.1	81.6	85.8	4.7
BCI without ErrP	Error rate [%]	33.0	27.5	34.3	39.0	33.5	4.7
	Rejection rate [%]	0.0	0.0	0.0	0.0	0.0	0.0
BCI with ErrP	Error rate [%]	8.3	4.5	8.2	12.7	8.4	3.4
	Rejection rate [%]	32.5	36.0	32.0	37.6	34.5	2.7
Performance	BpT initial	0.09	0.15	0.07	0.04	0.09	0.05
	BpT final	0.31	0.41	0.32	0.17	0.30	0.10
	Increase [%]	244	173	357	325	275	83

Subject 2 (Cz, C4, CP4 and 12 Hz, 24 Hz, 26 Hz)		I	II	III	IV	Average	S.D.
ErrP detection	Error [%]	94.8	76.6	76.5	80.2	82.0	8.7
	Correct [%]	68.0	88.5	86.1	91.4	83.5	10.6
BCI without ErrP	Error rate [%]	31.3	30.2	31.1	29.2	30.5	1.0
	Rejection rate [%]	0.0	0.0	0.0	0.0	0.0	0.0
BCI with ErrP	Error rate [%]	1.6	7.6	7.6	5.8	5.7	2.8
	Rejection rate [%]	51.6	32.5	33.1	29.5	36.7	10.1
Performance	BpT initial	0.10	0.12	0.11	0.13	0.12	0.01
	BpT final	0.38	0.36	0.33	0.42	0.37	0.04
	Increase [%]	280	200	200	223	226	38

For both subjects the ErrP detection rate is again around 80% and pretty stable over the different groups of recordings. Without the use of ErrP detection, the first subject shows a stable error rate of 34% for motor imagery, whereas for the second subject this rate is just above 30%. These rates are pretty high for a 2 tasks BCI, but keeping in mind that the subjects have very little BCI experience and that these are real online experiments, they are satisfactory. When

integrating ErrP detection, the error rates drop below 10% for both subjects with acceptable rejection rates around 35%. This is a first proof of the benefit of using ErrP detection to filter out wrong decisions. This benefit is clear in term of performance, indeed, the bit rate is multiplied by more than 3 for both subjects when using ErrP detection.

ErrP detection strongly interferes with the BCI so that it's tricky to really evaluate the gain in performance in terms of number of steps saved to reach the target. Indeed, ErrP detection can cancel errors and therefore decrease the number of steps required to reach the targets. In this case the BCI performance is increased, both because the direct effect of ErrP detection and because the subject feels comfortable and has the feeling of a good control. However, ErrP detection can also block correct commands (false negative) or let wrong commands through (false positive). In this case the BCI performance can be decreased both because of ErrP detection and because the frustration of the subject. In other words, if ErrP detection is usually directly and indirectly benefic, it can sometimes have bad effects.

Nevertheless, we can estimate the number of steps saved to reach the targets as following. On average, the two subjects showed a BCI accuracy without the integration of ErrP of 68% with no rejection. With the integration of ErrP, the average accuracy was 57.3% with a rejection rate of 35.6%. The two subjects were asked to reach 15 targets located 3 steps away from their initial position. This means that 45 pure correct movements are required. Then, depending on the error rates, additional correct movements are required to compensate the errors. Based on these observations and if x is the number of erroneous movements and y is the number of rejected movements, the total number of steps n required to reach 15 targets is:

$$\begin{aligned} n &= 45 + 2x && \text{without ErrP detection} \\ n &= 45 + 2x + y && \text{with ErrP detection} \end{aligned}$$

The values of n , x and y can be estimated using the average performances reported above. Table 7.2 summarizes these figures and reports the number of steps for both conditions, i.e. with and without ErrP detection. On average, the introduction of ErrP detection reduces by more than 2 the number of steps required to reach each target.

Table 7.2: Performances in term of number of steps.

The table shows the averaged BCI performance of both subjects. It also shows the number of steps required to reach 15 targets. The integration of ErrP detection reduces by more than 2 the number of steps required to reach a target.

	no ErrP detection	ErrP detection
BCI accuracy	68%	57.3%
Error rate	32%	7.1%
Rejection rate	0%	35.6%
Correct movements	85	51.5
Erroneous movements	40	6.5
Rejected movements	0	32
Total movements	125	90
Movements per target	8.3	6

7.3 Feature stability

In this section, we show the relative stability of the features used for classification of both motor imagery and ErrP detection for the two subjects. As expected, the features used for ErrP detection are slightly more stable than the features used for motor imagery. This translates in the classification rates presented in the previous section. Indeed, ErrP classification is on average above 80% whereas, without ErrP integration, motor imagery classification is under 70%. These classification rates are very similar to those obtained off-line in Chapter 6.

7.3.1 Motor imagery

We study here the stability of the features used for motor imagery classification, namely electrodes and frequency bands. Figure 7.2 shows the discriminant power (calculated with the DP algorithm described in Section 2.7.3) of the frequency bands (top) and of the EEG electrodes (bottom) for the four groups of online recordings for the first subject. Figure 7.3 shows the same plots for the second subject.

The selected features for the first subject were electrodes Cz, C2 and C4 and frequency bands 12 and 14 Hz. Figure 7.2 shows that even if the variability is relatively important, electrode C4 is always the most relevant one, and the frequencies between 10 and 15 Hz always show the best discriminant power. For the third group of recordings, higher frequency bands around 26 Hz (beta band) also show a good discriminant power. It is interesting to note that for this group of recordings, the focus on C4 is not as clear as for the other groups. Indeed, right frontal electrodes as well as electrodes around P3 show some discriminant power. For the second subject, the selected features were electrodes Cz, C4 and CP4 and frequency bands 12, 24 and 26 Hz. Again, even if the variability is relatively important, Figure 7.3 shows that electrode C4 is systematically relevant for classification and the selected frequency bands show the best discriminant power. The focus on Cz is less evident, it is not present for the second group of recordings and located under C1 for the third group and under FC1 for the fourth group.

7.3.2 Interaction ErrP

In this section, we focus on the stability of the features used for ErrP detection. Figure 7.4 shows the average of error and correct trials plus the difference error-minus-correct for channel FCz for the four groups of online recordings for the first subject (top) and for the second subject (bottom). Classification is based on a window including the 250 ms negative peak and the 330 ms positive peak. For both subjects these two peaks are very stable from a group of recordings to another, only small variations in amplitudes can be seen whereas the time at which they occur is very stable. This explains the fact that ErrP detection reaches higher classification rates than motor imagery. Furthermore, the potentials are very similar for both subjects. We could therefore imagine to use a general classifier for all subjects for ErrP detection, and no more a specific classifier for each subject. This classifier would be built with the data of several subjects. If the use of a general classifier for all subjects leads to sufficient classification rates, this would simplify our procedure. Indeed, in this specific case, preliminary recordings to build ErrP classifiers are not necessary anymore. It is to note that this procedure of a single general classifier could not be applied to motor imagery because the most discriminant features are different for all subjects as shown in Chapter 6.

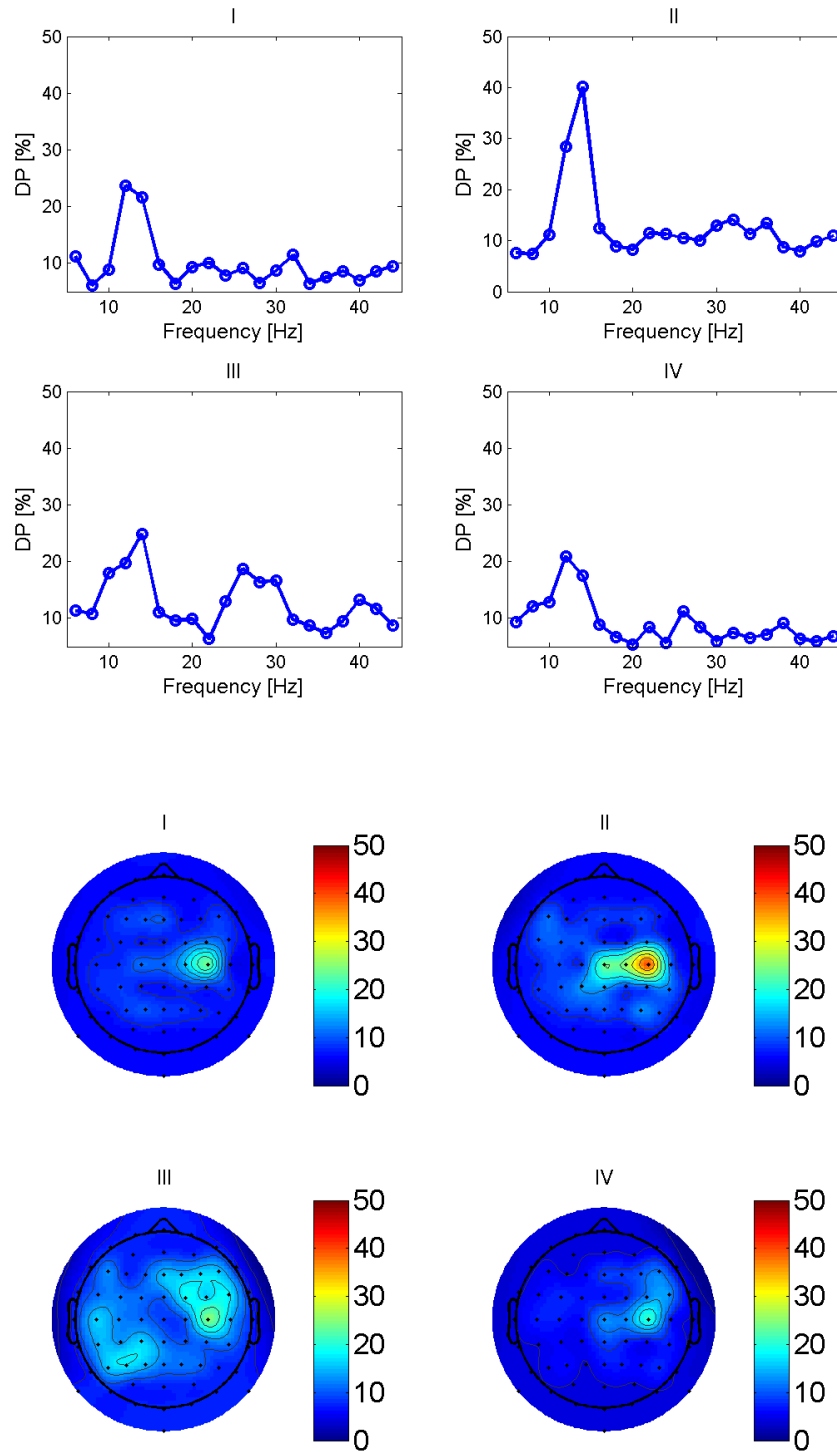


Figure 7.2: Stability of features for the first subject.

Discriminant power of the frequency bands (top) and of the EEG electrodes (bottom) for the four groups of online recordings for the first subject. The most relevant features match relatively well with those used according to the study of Chapter 6 (electrodes Cz, C2 and C4 and frequency bands 12 and 14 Hz).

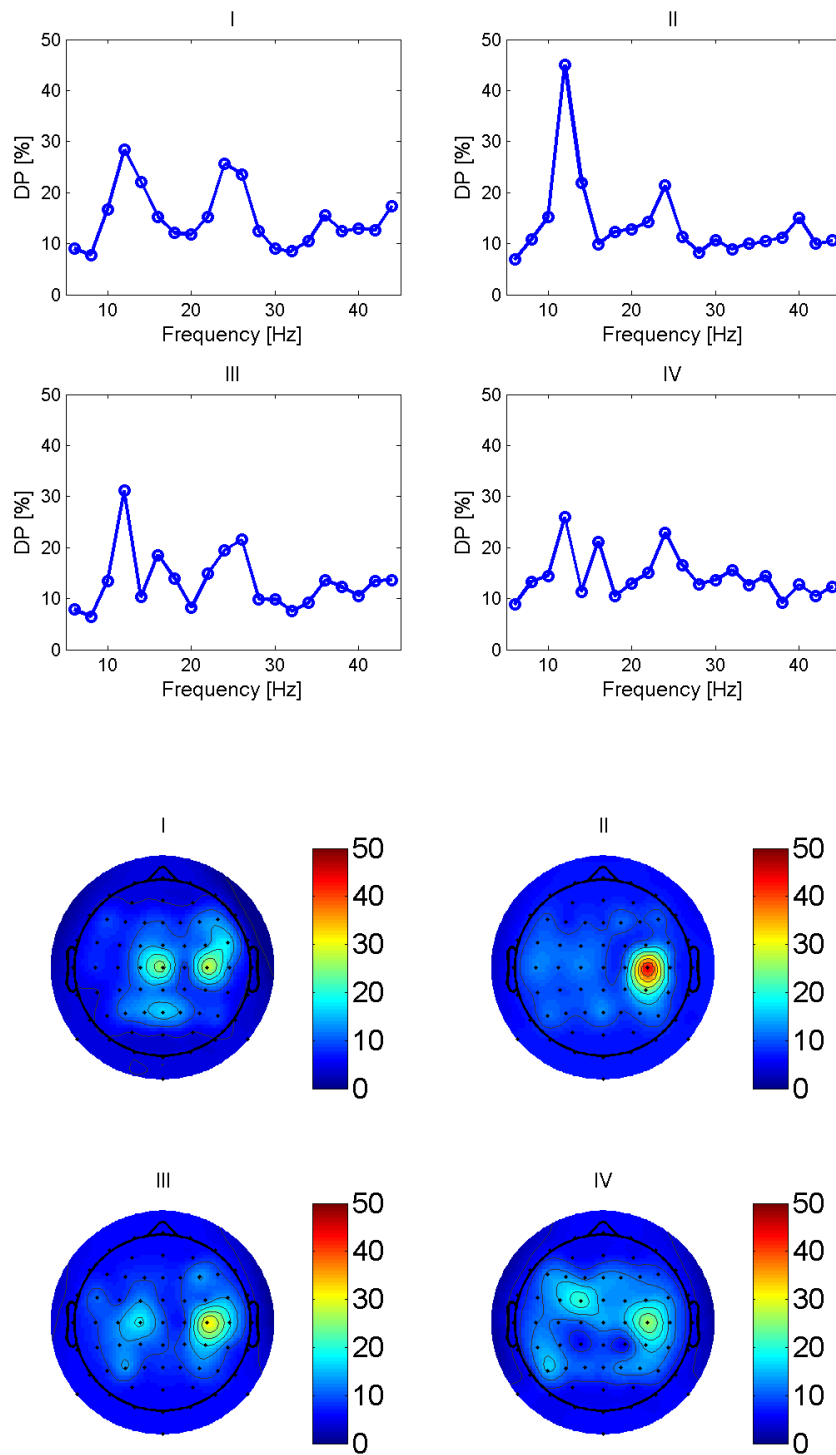


Figure 7.3: Stability of features for the second subject.

Discriminant power of the frequency bands (top) and of the EEG electrodes (bottom) for the four groups of online recordings for the second subject. The most relevant features match relatively well with those used according to the study of Chapter 6 (electrodes Cz, C4 and CP4 and frequency bands 12, 24 and 26 Hz).

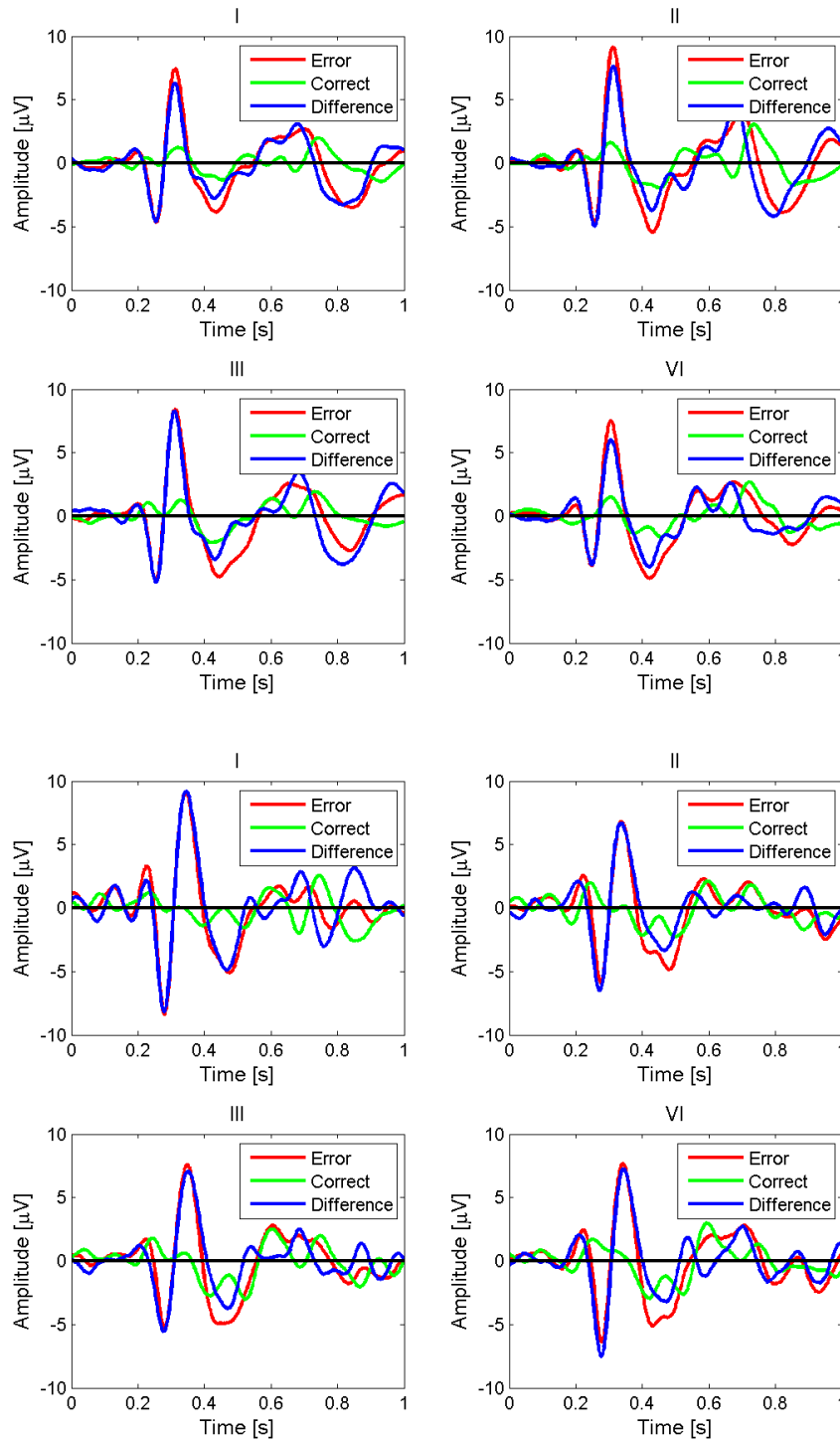


Figure 7.4: Stability of features for ErrP detection.

Average of error and correct trials plus the difference error-minus-correct for channel FCz for the four groups of online recordings for the first subject (top) and for the second subject (bottom). ErrP appear to be very stable over time, and very similar for both subjects.

7.4 Conclusion

In this chapter we have closed the loop using the protocol described in Chapter 6 for online experimentations, i.e. with statistical classifiers for motor imagery and ErrP detection running in real time and simultaneously. Two subjects were able to control the cursor using motor imagery with an average accuracy just below 70%. In parallel, the system was able to detect the presence of ErrP with an accuracy above 80% to improve the quality of the brain-computer interaction. Indeed, the integration of ErrP detection reduces the number of steps required to reach the target, and in terms of bit rate, the performance is multiplied by a factor 3. The features used for classification were those selected in Chapter 6. They show a relatively good stability, in particular the potentials used for ErrP detection. More generally, the ErrP potentials described in the chapters of this thesis are relatively similar for all subjects. We could therefore maybe build a general ErrP classifier that we would use for all subjects. This would simplify the training sessions, since no preliminary ErrP recordings to build classifiers would be needed anymore. The duration of the window used for motor imagery classification was 1 second. This could probably be shortened to 0.5 second or maybe even less without decreasing performances, so that if we reduce the delay after ErrP detection, we could be able to deliver a feedback almost every second.

Chapter 8

Integration of inverse models

In this chapter we report results on the use of inverse models to improve the detection of error-related potentials. The main goal of this study is to investigate whether inverse solutions such as the Cortical Current Density (CCD) and the estimated Local Field Potentials (eLFP) are useful tools to increase ErrP single trial classification. The localization of the source(s) responsible for ErrP generation is an interesting point, but it's not the main goal of this study. The aim of this study is to test the performances of the models when used for ErrP classification. The evaluation procedure is therefore very similar to the evaluation procedure of BCI performances. In this context, localization is only used to interpret classification results and is not a figure of merit in itself. Investigations were made in a brain-computer interface (BCI) framework to assess off-line the potential benefit of using an inverse solution for online ErrP single trial detection.

8.1 Experimental setup

Ten healthy volunteer subjects participated in this study. The task (previously described in Chapter 5) was to send manual commands to bring a cursor to a target, either to the left or to the right, and the system moved the cursor with an error rate of 20%; i.e., at each step, there was a 20% probability that the cursor moved in the opposite direction. Subjects performed 10 sessions of 3 minutes, corresponding to ~ 75 single trials per session. Five subjects did a single day of recordings (10 sessions) and the other five did two days of recordings (20 sessions in total).

The data processing procedure described in Section 2.7.2 was also applied for this study. The actual input vector for the statistical classifier was again a 150 ms window starting 250 ms after the feedback and ending 400 ms after the feedback. Channel selection was performed using the feature selection algorithm (DP) described in Section 2.7.3. For EEG, we selected again the best two electrodes, namely FCz and Cz, according to the feature selection results obtained in Section 5.3. For both CCD and eLFP analysis, the 10 most relevant solution points based on the feature selection algorithm ranking were selected. We showed in a previous study that the best classification rates are obtained with a relatively small number of solution points [Uldry et al., 2007]. It appears that adding more solution points doesn't improve the classification and in some cases it even decreases the performance by adding more noise.

Feature selection was performed on the 5 first sessions of all 10 subjects together, so that the 10 selected solution points are the same for all subjects. The same points were selected

for all subjects for comparison with EEG purposes. Indeed, the same electrodes are selected for all subjects, so we decided to select the same solution points for all subjects. Furthermore, and this was already mentioned, selecting the same solution points for all subjects will simplify the online procedure since no feature selection will be needed anymore. For each subject, the Gaussian classifier was trained using the 5 first sessions and performances were tested on the 5 last sessions. For the five subjects for whom a second day of recordings was available, we extended this procedure to the 20 sessions. The classifier was trained on 5 sessions, and tested on the next 5 sessions, leading for these subjects to 3 classification results that we averaged. In a BCI framework, this predictive procedure is much more realistic than cross-validation or leave-one-out because performances are evaluated on data unknown to the classifier and recorded after those used for the training of the classifiers. The data separation in 2 or 4 groups was performed to test stability of features because stability, more than very high classification rates, is a key issue for successful online BCI.

8.2 Grand averages

Figure 8.1 shows the average EEG for error and correct trials and the difference error-minus-correct at channel FCz for the grand average of the ten subjects for the 5 first sessions (top) and for the 5 last sessions (bottom) with an error rate of 20%. Feedback is delivered at time 0 seconds. A first positive peak shows up 200 ms after the feedback. A negative peak and a positive peak show up about 250 ms and 320 ms after the feedback, respectively. A second broader negative peak occurs after about 450 ms and is followed by a positive peak about 600 ms after the feedback. Averages are very similar for the two groups of recordings, only small changes in amplitudes can be observed whereas peaks occur at the same time for both groups.

Figure 8.2 shows the scalp potential topographies for the grand average EEG of the ten subjects, at the occurrence of the 5 described peaks for the 5 first sessions (top) and for the 5 last sessions (bottom). A fronto-central focus can be observed for all 5 peaks. As for the potentials, averages are very similar for the two groups of recordings, only small variations can be observed.

The reported potentials are once again very similar to the ones reported in the previous chapters of this thesis. The averages over ten subjects clearly filter out *noisy* potentials, and it seems that there are 5 main components in our so-called interaction ErrP. This tends to prove that we have a new kind of ErrP, since no other study reported so many ErrP components. In particular, the second negative peak (about 450 ms after the feedback) is the main component in terms of amplitude, but no study ever mentioned the presence of a second negative peak in any kind of ErrP. Another point to note is that the window used for classification only includes the first negative peak (250 ms after the feedback) and the second positive peak (320 ms after the feedback). There are 2 main reasons for that. The first reason is to try to shorten the window so that the ErrP detection can be performed as soon as possible after the feedback. Indeed, the detection can be performed now in 400 ms, whereas in the first studies, we had to wait until 650 ms. The second reason is the *quality* of the potentials. Indeed, those 2 peaks show lower variances and are therefore better candidates for stable classification performances. Furthermore, these 2 peaks only show foci on a reduced fronto-central area. The 3 other peaks also show foci on frontal and/or parietal areas. Therefore, selecting those 2 peaks probably excludes any systematical influence of both ocular or visual artifacts.

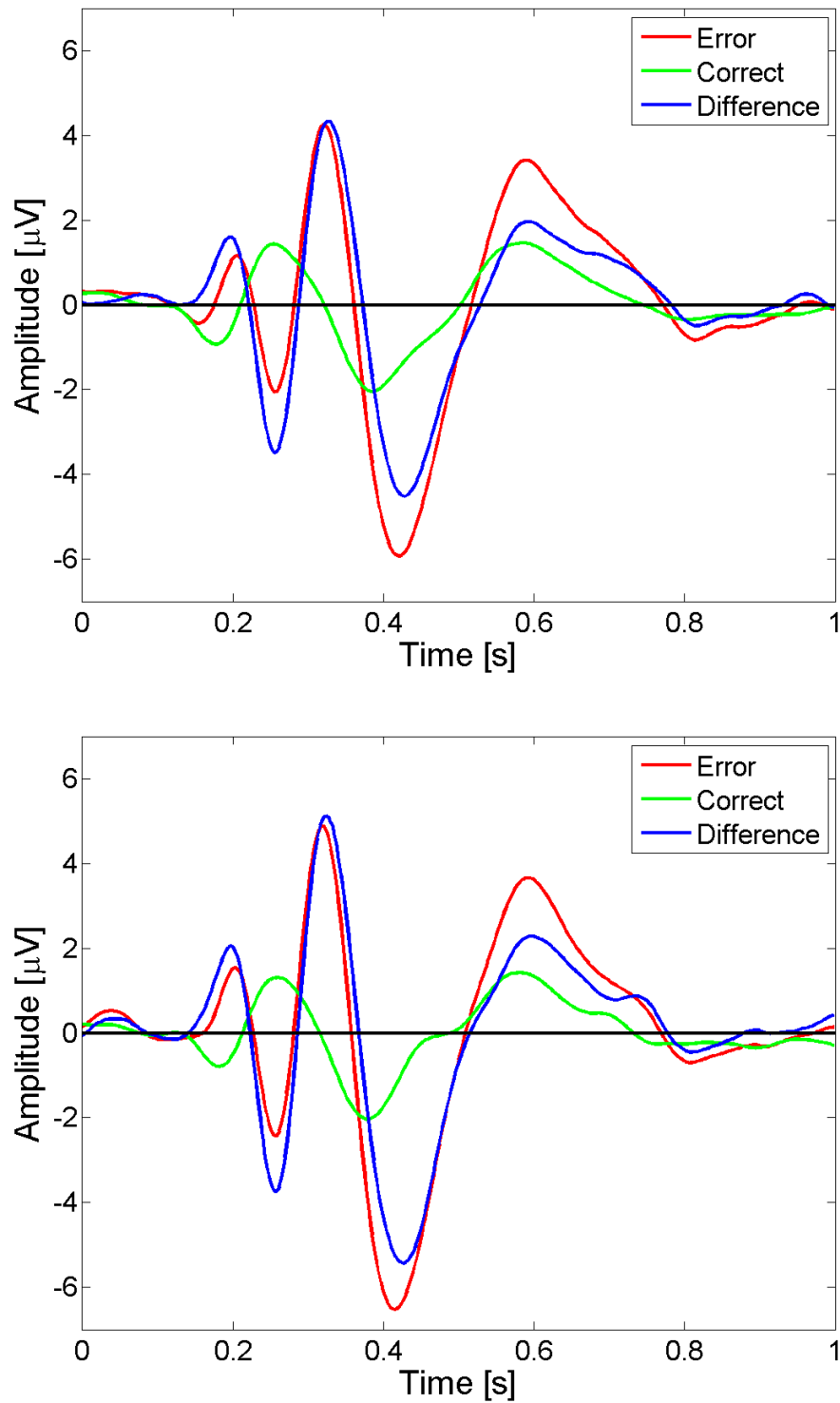


Figure 8.1: Interaction ErrP, 10 subjects.

Average EEG for error and correct trials and the difference error-minus-correct at channel FCz for the grand average of the ten subjects for the 5 first sessions (top) and for the 5 last sessions (bottom) with an error rate of 20%.

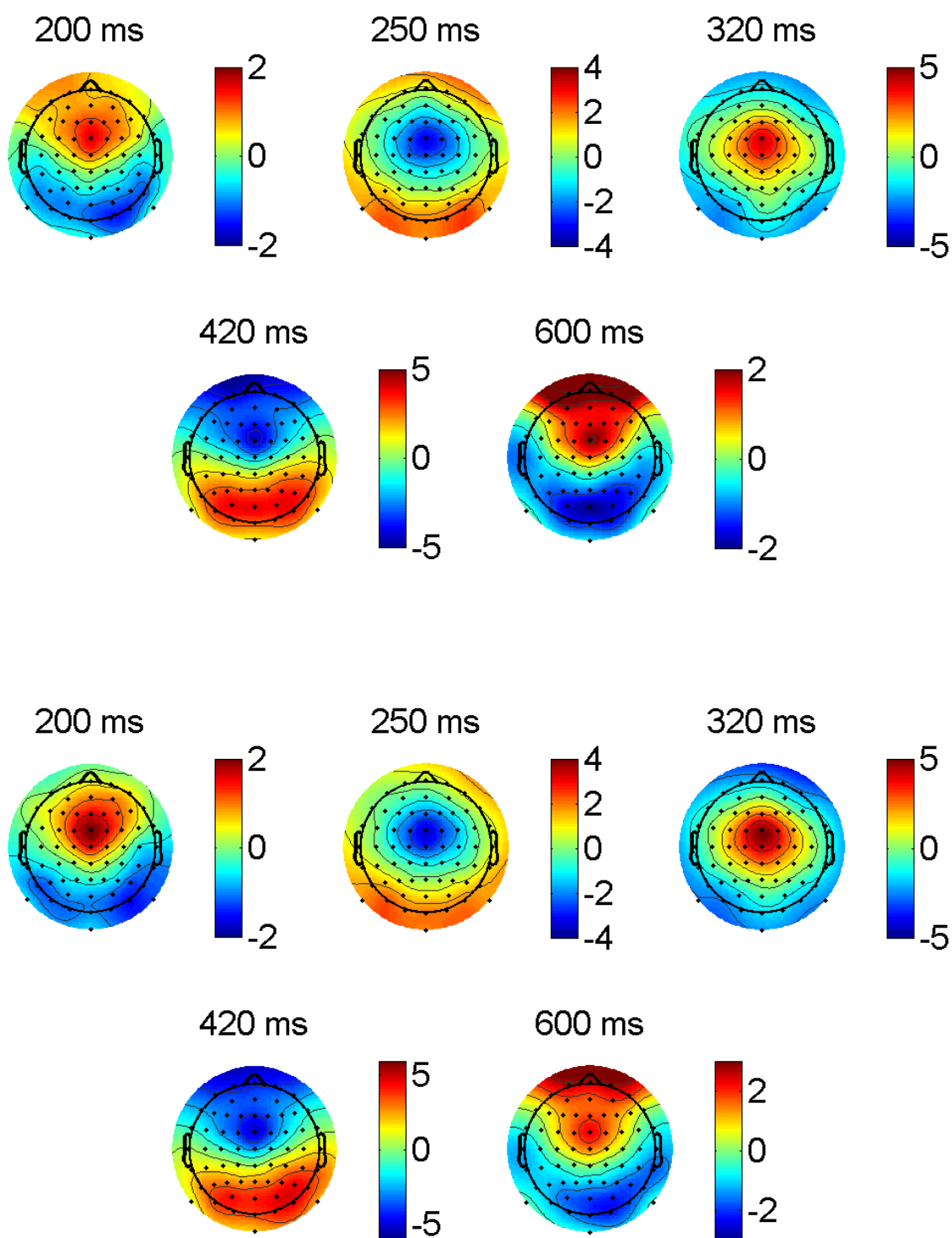


Figure 8.2: Scalp potentials topographies.

Scalp potential topographies for the grand average EEG of the ten subjects, at the occurrence of the 5 described peaks for the 5 first sessions (top) and for the 5 last sessions (bottom). A fronto-central focus can be observed for all 5 peaks.

8.3 Cortical current density (CCD) model

Figure 8.3 shows the 50 most relevant solution points of the CCD model for ErrP classification for the average of the ten subjects for the 5 first sessions (top) and for the 5 last sessions (bottom). Selection is very stable over the 2 groups of recordings, only a small proportion of points differ from a session to the other. Indeed, 78% of the solution points are common to the 2 groups. We show the 50 most relevant points for sake of visibility, since plotting only the 10 best points leads to very small clusters, sometimes quite hard to see. The average of the ten subjects was simply obtained by averaging the discriminant power (DP) score of each solution point over the ten subjects. Compact clusters partially localized on the pre-SMA (Brodmann area 6) and on the ventral anterior cingulate cortex (Brodmann area 24) can be seen on both hemispheres. Partial localization of the deep cluster on the posterior cingulate cortex (PCC, BA 23 & 31) could possibly be explained by the fact that the PCC is involved with the medial prefrontal cortex in anticipatory allocation of spatial attention [Kim et al., 1999, Mesulam et al., 2001]. Indeed, it's possible that the PCC has different activity when feedback correspond to anticipation (correct trials) and when feedback doesn't correspond to anticipation (erroneous trials). The anterior cingulate cortex mediates aspects such as performance monitoring whereas the posterior cingulate cortex is associated with the speed of detecting spatial targets, especially when attention is under the influence of a cue-induced anticipatory bias. However, the nature of the cingulate contribution to spatial attention is still not clear [Mesulam et al., 2001].

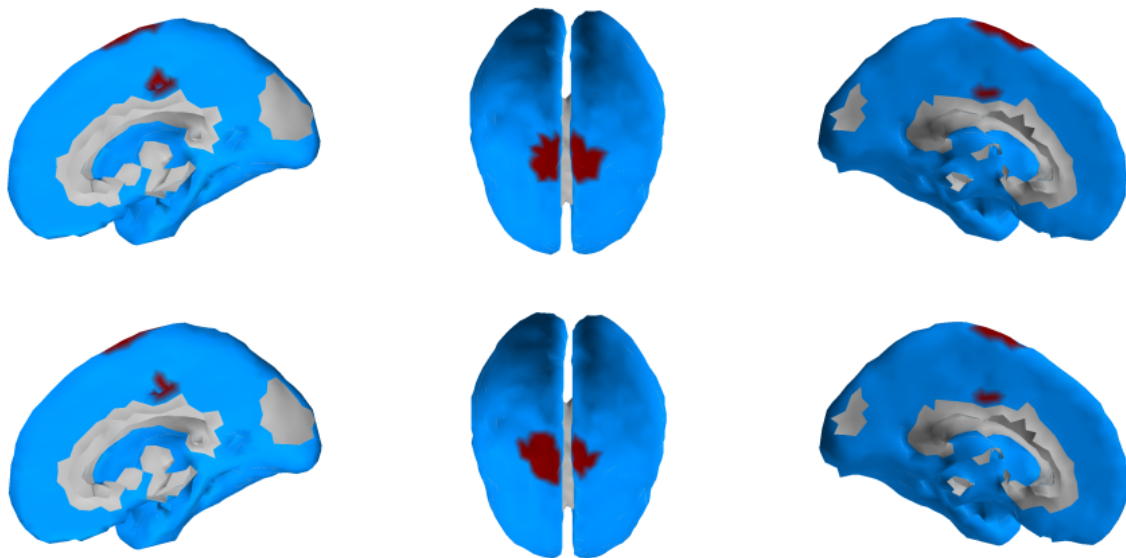


Figure 8.3: Most relevant solution points for the CCD, 10 subjects. Localization of the 50 most relevant solution points of the CCD model for ErrP classification for the average of the ten subjects for the 5 first sessions (top) and for the 5 last sessions (bottom). Clear clusters can be seen and localization is very stable from a group of sessions to the other, 78% of the points are common to the 2 groups.

Figure 8.4 shows the 50 most relevant solution points of the CCD model for ErrP classification for the average of the five subjects for the four groups of recordings (top to bottom). Since there were only five subjects in this parallel study, we can see a higher variability in the

localization of the 50 most relevant solution points. However, the same clusters can be seen for all 4 groups, only a very few number of points are outside the clusters. As for the analysis with ten subjects, the clusters are located over the pre-SMA and over small areas at the limit of the anterior and the posterior cingulate cortex.

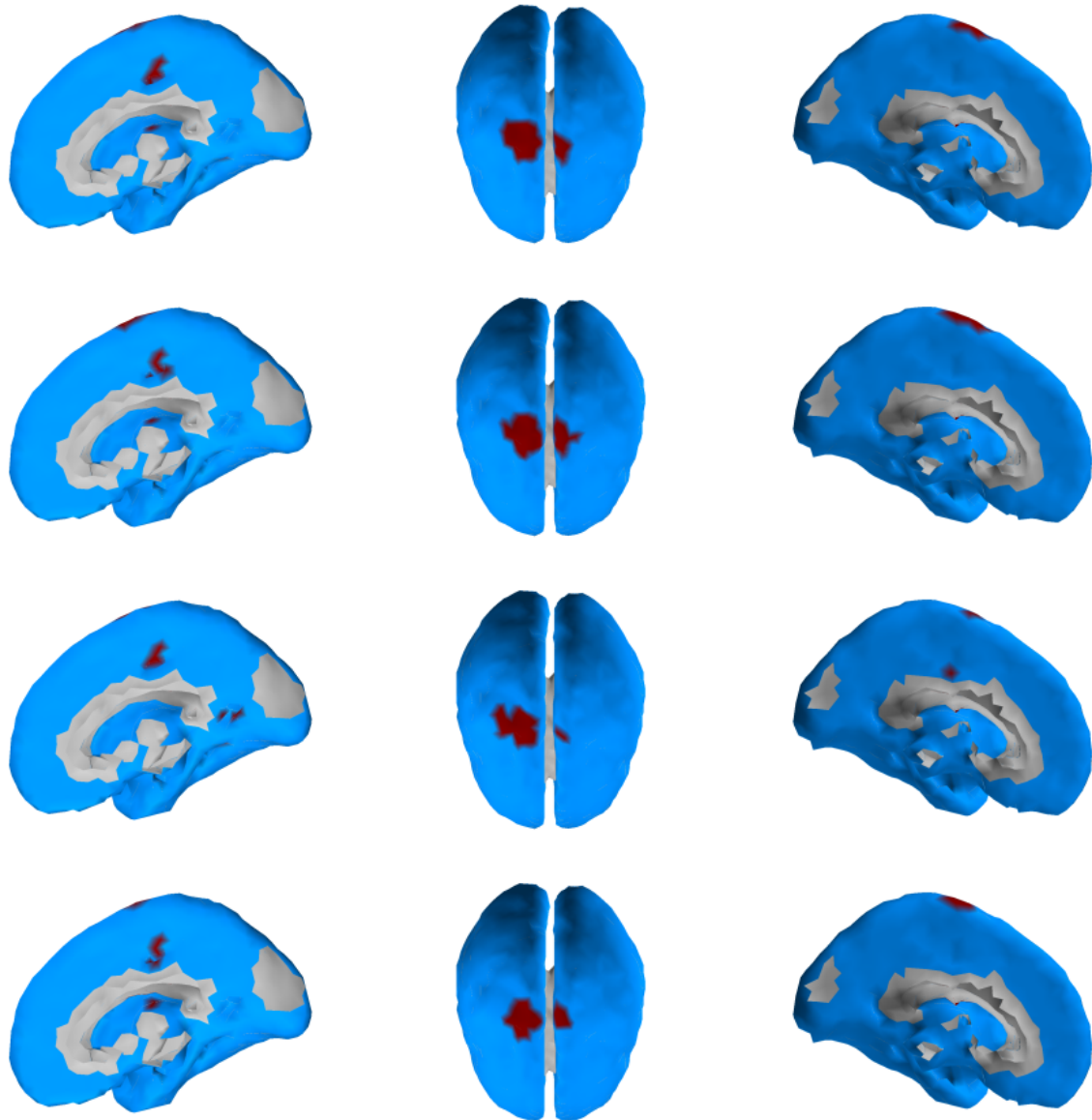


Figure 8.4: Most relevant solution points for the CCD, 5 subjects. Localization of the 50 most relevant solution points of the CCD model for ErrP classification for the average of the five subjects for the 4 groups of recording sessions. Clear clusters can be seen and the localization is very stable from a group of sessions to the other.

8.4 Estimated local field potentials (eLFP)

Figure 8.5 shows the 50 most relevant solution points of the eLFP model for ErrP classification for the average of the ten subjects for the 5 first sessions (top) and for the 5 last sessions (bottom). Even if the most relevant solution points seem to be mostly localized at the surface of a wide fronto-central area, no clear compact clusters can be seen on figure 8.5 and the variability is higher than with the CCD model. Indeed, only 68% of the solution points are common to the 2 groups. Finally it's interesting to note that for both CCD and eLFP, the localization is more important on the right hemisphere for the 5 first sessions whereas it is more important on the left hemisphere for the 5 last sessions (see central views of figures 8.3 and 8.5). This variability can be due to the fact that there are only 5 sessions for each group.

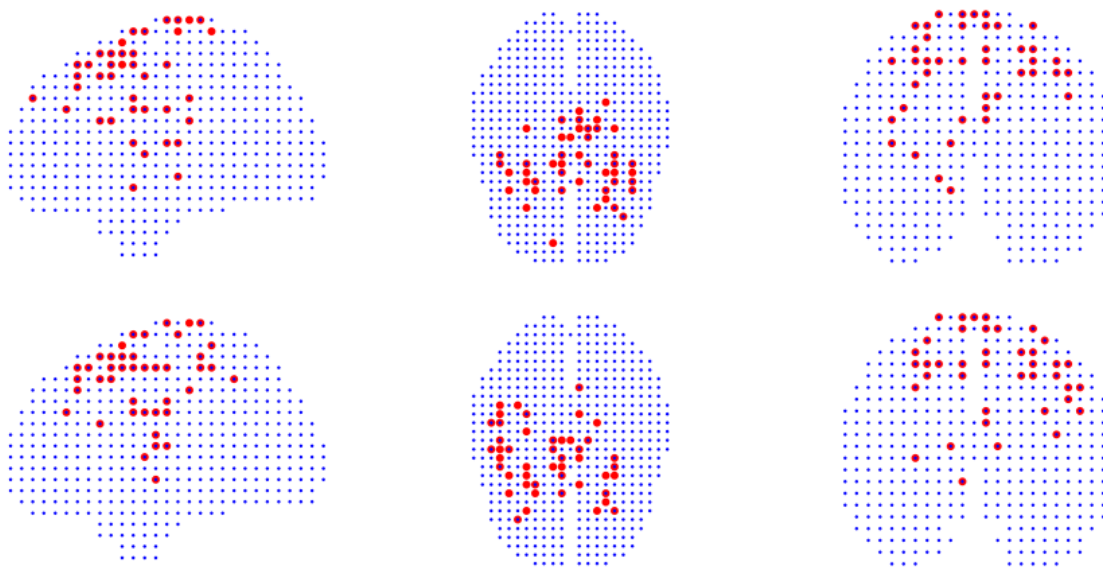


Figure 8.5: Most relevant solution points for eLFP, 10 subjects. Localization of the 50 most relevant solution points of the eLFP model for ErrP classification for the average of the ten subjects for the 5 first sessions (top) and for the 5 last sessions (bottom). No evident clusters can be seen and localization is not as stable as with the CCD, only 68% of the points are common to the 2 groups.

Figure 8.6 shows the 50 most relevant solution points of the eLFP model for ErrP classification for the average of the five subjects for the four groups of recordings (top to bottom). The most relevant solution points are mostly localized at the surface of a central area and a relatively stable cluster can be seen over the right hemisphere. However, this localization is not as clear and stable as the localization of the clusters obtained using the CCD model.

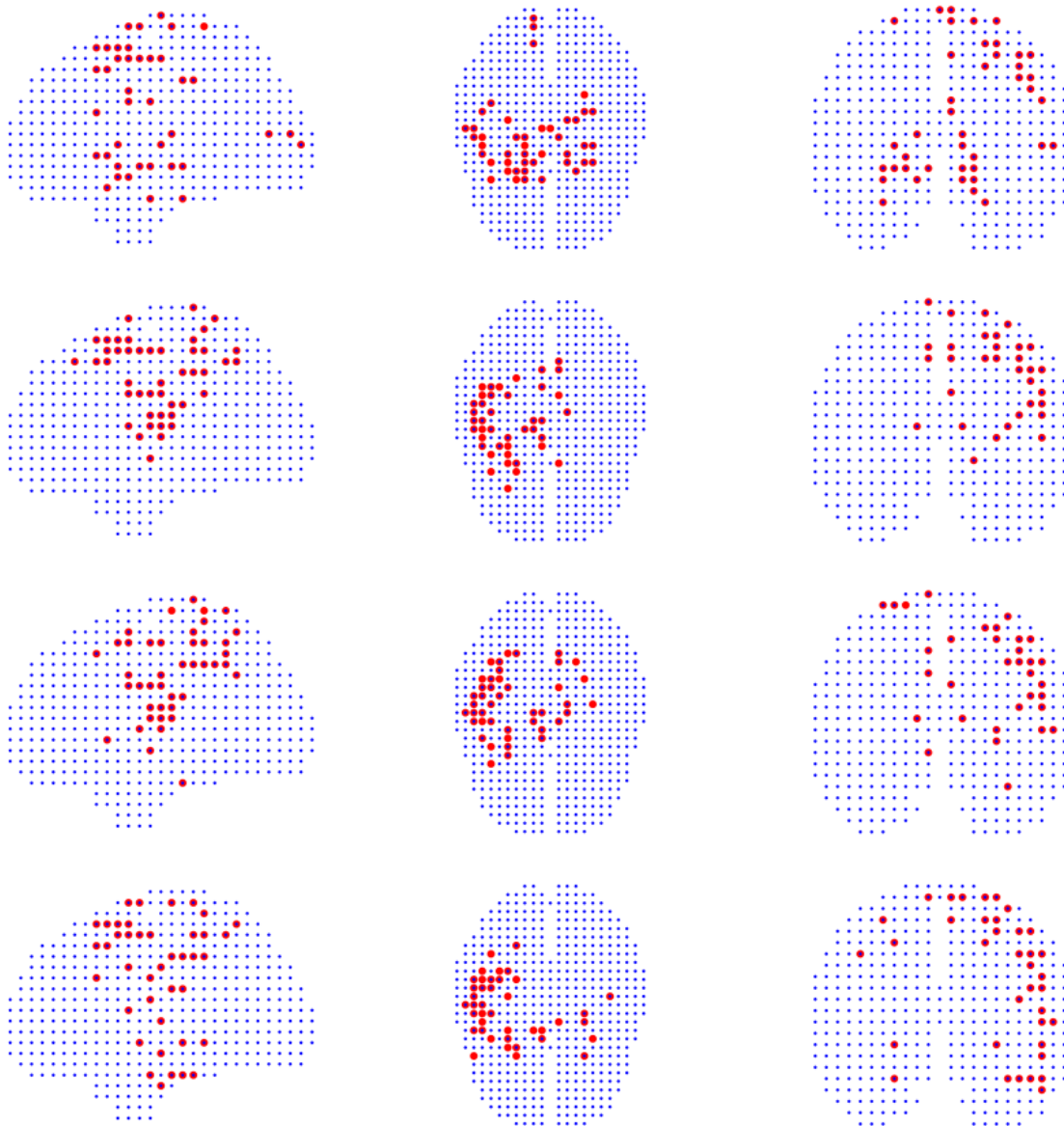


Figure 8.6: Most relevant solution points for eLFP, 5 subjects. Localization of the 50 most relevant solution points of the eLFP model for ErrP classification for the average of the five subjects for the 4 groups of recording sessions. Even if a cluster can be seen over the right hemisphere, the localization is not as clear and stable as with the CCD model.

8.5 Single trial classification

Table 8.1 reports the ErrP classification rates for the 10 subjects and for the average of them using EEG, the CCD and the eLFP and Figure 8.7 plots the average classification rates for the 10 subjects using the 3 techniques. Performances using EEG are very similar to performances reported in the previous chapters of this thesis, the average performance is around 80%. The average performance using the CCD model is just below 85% and is significantly better than the performance using EEG (Wilcoxon test, $p < 0.005$). Indeed, Table 8.1 and Figure 8.7 show

Table 8.1: Recognition rates using EEG, CCD and eLFP, 10 subjects. Percentages of correctly recognized error trials and correct trials for the 10 subjects and the average of them using EEG, CCD and eLFP. The CCD model is clearly the best for 8 subjects, whereas the eLFP is clearly the worst for 6 subjects.

	EEG			CCD			eLFP		
	Error	Correct	AVG	Error	Correct	AVG	Error	Correct	AVG
Subject 1	80.9	74.4	77.7	79.4	84.7	82.1	83.8	61.5	72.7
Subject 2	84.9	76.7	80.8	75.6	86.6	81.1	73.3	72.7	73.0
Subject 3	86.4	72.1	79.3	78.8	77.4	78.1	87.9	78.9	83.4
Subject 4	76.4	85.5	81.0	89.1	83.9	86.5	70.9	85.5	78.2
Subject 5	87.3	88.4	87.9	87.3	97.1	92.2	84.8	81.4	83.1
Subject 6	73.9	77.6	75.8	81.8	86.2	84.0	80.7	70.7	75.7
Subject 7	84.3	77.4	80.9	90.4	81.8	86.1	80.7	76.7	78.7
Subject 8	74.7	82.9	78.8	72.0	92.4	82.2	77.3	79.6	78.5
Subject 9	88.0	80.4	84.2	89.3	87.2	88.3	89.3	73.6	81.5
Subject 10	93.2	70.6	81.9	78.1	90.4	84.3	87.7	74.8	81.3
Average	83.0	78.6	80.8	82.2	86.8	84.5	81.6	75.5	78.6
S.D.	6.4	5.7	3.4	6.5	5.5	4.0	6.3	6.6	3.9

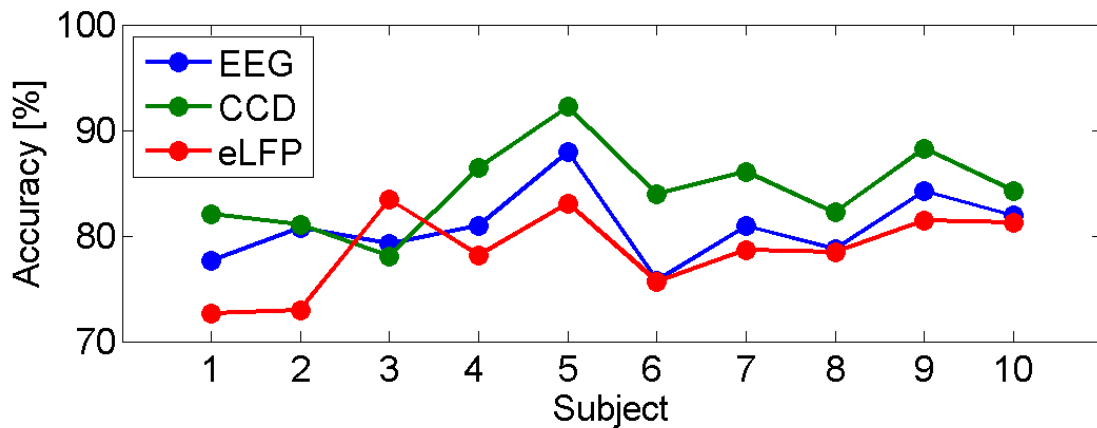


Figure 8.7: Average recognition rates, 10 subjects.

Average accuracy of the ErrP detection for the 10 subjects using EEG, CCD and eLFP. The CCD model is clearly the best for 8 subjects, whereas the eLFP is clearly the worst for 6 subjects.

that the CCD is better than EEG for 8 subjects. The average performance using eLFP is below 80%. The difference between EEG and eLFP is not as important as the difference between the CCD and EEG, however EEG is significantly better than eLFP (Wilcoxon test, $p < 0.05$).

Table 8.2 reports the ErrP classification rates for the five subjects and for the average of them using EEG, the CCD and the eLFP and Figure 8.8 plots the average classification rates for the five subjects using the 3 techniques. Once again, performances using EEG are around 80%, this

is similar to all other results obtained in the previous chapters of this thesis. The performance using the CCD model is again slightly above the performance of EEG and the difference is significant (wilcoxon test, $p < 0.05$). The performance using eLFP is again under 80% and EEG is again significantly better than eLFP (wilcoxon test, $p < 0.05$). In fact, Table 8.2 and Figure 8.8 show that the CCD model systematically provides the best classification rates whereas the eLFP model systematically provides the worst classification rates.

Table 8.2: Recognition rates using EEG, CCD and eLFP, 5 subjects
Percentages of correctly recognized error trials and correct trials for the 5 subjects and the average of them using EEG, CCD and eLFP. The CCD model systematically shows the best classification rates whereas the eLFP model systematically shows the worst.

	EEG			CCD			eLFP		
	Error	Correct	AVG	Error	Correct	AVG	Error	Correct	AVG
Subject 1	76.5	85.0	80.0	78.9	86.7	82.8	76.7	75.6	76.1
Subject 2	71.3	87.1	79.2	78.4	86.4	82.4	72.7	77.5	75.1
Subject 3	74.9	83.6	79.3	74.8	85.1	80.0	80.8	77.4	79.1
Subject 4	81.3	76.0	78.6	82.3	86.4	84.4	82.7	70.9	76.8
Subject 5	78.6	89.5	84.1	81.1	89.8	85.5	83.0	77.2	80.1
Average	76.5	84.2	80.4	79.1	86.9	83.0	79.2	75.7	77.5
S.D.	3.8	5.1	2.2	2.9	1.8	2.1	4.4	2.8	2.1

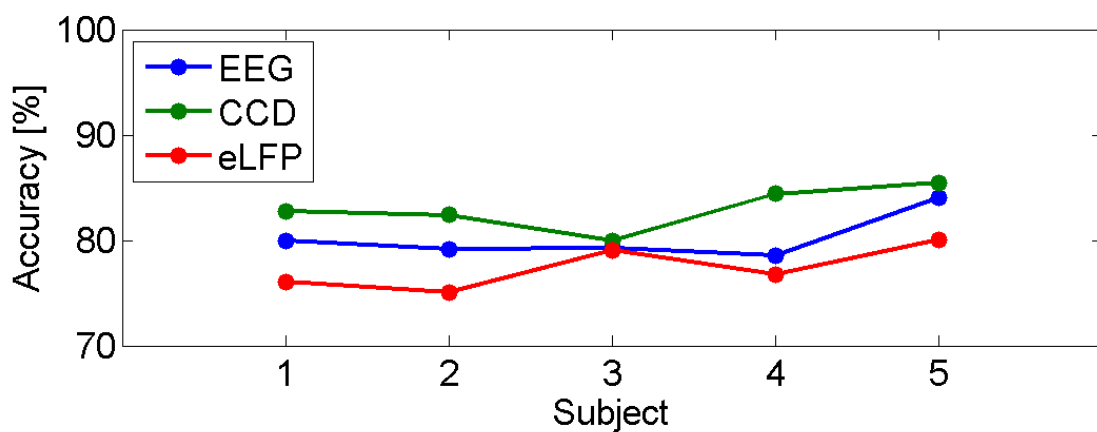


Figure 8.8: Average recognition rates, 5 subjects.
Average accuracy of the ErrP detection for the 5 subjects using EEG, CCD and eLFP. The CCD systematically provides the best classification rates whereas the eLFP model systematically provides the worst classification rates.

8.6 Source localization

The sLORETA software was once again used as a localization tool to estimate the focus of intracranial activity at the occurrence of the five described ErrP peaks. Figure 8.9 shows Talairach slices of localized activity for the grand average of the ten subjects for the 5 first sessions at the occurrence of the five previously described peaks and Figure 8.10 shows the same slices for the 5 last sessions. As expected, the areas involved in error processing, namely the pre-supplementary motor area (pre-SMA, Brodmann area 6) and the rostral cingulate zone (RCZ, Brodmann areas 24 & 32) are systematically activated. Indeed, Brodmann area 6 is the best match for the 200 ms positive peak, the 420 ms negative peak and the 600 ms positive peak. The X coordinate ranges between 0mm and 5mm, the Y coordinate ranges from -10mm to -25mm and the Z coordinate ranges between 45mm and 55mm. For the 250 ms negative peak, the best match is Brodmann area 24, the focus is located at X=0mm, Y=-5mm and Z=45mm or Z=50mm. Finally, for the 320 ms positive peak, the best match is Brodmann area 32 with a focus located at X=0mm, Y=5mm and Z=50mm. These localization are very similar to those described earlier in this thesis. As mentioned in the previous chapters, the general localization is at the limits of the anterior cingulate cortex, the pre-supplementary motor area and the posterior cingulate cortex, matching relatively well the clusters obtained using the CCD model.

8.7 Conclusion

In this chapter we investigated the potential benefit of using inverse solutions to increase ErrP detection rates. In particular, we tested two inverse models, namely the Cortical Current Density model (CCD) and the estimated Local Field Potentials (eLFP). The aim was of course not to make an evaluation of the quality of these models, it was simply to use these models as they are and evaluate their potential benefit in the particular framework of BCIs. The procedure was then to isolate the solution points of both models that are most relevant for classification. It is to note here that the most relevant solution points are not necessarily located in brain areas responsible for ErrP generation. Indeed, source localization and feature selection for classification are two different issues. The use of inverse solutions can potentially increase the classification rates, because these methods make it possible to unmix the signals measured at the scalp, attributing to each brain area its own estimated temporal activity. The best solution points can then be selected for classification. At the level of EEG, the classification may not be easy since a single electrode records the activity of a large population of neurons and the resulting signal represents the overlapping of several brain processes. The projection in a space with a much higher dimensionality provided by an inverse model spatially separates the processes and therefore may lead to an easier separability between the different classes.

According to the reported results, it seems that the use of the CCD model could significantly increase ErrP classification. Its high stability over sessions provides the best classification results, notably better than EEG. Furthermore the most relevant solution points for ErrP classification located at the surface are partly in the pre-SMA and other relevant points seem to be located in the cingulate cortex. The positions of these 2 clusters are quite consistent with the source localization provided by the sLORETA model. As previously explained, the most relevant points for classification are not necessarily located in brain areas responsible for ErrP generation, but in the case of the CCD model it seems to be relatively the case. On the contrary,

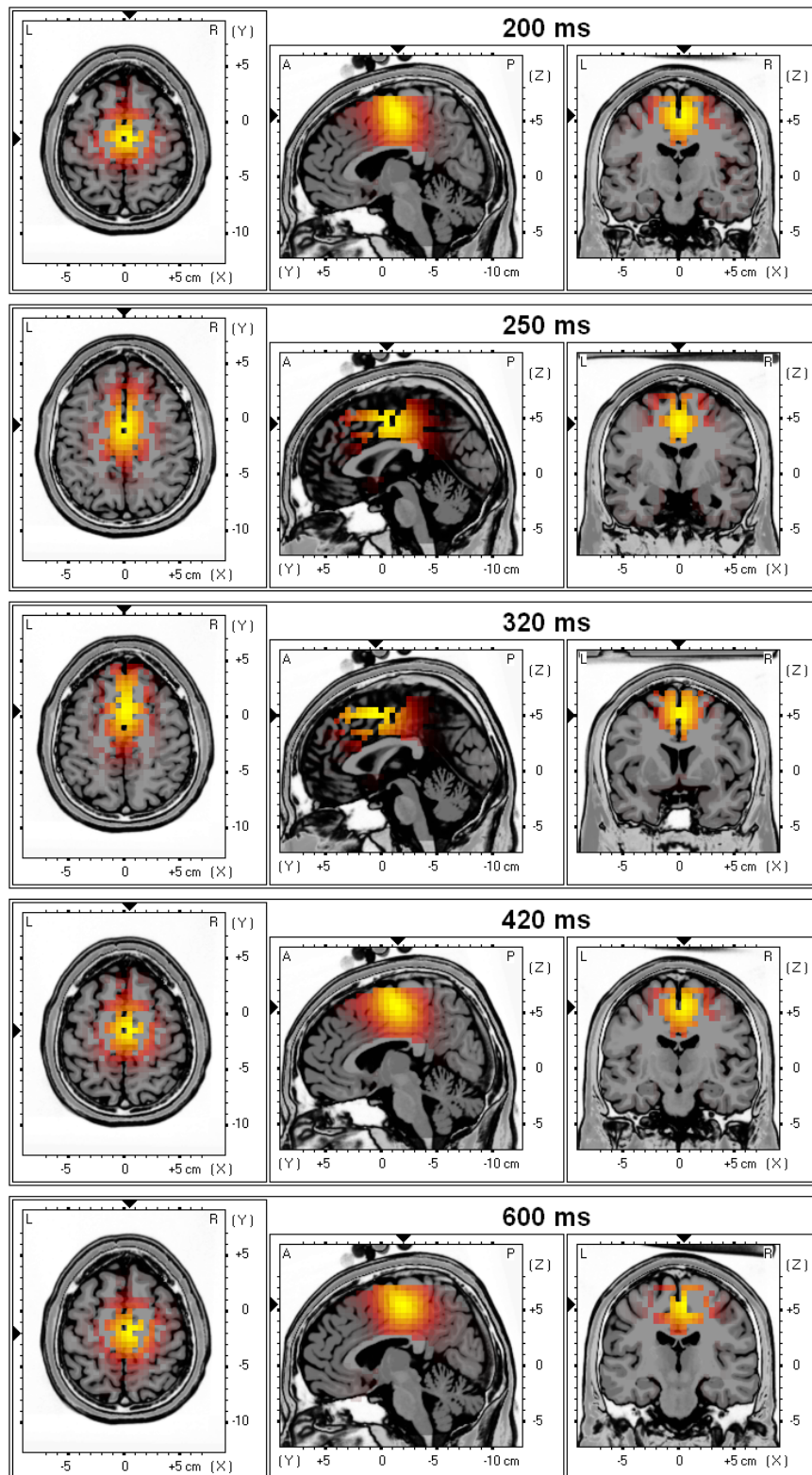


Figure 8.9: Source localization, first group

Talairach slices of localized activity for the grand average of the ten subjects for the 5 first sessions at the occurrence of the five described peaks.

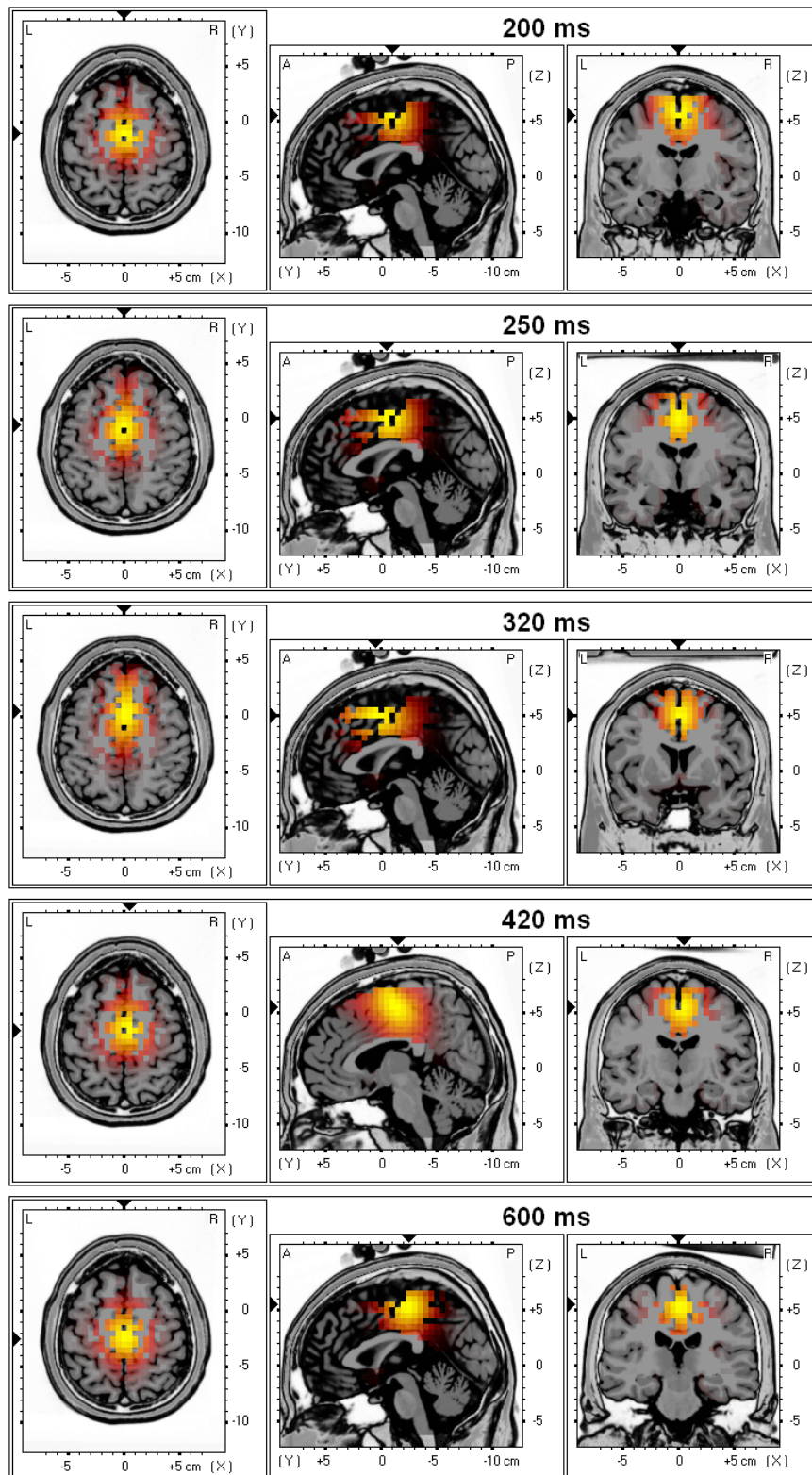


Figure 8.10: Source localization, second group
 Talairach slices of localized activity for the grand average of the ten subjects for the 5 last sessions at the occurrence of the five described peaks.

the eLFP seem to be too unstable to be successfully used in an online system and appear to perform significantly worse than both EEG and CCD. This instability is reflected in the ranking of the solution points according to their discriminant power. The 10 best points selected using the 5 first sessions were used for classification. For the CCD model, these 10 best points are all found in the list of the first 42 best points ranked using the 5 last sessions. For the eLFP, the 10 best points are all in the list of the first 87 best points ranked using the 5 last sessions. This implies that the overall discriminant power of the eLFP is lower, and thus explains the lower classification rates. Nevertheless, these results have to be confirmed in a real online application to assess if the use of the CCD model is really increasing the performances. Indeed, according to this off-line study, the increase of performance is less than 4% and drops to less than 3% when investigated on data sets recorded on 2 different days. If the increase of performance is not significant, it wouldn't make sense to use features that require more complicated processing than EEG channels for online ErrP classification.

Chapter 9

Conclusion

In this thesis we have reported results on the detection of the neural correlate of error awareness for improving the performance and reliability of BCI. In particular, we have shown the existence of a new kind of error-related potentials elicited in reaction to an erroneous recognition of the subject's intention. The main components of these *interaction ErrP* are a positive peak ~ 200 ms after the feedback, a negative peak after ~ 250 ms, a second larger positive peak after ~ 320 ms, a second larger negative peak after ~ 450 and a late positive peak after ~ 600 ms. The interaction ErrP are different from the other known ErrP. The response ErrP, typically observed in choice reaction tasks, are characterized by a negative component ~ 100 ms after an erroneous response of the subject and by a broader positive component peaking between 200 and 500 ms after the erroneous response. This is very different from the reported interaction ErrP. Both feedback ErrP and observation ErrP are only characterized by a small negative deflection ~ 250 ms after the feedback. Even if the interaction ErrP show a negative component ~ 250 ms after the feedback, there are other more important components. All interaction ErrP components show most of the time a fronto-central localization, with the highest amplitude under electrode FCz. The stability of the ErrP components and of their localization indicates a stability of the phenomena responsible for ErrP generation. Furthermore the couple FCz-Cz show a great stability in classification rates. Indeed, ErrP classification rates are always around 80% on average.

We have also confirmed that the pre-supplementary motor area and the anterior cingulate cortex (both ventral and dorsal) are important brain areas involved in error processing and responsible for ErrP generation. Most of the source localization performed in the different chapters of this thesis using the sLORETA software clearly show that the foci of activity at the occurrence of the different ErrP component are located in the posterior part of the anterior cingulate cortex and/or in the pre-supplementary motor area.

More importantly, we have shown the feasibility of simultaneously and satisfactorily detecting erroneous responses of the interface and classifying motor imagination for device control at the level of single trials. Indeed, the off-line analysis of Chapter 6 showed for the first time that subjects are able to focus on motor imagery tasks to control a cursor and to be simultaneously aware of the feedback delivered by the system at each single trial for ErrP detection. The average motor imagery classification rate before integration ErrP detection was 73% and the average ErrP detection was just under 80%. The integration of ErrP detection in this case leads to a great increase of the theoretical performance, the bit rate is doubled using ErrP detection. The online experiments of Chapter 7 confirmed these results. Two subjects successfully managed

to control in real time the cursor and were simultaneously sufficiently aware of the feedback to ensure good ErrP detection. For both subjects the ErrP detection was above 80% whereas motor imagery classification rate without ErrP integration were just under 70%. Again, the BCI performance is greatly improved by ErrP detection. The bit rate is multiplied by more than 3 for both subjects and the error rates drop under 10%.

Another issue explored in this thesis is the potential benefit of using inverse solutions to increase ErrP detection rates. In particular, we tested two inverse models, namely the Cortical Current Density model (CCD) and the estimated Local Field Potentials (eLFP). The goal was not to compare inverse solutions or to evaluate their general performances, it was simply to use these models as tools and evaluate their potential benefit in the particular framework of BCIs. Therefore, we tried to isolate solution points that are most relevant for ErrP classification. These points are not necessarily located in brain areas responsible for ErrP generation, source localization and feature selection for classification may lead to slightly different results. The analysis of Chapter 8 showed that the use of the CCD model could significantly increase ErrP classification. Its high stability over sessions provides the best classification results, notably better than EEG. Furthermore the most relevant solution points for ErrP classification located at the surface are partly in the pre-SMA and other relevant points seem to be located in the anterior cortex. The positions of these two clusters are quite consistent with the source localization provided by the sLORETA model. On the contrary, the eLFP seem to be too unstable to be successfully used in an online system and appear to perform significantly worse than both EEG and CCD. Even if the increase in performance is statistically significant using the CCD model, it is less than 4% and drops to less than 3% when the analysis is performed on data sets recorded on two different days. Further investigations are needed to confirm the fact that the CCD model is a useful tool for BCI. If the increase of performance is not significant, it would not make sense to use features that require more complicated processing than EEG channels for online ErrP classification. The key point using inverse models providing thousands of solution points is the stability of the list of the most relevant solution points for classification. Indeed, stability of features are more important than very high classification rates for a successful brain-computer interaction.

9.1 Future directions

In this thesis we have shown the feasibility of using a specific type of high-level cognitive states, namely error-related potentials, to improve the quality of the brain-computer interaction. Motor imagery is classified simultaneously and in addition to ErrP, the former is used for mental control to drive a device and the latter is incorporated as a verification tool. As shown in this thesis, the integration of ErrP detection in the BCI leads to great improvement of the performance of the BCI.

However, the introduction of an automatic verification procedure strongly interferes with the BCI. The user needs to process additional information which induces higher workload and slows down the interaction. Integrating ErrP or not, it's vital to design BCI protocols where the subject feels comfortable and has the feeling of control. The integration of ErrP is of great benefit especially for beginner users who normally reach moderate performances. Indeed, ErrP detection can be seen as the additional side wheels on a bicycle. They greatly help during the training phase, but they become useless once the biker has a good control of his bicycle.

For a subject reaching modest performance with a 2 tasks BCI, say 65%, the integration of ErrP detection (80% average accuracy) will multiply the performance by 4. However, for an experienced subject reaching 95% of correct classification using the BCI, the integration of ErrP becomes useless since the performance is in this situation 3% lower. The advantage of using ErrP detection for the training of beginner users will be achieved soon. To really demonstrate that ErrP detection accelerates learning we will have 2 groups of subjects, one group learning with ErrP detection and one group without. Another important aspect to avoid frustration of the subject is to minimize the amount of correct commands that are stopped by the ErrP detection, i.e. avoid classifying correct trials as error. Indeed, for the motivation of the subject it's better to transmit a few wrong commands than to block too many correct commands. For this we could force the ErrP classifier to reach a given accuracy for correct trials classification, even if it decreases the error trials detection.

Another topic that would be worth exploring is alternative data pre-processing for error-related potentials. So far we use a rather simple technique, we select a time window after the feedback and the sub-sampled time concatenated signal of electrodes FCz and Cz is sent for classification. However, it's clear that the ErrP waveform varies in time and amplitude among all single trials. Therefore, a more adaptive technique for ErrP detection could lead to better recognition rates. In this respect, the use of wavelet transform provides a time-frequency decomposition that has been recently reported to be very suitable for event-related potentials analysis [Quiñero Quiroga and Garcia, 2003, Subasi et al., 2005].

As already mentioned, further experiments using the CCD model, and also eventually with other inverse models, are needed to confirm the benefit of using inverse solutions to increase ErrP detection rates. Real time experiments using the CCD model should be performed, on-line experiments are the ultimate and the best way to validate a technique. As mentioned in Chapter 8, the increase of performance using the CCD model is under 5%. If during real time experiments the classification rates don't remain significantly higher, it wouldn't make sense to continue working with features requiring more complicated processing than EEG for ErrP classification. Finally, the use of the CCD model could be extended to motor imagery. Indeed, if the CCD is useful to discriminate error and correct trials, it could also be a benefic tool for motor imagery tasks classification.

In this thesis, the movement of the cursor eliciting ErrP occurs at precise moments, the cursor doesn't have a continuous movement. The discrete behavior of the cursor provides the system with the precise time of the occurrence of the feedback (trigger) and this makes it possible to select the correct time window after the feedback to perform the ErrP detection. However, this discrete behavior is not always appropriate. The recent studies showing that subjects are able to control a miniature robot or a wheelchair they sit on in an indoor environment only using the signals derived from a EEG-based brain-computer interface suggest a continuous behavior of the driven device [Millán et al., 2004a, Philips et al., 2007]. In this case the feedback is no more delivered at precise moments but is continuous, so that the presence of traditional error-related potentials is doubtful. Trying to detect the neural correlate of error awareness in a continuous framework and using *continuous ErrP* for improving the performance and reliability of BCI opens a whole field of investigation. Continuous ErrP will most probably have a different waveform, their amplitude should vary according to the importance of the error, and the precise time at which they occur is unknown, making the detection more complicated.

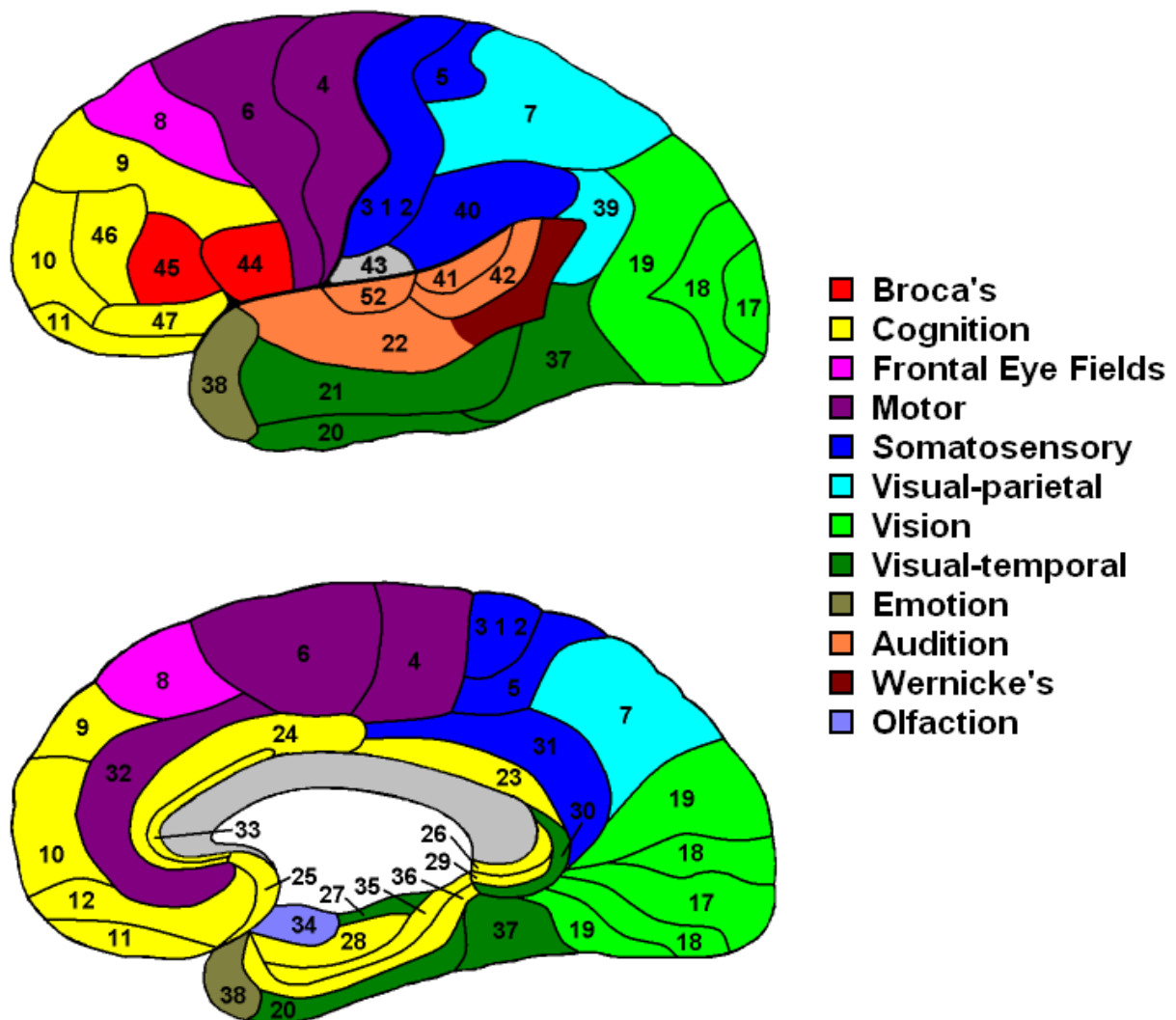
In this thesis, error-related potentials were used simultaneously and in addition to motor imagery to control in real time a cursor on a computer screen, i.e. ErrP were successfully

exploited in a BCI framework. However, this is not the only way to take benefit of ErrP. Indeed, recent studies conducted at the IDIAP Research Institute show that with ErrP, subjects are able to monitor an external system upon which she/he has no control whatsoever. In this new approach, the subject acts as a critic who observes the behavior of an autonomous agent and emits monitoring signals about its performance. Preliminary experimental results show that single-trial recognition of correct and error trials can be achieved. Furthermore, ErrP could also be used as learning signals to infer the user's intended strategy and to modify the performance of the system being monitored according to this strategy [Chavarriaga et al., 2007].

More generally, the work described in this thesis suggests that it could be possible to recognize in real time high-level cognitive and emotional states from EEG (as opposed, and in addition, to motor commands) such as alarm, fatigue, frustration, confusion, or attention that are crucial for an effective and purposeful interaction. Indeed, the rapid recognition of these states will lead to truly adaptive interfaces that customize dynamically in response to changes of the cognitive and emotional/affective states of the user.

Appendix A

Brodmann areas



Areas 1, 2 & 3 - Primary Somatosensory Cortex (frequently referred to as Areas 3, 1, 2 by convention)

Area 4 - Primary Motor Cortex

Area 5 - Somatosensory Association Cortex

Area 6 - Pre-Motor and Supplementary Motor Cortex

Area 7 - Somatosensory Association Cortex

Area 8 - Includes Frontal Eye Field

Area 9 - Dorsolateral prefrontal cortex

Area 10 - Frontopolar area (most rostral part of superior and middle frontal gyri)

Area 11 - Orbitofrontal area (orbital and rectus gyri, plus part of the rostral part of the superior frontal gyrus)

Area 12 - Orbitofrontal area (used to be part of BA11, refers to the area between the superior frontal gyrus and the inferior rostral sulcus)

Area 13 - Insular cortex

Area 17 - Primary Visual Cortex (V1)

Area 18 - Visual Association Cortex (V2)

Area 19 - Visual Association Cortex (V3)

Area 20 - Inferior Temporal gyrus

Area 21 - Middle Temporal gyrus

Area 22 - Superior Temporal Gyrus, of which the rostral part participates to Wernicke's area

Area 23 - Ventral Posterior cingulate cortex

Area 24 - Ventral Anterior cingulate cortex

Area 25 - Subgenual cortex

Area 26 - Ectosplenial area

Area 28 - Posterior Entorhinal Cortex

Area 29 - Retrosplenial cingular cortex

Area 30 - Part of cingular cortex

Area 31 - Dorsal Posterior cingular cortex

Area 32 - Dorsal Anterior cingulate cortex

Area 34 - Anterior Entorhinal Cortex (on the Parahippocampal gyrus)

Area 35 - Perirhinal Cortex (on the Parahippocampal gyrus)

Area 36 - Parahippocampal cortex (on the Parahippocampal gyrus)

Area 37 - Fusiform gyrus

Area 38 - Temporopolar area (most rostral part of the superior and middle temporal gyri)

Area 39 - Angular gyrus, part of Wernicke's area

Area 40 - Supramarginal gyrus part of Wernicke's area

Area 41 & 42 - Primary and Auditory Association Cortex

Area 43 - Subcentral area (between insula and post/precentral gyrus)

Area 44 - pars opercularis, comprises Broca's area

Area 45 - pars triangularis, comprises Broca's area

Area 46 - Dorsolateral prefrontal cortex

Area 47 - Inferior prefrontal gyrus

Area 48 - Retrosubicular area (a small part of the medial surface of the temporal lobe)

Area 52 - Parainsular area (at the junction of the temporal lobe and the insula)

Bibliography

- F. Babiloni, C. Babiloni, L. Locche, F. Cincotti, P.M. Rossini, and F. Carducci. High-resolution EEG: Source estimates of Laplacian-transformed somatosensory-evoked potentials using realistic subject head model constructed from magnetic resonance images. *Medical & Biological Engineering and Computing*, 38:512–519, 2000a.
- F. Babiloni, F. Cincotti, L. Lazzarini, J. del R. Millán, J. Mouriño, M. Varsta, J. Heikkonen, L. Bianchi, and M.G. Marciani. Linear classification of low-resolution EEG patterns produced by imagined hand movements. *IEEE Transactions on Rehabilitation Engineering*, 8: 186–188, 2000b.
- F. Babiloni, C. Babiloni, F. Carducci, G.L. Romani, P.M. Rossini, L.M. Angelone, and F. Cincotti. Multimodal integration of high resolution EEG and functional magnetic resonance: A simulation study. *NeuroImage*, 19:1–15, 2003.
- F. Babiloni, F. Cincotti, C. Babiloni, F. Carducci, D. Mattia, L. Astolfi, A. Basilico, P.M. Rossini, L. Ding, Y. Ni, J. Cheng, K. Christine, J. Sweeney, and B. He. Estimation of the cortical functional connectivity with the multimodal integration of high-resolution EEG and fMRI data directed by transfer function. *NeuroImage*, 24:118–131, 2005.
- R.D. Badgaiyan and M.I. Posner. Mapping the cingulate cortex in response selection and monitoring. *NeuroImage*, 7:255–260, 1998.
- M.F. Bear, B.W. Connors, and M.A. Paradiso. *Neuroscience: Exploring the brain*. Lippincott Williams & Wilkins, 2001.
- H. Berger. Über das Elektrenkephalogramm des Menschen. *Archive für Psychiatrie und Nervenkrankheiten*, 87:527–570, 1929.
- P.S. Bernstein, M.K. Scheffers, and M.G.H. Coles. Where did I go wrong? A psychophysiological analysis of error detection. *Journal of Experimental Psychology: Human Perception and Performance*, 21:1312–1322, 1995.
- N. Birbaumer, N. Ghanayim, T. Hinterberger, I. Iversen, B. Kotchoubey, A. Kubler, J. Perelmouter, E. Taub, and H. Flor. A spelling device for the paralysed. *Nature*, 398:297–298, 1999.
- G.E. Birch, Z. Bozorgzadeh, and S.G. Mason. Initial on-line evaluation of the LF-ASD brain-computer interface with able-bodied and spinal-cord subjects using imagined voluntary motor potentials. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 10: 219–224, 2002.

- B. Blankertz, C. Schäfer, G. Dornhege, and G. Curio. Single trials detection of EEG error potentials: A tool for increasing BCI transmission rates. *Artificial Neural Networks ICANN 2002*, pages 1137–1143, 2002.
- B. Blankertz, G. Dornhege, C. Schäfer, R. Krepi, J. Kohlmorgen, K.-R. Müller, V. Kunzmann, F. Losch, and G. Curio. Boosting bit rates and error detection for the classification of fast-paced motor commands based on single-trial EEG analysis. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11:127–131, 2003.
- 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*, 14:147–152, 2006.
- K. Brodmann. *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues*. Johann Ambrosius Barth Verlag, Leipzig, 1909.
- G. Bush, P. Luu, and M.I. Posner. Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Sciences*, 4:215–222, 2000.
- A. Buttfield, P.W. Ferrez, and J. del R. Millán. Towards a robust BCI: Error potentials and online learning. *IEEE Transactions on Neural Systems and Rehabilitation*, 14:164–168, 2006.
- J.M. Carmena, M.A. Lebedev, R.E. Crist, J.E. O’Doherty, D.M. Santucci, D.F. Dimitrov, P.G. Patil, C.S. Henriquez, and M.A.L. Nicolelis. Learning to control a brain-machine interface for reaching and grasping by primates. *PLoS Biology*, 1:193–208, 2003.
- C.S. Carter, T.S. Braver, D.M. Barch, M.M Botvinick, D. Noll, and J.D. Cohen. Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science*, 280: 747–749, 1998.
- C.S. Carter, M.M. Botvinik, and J.D. Cohen. The contribution of the anterior cingulate cortex to executive processes in cognition. *Reviews in the Neurosciences*, 10:49–57, 1999.
- J.K. Chapin, K.A. Moxon, R.S. Markowitz, and M.A.L. Nicolelis. Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex. *Nature Neuroscience*, 2: 664–670, 1999.
- R. Chavarriaga, P.W. Ferrez, and J. del R. Millán. To err is human: Learning from error potentials in brain-computer interfaces. In *1st International Conference on Cognitive Neurodynamics*, 2007.
- F. Cincotti. *Multimodal integration of neuroelectromagnetic and fMRI data: The role of different metrics in the solution of the linear inverse problem*. PhD thesis, Università degli Studi di Roma La Sapienza, 2002.
- M.G.H. Coles, G. Gratton, T.R. Bashore, C.W. Eriksen, and E. Donchin. A psychophysiological investigation of the continuous flow model of human information processing. *Journal of Experimental Psychology: Human Perception and Performance*, 11:529–553, 1985.

- M.G.H. Coles, M.K. Scheffers, and C.B. Holroyd. Why is there an ERN/Ne on correct trials? Response representations, stimulus-related components, and the theory of error processing. *Biological Psychology*, 56:173–189, 2001.
- A.M. Dale and M.I. Sereno. Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: A linear approach. *Journal of Cognitive Neuroscience*, 5:162–176, 1993.
- F. Darvas, D. Pantazis, E. Kucukaltun-Yildirim, and R.M. Leahy. Mapping human brain function with MEG and EEG: Methods and validation. *NeuroImage*, 23:S289–S299, 2004.
- O. Devinsky, M.J. Morrell, and B.A. Vogt. Contributions of anterior cingulate cortex to behaviour. *Brain*, 118:279–306, 1995.
- E. Donchin, G. Gratton, D. Dupree, and M. Coles. After a rash action: Latency and amplitude of the P300 following fast guesses. In G.C. Galbraith, M.L. Kietzman, and E. Donchin, editors, *Neurophysiology and Psychophysiology: Experimental and Clinical Applications*, pages 173–188. Lawrence Erlbaum Associates, Hillsdale, NJ, 1988.
- 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*, 8:174–179, 2000.
- O. Donchin, A. Gribova, O. Steinberg, H. Bergman, S. Cardoso de Oliveira, and F. Vaadia. Local field potentials related to bimanual movements in the primary and supplementary motor cortices. *Experimental Brain Research*, 140:46–55, 2001.
- T. Elbert, B. Rockstroh, W. Lutzenberger, and N. Birbaumer. Biofeedback of slow cortical potentials. *Electroencephalography and Clinical Neurophysiology*, 48:293–301, 1980.
- M. Falkenstein, J. Hohnsbein, J. Hoormann, and L. Blanke. Effects of errors in choice reaction tasks on the ERP under focused and divided attention. In C.H.M. Brunia, A.W.K. Gaillard, and A. Kok, editors, *Psychophysiological Brain Research*, pages 192–195. Tilburg University Press, Tilburg, 1990.
- M. Falkenstein, J. Hoormann, S. Christ, and J. Hohnsbein. ERP components on reaction errors and their functional significance: A tutorial. *Biological Psychology*, 51:87–107, 2000.
- 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*, 70:510–523, 1998.
- P.W. Ferrez and J. del R. Millán. You are wrong!—Automatic detection of interaction errors from brain waves. In *19th International Joint Conference on Artificial Intelligence*, pages 1413–1418, 2005.
- P.W. Ferrez and J. del R. Millán. Error-related EEG potentials in brain-computer interfaces. In G. Dornhege, J. del R. Millán, T. Hinterberger, D. McFarland, and K.-R. Müller, editors, *Toward Brain-Computer Interfacing*. The MIT Press, 2007.

- K. Fiehler, M. Ullsperger, and Y. von Cramon. Neural correlates of error detection and error correction: Is there a common neuroanatomical substrate? *European Journal of Neuroscience*, 19:3081–3087, 2004.
- M. Fuchs, M. Wagner, T. Kohler, and H.-A. Wischmann. Linear and nonlinear current density reconstructions. *Journal of Clinical Neurophysiology*, 16:267–295, 1999.
- V.V. Galván, J. Chen, and N.M. Weinberger. Differential thresholds of local field potentials and unit discharges in rat auditory cortex. *Hearing Research*, 167:57–60, 2002.
- X. Gao, X. Dingfeng, M. Cheng, and S. Gao. A BCI-based environmental controller for the motion-disabled. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11: 137–140, 2003.
- W.J. Gehring and D.E. Fencsik. Functions of the medial frontal cortex in the processing of conflict and errors. *The Journal of Neuroscience*, 21:9430–9437, 2001.
- W.J. Gehring and R.T. Knight. Prefrontal-cingulate interactions in action monitoring. *Nature Neuroscience*, 3:516–520, 2000.
- W.J. Gehring and A.R. Willoughby. The medial frontal cortex and the rapid processing of monetary gains and losses. *Science*, 295:2279–2282, 2002.
- W.J. Gehring, M.G.H. Coles, D.E. Meyer, and E. Donchin. The error-related negativity: An event-related brain potential accompanying errors. *Psychophysiology*, 27:S34, 1990.
- W.J. Gehring, B. Goss, M.G.H. Coles, D.E. Meyer, and E. Donchin. A neural system for error detection and compensation. *Psychological Science*, 4:385–390, 1993.
- S.L. Gonzalez Andino, R. Grave de Peralta Menendez, G. Thut, J. del R. Millán, P. Morier, and T. Landis. Very high frequency oscillations (VHFO) as a predictor of movement intentions. *NeuroImage*, 32:170–179, 2006.
- R. Grave de Peralta Menendez, S.L. Gonzalez Andino, S. Morand, C.M. Michel, and T. Landis. Imaging the electrical activity of the brain: ELECTRA. *Human Brain Mapping*, 91:1–12, 2000.
- R. Grave de Peralta Menendez, S.L. Gonzalez Andino, G. Lantz, C.M. Michel, and T. Landis. Noninvasive localization of electromagnetic epileptic activity. I. Method descriptions and simulations. *Brain topography*, 14:131–137, 2001.
- R. Grave de Peralta Menendez, M.M. Murray, C.M. Michel, R. Martuzzi, and S.L. Gonzalez Andino. Electrical neuroimaging based on biophysical constraints. *NeuroImage*, 21:527–539, 2004.
- R. Hester, C. Fassbender, and H. Garavan. Individual differences in error processing: A review and reanalysis of three event-related fMRI studies using the GO/NOGO task. *Cerebral Cortex*, 14:986–994, 2004.

- L.R. Hochberg, M.D. Serruya, G.M. Friehs, J.A. Mukand, M. Saleh, A.H. Caplan, A. Branner, D. Chen, R.D. Penn, and J.P. Donoghue. Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature*, 442:164–171, 2006.
- C.B. Holroyd and M.G.H. Coles. The neural basis of human error processing: Reinforcement learning, dopamine and the error-related negativity. *Psychological Review*, 109:679–709, 2002.
- C.B. Holroyd, S. Nieuwenhuis, N. Yeung, and J.D. Cohen. Errors in reward prediction are reflected in the event-related brain potential. *Cognitive Neuroscience and Neuropsychology*, 14:2481–2484, 2003.
- P.R. Kennedy and R.A. Bakay. Restoration of neural output from a paralyzed patient by a direct brain connection. *Neuroreport*, 9:1707–1711, 1998.
- P.R. Kennedy, R.A. Bakay, M.M. Moore, K. Adams, and J. Goldwaithe. Direct control of a computer from the human central nervous system. *IEEE Transactions on Rehabilitation Engineering*, 8:198–202, 2000.
- Y.-H. Kim, D.R. Gitelman, A.C. Nobre, T.B. Parrish, K.S. Labar, and M.M. Mesulam. The large scale neural network for spatial attention displays multi-functional overlap but differential asymmetry. *NeuroImage*, 9:269–277, 1999.
- K. Kira and L.A. Rendell. The feature selection problem: Traditional methods and a new algorithm. In *Proceedings of the Ninth National Conference on Artificial Intelligence*, 1992.
- T. Kohonen. *Self-Organizing Maps*, 2nd ed. Springer-Verlag, Berlin, 1997.
- A. Kostov and M. Polak. Parallel man-machine training in development of EEG-based cursor control. *IEEE Transactions on Rehabilitation Engineering*, 8:203–204, 2000.
- P. Luu, D.M. Tucker, D. Derryberry, M. Reed, and C. Poulsen. Electrophysiological responses to errors and feedback in the process of action regulation. *Psychological Science*, 14:47–53, 2003.
- D. McFarland and J.R. Wolpaw. Sensorimotor rhythm-based brain-computer interface (BCI): Feature selection by regression improves performance. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 13:372–379, 2005.
- M.M. Mesulam, A.C. Nobre, Y.-H. Kim, T.B. Parrish, and D.R. Gitelman. Heterogeneity of cingulate contributions to spatial attention. *NeuroImage*, 13:1065–1072, 2001.
- A.C. Metting Van Rijn, A. Peper, and C.A. Grimbergen. High quality recording of bioelectric events (Part 1): Interference reduction, theory and practice. *Medical and Biological Engineering and Computing*, 28:389–397, 1990.
- C.M. Michel, M. Murray, G. Lantz, S.L. Gonzalez Andino, L. Spinelly, and R. Grave de Peralta Menendez. EEG source imaging. *Clinical Neurophysiology*, 115:2195–2222, 2004.

- M. Middendorff, G. McMillan, G. Calhoun, and K.S. Jones. Brain-computer interfaces based on steady-state visual evoked response. *IEEE Transactions on Rehabilitation Engineering*, 8:211–213, 2000.
- J. del R. Millán. Brain-computer interfaces. In Michael A. Arbib, editor, *Handbook of Brain Theory and Neural Networks*. The MIT Press, 2002.
- J. del R. Millán, M. Franzé, J. Mouriño, F. Cincotti, and F. Babiloni. Relevant EEG features for the classification of spontaneous motor-related tasks. *Biological Cybernetics*, 86:89–95, 2002a.
- J. del R. Millán, J. Mouriño, M. Franzé, F. Cincotti, M. Varsta, J. Heikkonen, and F. Babiloni. A local neural classifier for the recognition of EEG patterns associated to mental tasks. *IEEE Transactions on Neural Networks*, 13:678–686, 2002b.
- J. del R. Millán, F. Renkens, J. Mouriño, and W. Gerstner. Non-invasive brain-actuated control of a mobile robot by human EEG. *IEEE Transactions on Biomedical Engineering*, 51:1026–1033, 2004a.
- J. del R. Millán, F. Renkens, J. Mouriño, and W. Gerstner. Brain-actuated interaction. *Artificial Intelligence*, 159:241–259, 2004b.
- W.H.R. Miltner, C.H. Braun, and M.G.H. Coles. Event-related brain potentials following incorrect feedback in a time-estimation task: Evidence for a “generic” neural system for error detection. *Journal of Cognitive Neuroscience*, 9:788–798, 1997.
- J. Mouriño. *EEG-based analysis for the design of adaptive brain interfaces*. PhD thesis, Centre de Recerca en Enginyeria Biomèdica, Universitat Politècnica de Catalunya, Barcelona, Spain, 2003.
- S. Musallam, B.D. Corneil, B. Greger, H. Scherberger, and R.A. Andersen. Cognitive control signals for neural prosthetics. *Science*, 305:258–262, 2004.
- M.A.L. Nicolelis. Actions from thoughts. *Nature*, 409:403–407, 2001.
- M.A.L. Nicolelis. Brain-machine interfaces to restore motor function and probe neural circuits. *Nature Reviews Neuroscience*, 4:417–422, 2003.
- S. Nieuwenhuis, K.R. Ridderinkhof, J. Blom, G.P.H. Band, and A. Kok. Error-related brain potentials are differently related to awareness of response errors: Evidence from an antisaccade task. *Psychophysiology*, 38:752–760, 2001.
- B. Obermaier, G.R. Muller, and G. Pfurtscheller. “Virtual keyboard” controlled by spontaneous EEG activity. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11: 422–426, 2003.
- L.C. Parra, C.D. Spence, A.D. Gerson, and P. Sajda. Response error correction—A demonstration of improved human-machine performance using real-time EEG monitoring. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11:173–177, 2003.

- R.D. Pascual-Marqui. Standardized low resolution brain electromagnetic tomography (sLORETA): Technical details. *Methods & Findings in Experimental & Clinical Pharmacology*, 24D:5–12, 2002.
- R.D. Pascual-Marqui, C.M. Michel, and D. Lehmann. Low resolution electromagnetic tomography: A new method for localizing electrical activity in the brain. *International Journal of Psychophysiology*, 18:49–65, 1994.
- R.D. Pascual-Marqui, M. Esslen, K. Kochi, and D. Lehmann. Functional imaging with low resolution brain electromagnetic tomography (LORETA): A review. *Methods & Findings in Experimental & Clinical Pharmacology*, 24C:91–95, 2002.
- W. Penfield and T. Rasmussen. *The Cerebral Cortex of Man*. MacMillan, New York, 1950.
- W.D. Penny, S.J. Roberts, E.A. Curran, and M.J. Stokes. EEG-based communication: A pattern recognition approach. *IEEE Transactions on Rehabilitation Engineering*, 8:214–215, 2000.
- G. Pfurtscheller and F.H. Lopes da Silva. Event-related EEG/EMG synchronization and desynchronization: Basic principles. *Clinical Neurophysiology*, 110:1842–1857, 1999.
- G. Pfurtscheller and C. Neuper. Motor imagery and direct brain-computer communication. *Proceedings of the IEEE*, 89:1123–1134, 2001.
- G. Pfurtscheller, C. Guger, G. Müller, G. Krausz, and C. Neuper. Brain oscillations control hand orthosis in a tetraplegic. *Neuroscience Letters*, 292:211–214, 2000.
- J. Philips, J. del R. Millán, G. Vanacker, E. Lew, F. Galán, P.W. Ferrez, H. Van Brussel, and M. Nuttin. Adaptive shared control of a brain-actuated simulated wheelchair. In *Proceedings of the 10th International Conference on Rehabilitation Robotics*, 2007.
- R. Plonsey. The nature of sources of bioelectric and biomagnetic fields. *Biophysical Journal*, 39:309–312, 1982.
- R. Quian Quiroga and H. Garcia. Single-trial event-related potentials with wavelet denoising. *Clinical Neurophysiology*, 114:376–390, 2003.
- J. Rickert, S. Cardoso de Oliveira, E. Vaadia, A. Aertsen, S. Rotter, and C. Mehring. Encoding of movement direction in different frequency ranges of motor cortical local field potentials. *The Journal of Neuroscience*, 25:8815–8824, 2005.
- K.R. Ridderinkhof, S. Nieuwenhuis, and T.R. Bashore. Errors are foreshadowed in brain potentials associated with action monitoring in cingulate cortex in humans. *Neuroscience Letters*, 348:1–4, 2003.
- M. Robnik-Sikonja and I. Kononenko. Theoretical and empirical analysis of Relief and RRelief. *Machine Learning*, 53:23–69, 2003.
- A. Rodriguez-Fornells, A.R. Kurzbuch, and T.R. Münte. Time course of error detection and correction in humans: Neurophysiological evidence. *The Journal of Neuroscience*, 22:9990–9996, 2002.

- G. Schalk, J.R. Wolpaw, D.J. McFarland, and G. Pfurtscheller. EEG-based communication: Presence of and error potential. *Clinical Neurophysiology*, 111:2138–2144, 2000.
- M.K. Scheffers, M.G.H. Coles, P. Bernstein, W.J. Gehring, and E. Donchin. Event-related brain potentials and error-related processing: An analysis of incorrect responses to Go and No-Go stimuli. *Psychophysiology*, 33:42–54, 1996.
- H. Scherberger and M.R. Jarvis. Cortical local field potentials encodes movement intentions in the posterior parietal cortex. *Neuron*, 46:347–354, 2005.
- R. Scherer, G.R. Müller, C. Neuper, B. Graimann, and G. Pfurtscheller. An asynchronously controlled EEG-based virtual keyboard: Improvement of the spelling rate. *IEEE Transactions on Biomedical Engineering*, 51:979–984, 2004.
- M. Scherg. Fundamentals of dipole source potentials analysis. *Advances in Audiology*, 6: 40–69, 1990.
- M.D. Serruya, N.G. Hatsopoulos, L. Paninski, M.R. Fellows, and J.P. Donoghue. Instant neural control of a movement signal. *Nature*, 416:141–142, 2002.
- C.E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27: 379–423 and 623–656, 1948.
- A. Subasi, A. Alkan, E. Koklukaya, and M.K. Kiymik. Wavelet neural network classification of EEG signals by using AR model with MLE preprocessing. *Neural Networks*, 18:985–997, 2005.
- E.E. Sutter. The brain response interface: Communication through visually induced electrical brain responses. *Journal of Microcomputer Applications*, 15:31–45, 1992.
- J. Talairach and P. Tournoux. *Co-planar stereotaxic atlas of the human brain*. Thieme, 1988.
- D.M. Taylor, S.I. Helms Tillery, and A.B. Schwartz. Direct cortical control of 3D neuroprosthetic devices. *Science*, 296:1829–1832, 2002.
- L. Uldry, P.W. Ferrez, and J. del R. Millán. Feature selection methods on distributed linear inverse solutions for a non-invasive brain-machine interface. *IDIAP-Com 07-04*, available at <ftp://ftp.idiap.ch/pub/reports/2007/uldry-idiap-com-07-04.pdf>, 2007.
- M. Ullsperger and D.Y. von Cramon. Error monitoring using external feedback: Specific roles of the habenular complex, the reward system, and the cingulate motor area revealed by functional magnetic resonance imaging. *The Journal of Neuroscience*, 23:4308–4314, 2003.
- H.T. van Schie, R.B. Mars, M.G.H. Coles, and H. Bekkering. Modulation of activity in medial frontal and motor cortices during error observation. *Nature Neuroscience*, 7:549–554, 2004.
- G. Vanacker, J. del R. Millán, E. Lew, P.W. Ferrez, F. Galán, J. Philips, H. Van Brussel, and M. Nuttin. Context-based filtering for assisted brain-actuated wheelchair driving. *Computational Intelligence and Neuroscience*, 2007.

- J.D. Victor, K. Purpura, E. Katz, and B. Mao. Population encoding of spatial frequency, orientation, and color in macaque V1. *Journal of Neurophysiology*, 72:2151–2166, 1994.
- F. Vidal, T. Hasbroucq, J. Grapperon, and M. Bonnet. Is the “error negativity” specific to errors? *Biological Psychology*, 51:109–128, 2000.
- J.J. Vidal. Toward direct brain-computer communication. *Annual Review of Biophysics and Bioengineering*, 2:157–180, 1973.
- J.J. Vidal. Real-time detection of brain events in EEG. *Proceedings IEEE*, 65:633–641, 1977.
- J. Wessberg, C.R. Stambaugh, J.D. Kralik, P.D. Beck, M. Laubach, J.K. Chapin, J. Kim, S.J. Biggs, M.A. Srinivasan, and M.A.L. Nicolelis. Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. *Nature*, 408:361–365, 2000.
- I. Wickelgren. Tapping the mind. *Science*, 299:496–499, 2003.
- J.R. Wolpaw and D.J. McFarland. Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *PNAS*, 101:17849–17854, 2004.
- J.R. Wolpaw, H. Ramoser, D.J. McFarland, and G. Pfurtscheller. EEG-based communication: Improved accuracy by response verification. *IEEE Transactions on Rehabilitation Engineering*, 6:326–333, 1998.
- J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, and T.M. Vaughan. Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 113:767–791, 2002.
- J.R. Wolpaw, D.J. McFarland, T.M. Vaughan, and G. Schalk. The Wadsworth Center brain-computer interface (BCI) research and development program. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11:204–207, 2003.

Curriculum Vitae

Pierre Ferrez
1936 Verbier

Swiss
Born January 4, 1975

Tel.: +41 78 611 96 76
pierre.ferrez@gmail.com



EDUCATION

École Polytechnique Fédérale de Lausanne (EPFL) <i>Materials Science Engineer</i>	1994-2001
Lycée Collège des Creusets, Sion <i>Scientific Maturity (University entrance)</i>	1989-1994

PROFESSIONAL EXPERIENCE

IDIAP Research Institute, Martigny <i>Research assistant</i>	since 2004
Institute for Research in Ophthalmology, Sion <i>Research assistant</i>	2002 - 2003
Practical Training - Polykem (PSE / EPFL), Lausanne	2000
Practical Training - Asulab (Swatch), Marin	1999

LANGUAGES

French	Mother tongue
English	Very good level, reading, speaking and writing, a stay of one month in England
German	Average level, two stays of one month in Germany
Swiss German	Basic understanding
Italian	Basic knowledge

PUBLICATIONS

Original peer-reviewed articles

P.W. Ferrez and J. del R. Millán. Error-Related EEG Potentials Generated During Simulated Brain-Computer Interaction. *IEEE Transactions on Biomedical Engineering*, 2007, to appear.

G. Vanacker, J. del R. Millán, E. Lew, P.W. Ferrez, F. Galán, J. Philips, H. Van Brussel and M. Nuttin. Context-based Filtering for Assisted Brain-Actuated Wheelchair Driving. *Computational Intelligence and Neuroscience*, 2007, to appear.

A. Buttfield, P.W. Ferrez and J. del R. Millán. Towards a Robust BCI: Error Recognition and Online Learning. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2006, vol 14, pp. 164-168.

R. Grave de Peralta, S.L. Gonzalez, L. Perez, P.W. Ferrez and J. del R. Millán. Non-Invasive Estimation of Local Field Potentials for Neuroprosthesis Control. *Cognitive Processing*, 2005, vol. 6, pp. 59-64.

P.W. Ferrez, S.R. Chamot, B.L. Petrig, C.J. Pournaras and C.E. Riva. Effect of visual stimulation on blood oxygenation in the optic nerve head of miniature pigs: A pilot study. *Klinische Monatsblätter für Augenheilkunde*, 2004, vol. 221, pp. 364-366.

Book chapters

P.W. Ferrez and J. del R. Millán. Error-Related EEG Potentials in Brain-Computer Interfaces. In G. Dornhege et al. (eds.), *Toward Brain-Computer Interfacing*, MIT Press, Cambridge, MA, 2007, pp. 291-301.

R. Grave de Peralta, S.L. Gonzalez, P.W. Ferrez and J. del R. Millán. Non-Invasive Estimates of Local Field Potentials for Brain-Computer Interfaces. In G. Dornhege et al. (eds.), *Toward Brain-Computer Interfacing*, MIT Press, Cambridge, MA, 2007, pp. 279-290.

J. del R. Millán, P.W. Ferrez and A. Buttfield. The IDIAP Brain-Computer Interface: An Asynchronous Multi-Class Approach. In G. Dornhege et al. (eds.), *Toward Brain-Computer Interfacing*, MIT Press, Cambridge, MA, 2007, pp. 103-110.

Conferences

F. Galán, M. Nuttin, E. Lew, P.W. Ferrez, G. Vanacker, J. Philips, H. Van Brussel and J. del R. Millán. An Asynchronous and Non-Invasive Brain-Actuated Wheelchair. *13th International Symposium on Robotics Research*, 2007.

J. del R. Millán, P.W. Ferrez, F. Galán, E. Lew, and R. Chavarriaga. Non-Invasive Brain-Actuated Interaction. *2nd International Symposium on Brain, Vision and Artificial Intelligence*, 2007.

F. Galán, P.W. Ferrez, F. Oliva, J. Guàrdia and J. del R. Millán. Feature Extraction for Multi-class BCI using Canonical Variates Analysis. *IEEE International Symposium on Intelligent Signal Processing*, 2007.

F. Galán, J. Palix, R. Chavarriaga, P.W. Ferrez, E. Lew, C.A. Hauert and J. del R. Millán. Visuo-Spatial Attention Frame Recognition for Brain-Computer Interfaces. *1st International Conference on Cognitive Neurodynamics*, 2007.

R. Chavarriaga, P.W. Ferrez and J. del R. Millán. To Err is Human: Learning from Error Potentials in Brain-Computer Interfaces. *1st International Conference on Cognitive Neurodynamics*, 2007.

J. Philips, J. del R. Millán, G. Vanacker, E. Lew, F. Galán, P.W. Ferrez, H. Van Brussel and M. Nuttin. Adaptive Shared Control of a Brain-Actuated Simulated Wheelchair. *10th IEEE International Conference on Rehabilitation Robotics*, 2007.

C. Menon, C. de Negueruela, J. del R. Millán, O. Tonet, F. Carpi, M. Broschart, P.W. Ferrez, A. Buttfeld, P. Dario, L. Citi, C. Laschi, M. Tombini, F. Sepulveda, R. Poli, R. Palaniappan, F. Tecchio, P.M. Rossini. and D. de Rossi. Prospects of Brain-Machine Interfaces for Space System Control. *57th International Astronautical Conference*, 2006.

A. Buttfeld, P.W. Ferrez and J. del R. Millán. Online Classifier Adaptation in High Frequency EEG. *3rd International Brain-Computer Interface Workshop*, 2006.

P.W. Ferrez, F. Galán, A. Buttfeld, S.L. Gonzalez, R. Grave de Peralta and J. del R. Millán. High Frequency Bands and Estimated Local Field Potentials to Improve Single-Trial Classification of Electroencephalographic Signals. *3rd International Brain-Computer Interface Workshop*, 2006.

E. Lew, M. Nuttin, P.W. Ferrez, A. Degeest, A. Buttfeld, G. Vanacker and J. del R. Millán. Non-Invasive Brain Computer Interface for Mental Control of a Simulated Wheelchair. *3rd International Brain-Computer Interface Workshop*, 2006.

J. del R. Millán, P.W. Ferrez, S.L. Gonzalez and R. Grave de Peralta. Non-Invasive Accurate Prediction of Arm Movements: A Harmless Approach to Neuroprosthetic Devices. *12th International Conference on Functional Mapping of the Human Brain*, 2006

P.W. Ferrez and J. del R. Millán. You Are Wrong!—Automatic Detection of Interaction Errors from Brain Waves. *19th International Joint Conference on Artificial Intelligence*, 2005, pp. 1413-1418.