From the Institute for Signal Processing

of the University of Lübeck

Director: Prof. Dr. Alfred Mertins

On Human-Machine Interfaces based on Electrical Brain Signals

Dissertation for

Fulfillment of Requirements for the

Doctoral Degree of the University of Lübeck

from the Department of Computer Sciences/Engineering

Submitted by Mehrnaz Khodam Hazrati

Born in Tehran, Iran

Lübeck, 2013

First referee: Prof. Dr. rer. nat. Ulrich G. Hofmann

Second referee: Prof. Dr. Rolf Verleger

Date of oral examination: October 14th, 2013

Approved for printing, Lübeck: March 18th, 2014

ACKNOWLEDGEMENTS

The last four years have been a memorable journey for me. I owe a debt of gratitude to many people who have supported me during my research life in Lübeck. This dissertation would never have been realized without the work of several experts, to whom I am deeply indebted.

In particular, I gratefully acknowledge the contribution of my committee members Prof. Dr. Verleger and Prof. Dr. Schrader and Prof. Dr. rer. nat Ulrich G. Hofmann for their time and input. I extend my sincere appreciation to Prof. Dr. Verleger for his superb fast and accurate proofreading and his precious comments, which have intensively increased the quality of my dissertation.

I deeply thank my supervisor for giving me the chance to pursue my PhD in the graduate school at Lübeck. This work owns its existence and success largely to the guiding inspiration and energy of my Supervisor, Prof Ulrich G. Hofmann for keeping me in high spirits with his unwavering good humor and for giving me space to follow my ideas. I cordially admire his wit and sharp mind in every single step of my project.

I would like to thank my master supervisor Prof. Dr. Abbas Erfanian for encouraging me to pursue a PhD degree abroad.

I am indebted to all ISIPers, my friends and colleagues in the institute for signal processing (ISIP), who helped me through this: Lei Chen, Yijing Xie, Simon Vogt, Mattias Pohl, Olaf Christ and Florian Müller. It has been pleasure to work with you.

I am grateful to all my former RISE interns during the summers of 2010-2012. They include Vivek Subramanian, Ethan Coffel and Jeffrey Weiss, as well as Haliza Mat Husin and Damar Adhika and Akshay Rangamani master students who worked intensively with me.

A number of people read significant portions of the text and provided valuable comments, criticism, and insight. In writing this manuscript I benefitted from conversation or correspondence with Prof. Jose C. Principe, Austin Brockmeier, Angelos Barmpoutis, and Bilal Fadlallah of the computational neuroengineering laboratory at university (CNEL) at the University of Florida.

Very special thanks to my husband Abolfazl for his unlimited love, patience and encouragement. Many thanks to my parents and my little sister Mehrnoosh for keeping me on track with their unconditional love and support and to my friends Aizhan, Katia and Jelena who inspired me to write and produce of my dissertation.

This work was supported by the Graduate School for Computing in Medicine and Life Sciences funded by Germany's Excellence Initiative [DFG GSC 235/1].

Mehrnaz Kh. Hazrati

Lübeck, August 2013

To my parents who dreamed it

To my husband who helped me realize it

On Human-Machine Interfaces based on Electrical Brain Signals

MEHRNAZ KHODAM HAZRATI

ABSTRACT

Nowadays brain machine interfacing (BMI) and its non-invasive branch, brain computer interfacing (BCI), is a marvelous scientific concept which describes any communication channel established directly from the brain, to the surrounding physical world by bypassing the natural neural-muscular peripheral system. The proliferation of sophisticated communication and electronic devices paved the way for a renaissance in brain computer interfacing. Each BCI is tailored for a specific application regarding the hardware employed and the computational techniques and algorithmic approaches used to form its elements. In this dissertation, novel mathematical methods as well as design considerations are combined to shape the imaginary-based BCI for control applications. For each application an overview of each component is presented and the entire system is evaluated. Each dataset was evaluated offline by calculating evoked potential (EP), power band spectrum and event-related spectral perturbation (ERSP) analysis.

In order to design a BCI several mental strategies can be used. This project addresses EEG classification during limb movement imagination and relaxation. For this purpose two essential aspects should be considered. On one side, the quality of imaginary tasks performed by the subjects and their ability to control their mind affects highly the performance of EEG-based BCI. On the other side, application of BCI systems heavily depends on the online EEG signal processing feasibility. During online experiments, the moving windows starting from the beginning of the trial were used for feature extraction and classification. Each trial session consisted of two consecutive phases, relaxation and hand movement imagination.

EEG-based brain-actuated control requires a classifier to discriminate between mental states. In this way, online feature extraction and classification of EEG signals are the most important parts. I investigated in particular a selection of feature extraction schemes exploited for the proposed BCI paradigms. These feature sets can be categorized under three groups:

Time-frequency analysis, fractal components, and higher order statistics. Initially the power spectrum was estimated to extract online features, which are theta, alpha, beta, and gamma power bands. The feature vectors were formed and fed into the classifier. There are many different types of classifiers, with some being far more effective than others. In this work, I investigated the effect of classification methods and their adaptation on the performance of BCI. The LDA, QDA and soft-SVM in the current work were implemented in a special architecture. The results demonstrated that the classification accuracy of ensemble SVM was significantly higher than that of the single SVM. The output of the decision making algorithm was presented to the subject during experiments as bio-feedback using simulated hand demonstration, a robotic hand prosthetics, a robot and an avatar in the virtual world. All experiments were carried out on human subjects.

PUBLICATIONS

1. **M. Kh. Hazrati**, A. Rangamani, J. C. Principe, and U. G. Hofmann, “A Robust Online Ocular Artifact Removal for EEG – An information-based Approach”, *Journal of Neurology*, under preparation, 2014
2. **M. Kh. Hazrati**, R. K. Almajid, J. Weiss, S. Oung, and U. G. Hofmann, “Controlling a simple hand prosthesis using brain signals”, Biomedical Technology congress BMT, 2014, submitted
3. **M. Kh. Hazrati**, and U. G. Hofmann, “Avatar Navigation in Second Life using Brain Signals”, *8th IEEE international symposium on intelligent signal processing (WISP)*, 2013
4. **M. Kh. Hazrati**, V. Subramanian, and U. G. Hofmann,” A Four-Class Brain Computer Interface for Robot Control”, *ISSNIP Biosignals and Biorobotic Conference, Vitoria, Brazil, January 2011.(Invited to submit extension in Journal of Medical and Biological Engineering)*
5. **M. Kh. Hazrati**, A. Erfanian and U. G. Hofmann, “Fractal Components From Electroencephalogram Provide Features For Brain Computer Interface”, *20^{th} Biennial International EURASIP, Biosignal, June 2010. (winner the student competition)*

The papers below have been written or co-authored by me during my PhD study, but they are not considered as part of this dissertation:

1. **M. Kh. Hazrati** , H. Mat Husin, and U. G. Hofmann, “Wireless brain signal recording based on capacitive electrodes”, *8th IEEE international symposium on intelligent signal processing (WISP)*, 2013
2. **M. Kh. Hazrati**, and U. G. Hofmann, “Decoding finger movements from ECoG signals using empirical mode decomposition”, *Biomedical Technology congress BMT*, 2012
3. D. Adhika, **M. Kh. Hazrati**, and U. G. Hofmann, “A new design for a brain computer interface by near infrared spectroscopy”, *Biomedical Technology congress BMT*, 2012
4. A. Brockmeier, **M. Kh. Hazrati**, W. J. Freeman, J. Principe, “Locating Spatial Patterns of Waveforms During Sensory Perception in Scalp EEG”, *34^{th} International Proc. IEEE EMBS*, Canada, 2012
5. M. Shabroo, M. A. Tootoonchian, and **M. Kh. Hazrati**, Rapid Prototyping Environment for Power Line Modem Design, Implementation, Verification, and Optimization, *International Journal of Engineering and Advanced Technology (IJEAT)*, Vol. 2, Issues. 1, 2012
6. K. K. Kandaswamy, G. Pugalenthi, **M. Kh. Hazrati,** K. U. Kalies, T. Martinetz, BLProt: prediction of bioluminescent proteins based on support vector machine and relieff feature selection. *BMC Bioinformatics*, Vol. 12, Issues. 345, 2011
7. **M. Kh. Hazrati**, and A. Erfanian, ”An On-line BCI Without Training for Controlling the Sequence of Hand Grasping, Opening, and Holding Using Adaptive Probabilistic Neural Network ,” *Journal of Medical Engineering and Physics*, April 2010.
8. **M. Kh. Hazrati**, and A. Erfanian, ”An On-line BCI for Control of Hand Grasping Sequence and Holding Using Adaptive Probabilistic Neural Network,” *30^{th} International Proc. IEEE EMBS, Canada, 2008.*
9. **M. Kh. Hazrati**, N. Hamedi, and A. Erfanian, ” Bicoherence Analysis of EEG signals During Hand Movement Imagination in Bain-Computer Interfacing,” *ICBME,* 2008.

Table of Contents

Chapter1
Overview and introduction[1]

1.1. Motivation: Communication in a wider sense

Theoretically, there are a myriad of possible communication methods between two humans (Arndt, 2001). Just using one's five senses, the amount of information one can send or receive through each of their natural communication channels is amazingly high. In a primary sense, a body gesture or an obvious physical behavior can be interpreted as a communication signal and consequently can be understood easily by another person (Tomasello, 2008). On a more abstract level i. e. within the linguistic domain, as Noam Chomsky (1966) explains, a human is able to construct an infinite variety of sentences using a finite set of rules to convey an almost uncountable variety of messages (Chomsky, 2006).

Beyond the capabilities of our conventional senses, we can also extend our communication with the outside world by exploiting technology (Daly I., 2012). Developments in computer science have changed our perspective towards machines and their usability. Nowadays machines are already inseparable parts of our lives within our wired world (Torrance, 1984); peer-to-peer computing is changing the way people interact (Loo, 2003). Thus it may indeed be seen desirable to "upgrade" the human body to improve or augment conventional forms of human communication by establishing new electrical/mechanical control and communication channels (Coyle, 2010). It can be notably useful for people who undergo serious physical communication impediments, e. g. paralysis, blindness or deafness, but it could also give the healthy ones new communication opportunities. The plasticity of the brain is the key to its learning and adaptation capabilities which will be vital for the extension of both afferent (sensory) and efferent (motor) nervous structures in the future (Dennett, 1992) (Coyle, 2010). It does not, however, convey that we should all strive to become Cybernetic organisms, or, to use the term introduced by Professor Kevin Warwick (1998), Cyborgs, which are theoretical constructs that intimately and

[1] Some parts of the present chapter have been peer reviewed and published in the papers written by **Mehrnaz. Kh. Hazrati**.

irreversibly integrate technology and biology. Yet, this construct offers an extended and novel sense of what it implies to be human (Daly I., 2012).

The interface between humans and computers in any case seems to be more complicated than thought before (Kuebler et al, 2001). The ability to communicate decreases with the information transfer rate dramatically if in our natural communication model above a computer is substituted on the other side (Wolpaw et al., 2002). The reason for this is described well in a recent paper from Schalk, who discusses the theoretical and practical possibility for direct communication between the brain and computers (Schalk, 2008). Our central and peripheral nervous system is a massively interactive computational system with the ability to process extremely diverse and dense sensory inputs, store coded and selected information, express thought and control the available output system (Scarabino et al., 2003). The lack of intellect and consciousness in machines, on one hand and the different computational policy on the other, keeps the speed of this improvement down. Looking closer, there are even more substantial differences between the human mind and computers. Apart from the limitations in input/output structures, the brain is not simply a binary decoder and interpretation and comprehension are concepts beyond simple programming. According to the current view, the human brain seems to be a vastly parallel system which continuously receives information from the environment (Hubel, 1979). It has a high degree of adaptation along with prediction abilities which could be interpreted coarsely as intelligence (Torrance, 1984). The human brain, in contrast to any computer, can plan, predict and postpone for a better or more effective result in the future (Tomasello, 2008). In a stunning paper, Alan Turing (1950) brings up a basic discussion about the computability of the brain's actions and wonders whether an analogous model can be implemented on a computer to create a universal machine (Turing, 1950). The question entails challenges and remains unexplored.

In bioengineering, human machine interfaces (HMI) or, equivalently, man machine interfaces (MMI) comprise literally any interaction between humans and machines other than those based on manual (or natural) contact. An interaction can involve one or more physical or cognitive aspects of humans such as the visual, auditory or tactile systems. Eye blinks, muscular signals and brain signal are common signals used in this area of research (Cannan et al., 2010). In practice each HMI system has a particular aim, performance level and set of limitations, but in theory is an upgrade to the human body. Among all disciplines, the symbiosis between the brain and computers is completely amazing. Brain machine interfacing (BMI) and its non-invasive branch brain computer interfacing (BCI) is a marvelous scientific concept to describe any communication channel established directly from the brain, to the surrounding physical world by bypassing the natural neural-muscular peripheral system, either for control, communication (Graimann et al, 2010) or for conveying an intention, as Schwartz describes (Schwartz, 2007). The relationship between these concepts is highlighted in Figure 1.1. Wolpow et al. defines the direct connection of the brain to its environment as an "effective bypass" (Wolpaw et al., 2002).

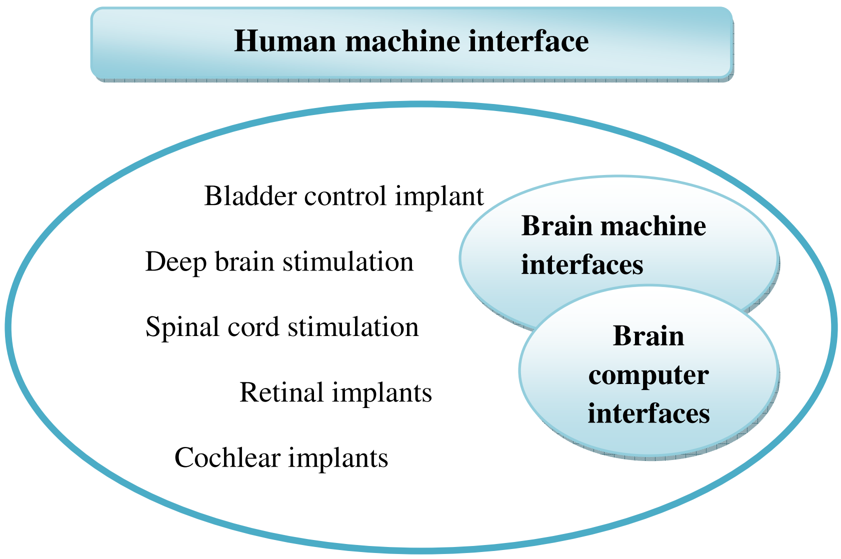


Figure 1.1: Hierarchical nomenclature of human machine interfacing: HMI comprises literally any interaction between the human and a machine other than those based on manual contact. Brain Computer interfacing is a non-invasive subgroup of Brain machine interfacing (Graimann et al, 2010).

In the course of development up till now, computational techniques, physiological discoveries and hardware design have been used in combination giving ascent to some type of brain-machine wiring. There are complex and subtle concepts of mind consciousness and brain functional structure, whose boundaries are to be discovered. So the task is not mundane; the topic remains challenging with open problems in modern science and related philosophical debates. Brain computer interfacing comes in various guises and structures. It can be categorized based on the type of brain signal, whether the computation strategy is online or offline, and whether the scenarios are cue-based or self-paced (Cannan et al., 2010) (Graimann et al, 2010).

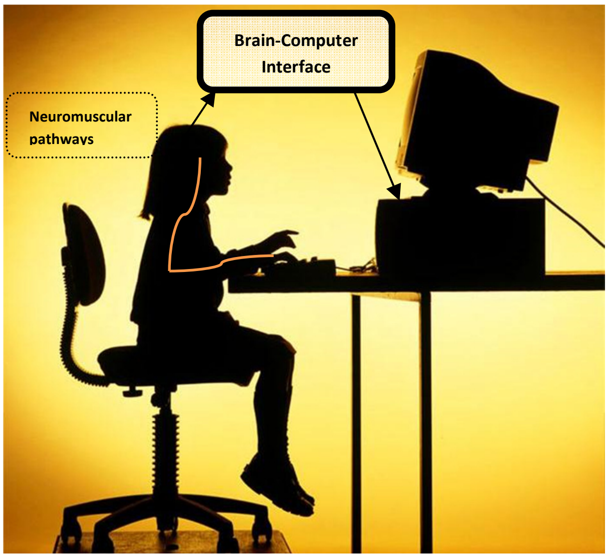
1.2. Brain computer interface (BCI)

Brain computer interfacing is an augmenting communication channel, which extends the ranges of human senses by establishing a direct connection between the brain and the outside virtual or physical world (Wolpaw et al., 2002). BCI is based on superficial brain signals alone (Figure 1.2), so no other biological signal is supposed to be any help to this type of communication (Graimann et al, 2010). Its initial goal is to establish a rehabilitation facility to patients who suffer from severe kind of motor impairments (Wolpaw et al., 2002).

Forfeiture of movement control in humans is a problem that is common worldwide and has a deeply negative effect on the patient's quality of life. It is estimated that nearly five million people suffer from loss of voluntary muscle control around the world which impair them to communicate in a normal way (Kuebler et al, 2001). The causes are diverse and may vary from spinal cord injury, stroke to amyotrophic lateral sclerosis (ALS) disease (Society of Neuroscience, 2008).

For the present purpose of BCI, two classes of syndromes may be distinguished: 1) Incomplete impairments of movement control when the patient is still able to communicate and BCI is used as a means to overcome the blocking of neural efference and control, e.g., to control movements of the right hand. Such syndromes are due to spinal cord lesions, which may lead to palsy of the two legs or (if the lesion is higher) to tetraplegia (palsy of all four limbs) or due to stroke (cerebral palsy) which may lead to hemiplegia (palsy of one side of the body) 2) Complete impairment of movement control when BCI is mainly used to provide a means of communication for the patient to the external world. This "locked-in syndrome" may be due to ALS or due to the stroke in the brain stem (Scarabino et al., 2003). ALS happens for unknown reasons, when brain and spinal motor neurons in the spinal cord begin to disintegrate (Society of Neuroscience, 2008). In people with ALS the motor neurons gradually degenerate, while their brain would maintain its intact cognitive function. When the disease reaches the last stage, they might lose all voluntary muscle control. Patients unable to communicate and unable to perform simple bodily functions are thus called "locked-in“ based on their mandatory need for full-time care (Townsend et al., 2010).

Using technology one might communicate in a different manner with either the surrounding physical world or any virtual environment. BCI as a branch under HMI must fulfill several conditions (Cannan et al., 2010): The BCI-system needs to receive direct input from the brain by evaluating brain waves of the user in real-time to make the decision based on the information extracted from each segment of the data. It leads to a control command, a physical action on a real or virtual object or any communication sent to a device, while the subject observes its result as biofeedback. The entire scenario should rely on the subject’s intention or a time locked paradigm. Once restricted to specially equipped laboratories, BCI today is almost ubiquitous. There are several labs and BCI research groups, who actively work on new ideas all over the world (Allison, 2012).



Edited from original photo: M12156RG [RM] © www.visualphotoes.com

Figure 1.2: Brain computer interfacing establishes a direct communication between the brain and the surrounding world (computer) and bypasses the afferent/efferent system. It may indeed be seen as a desirable way to "upgrade" the human body to improve or augment conventional forms of human communication by establishing new electrical/mechanical control and communication channels (Graimann et al, 2010)

1.3. A review on the current trend

Emerging from science-fiction stories like "Forbidden planet" in 1956 (Thomas, 2008) and well-known TV series like "Star Trek" in 1966 (Graimann et al, 2010) to the "Matrix" movies (Coyle, 2010), the ambitious idea of hooking up the human brain to a machine and translating the mere thinking power into physical or virtual actions has fascinated humans. Meanwhile scientist dealt with it as a serious research topic. In 1964 the neurophysiologist, Dr. Grey Walter investigated the brain signals of volunteer subjects while they were pressing a button. He found that the control signal actually appears before the movement happens. He introduced the concept of contingent negative variation (CNV) that appears following the conditional response (Walter W. G. et al., 1964). Later his research group reported that cerebral events resembling evoked potentials could also occur in the absence of expected stimuli (Weinberg et al, 1970). These early ideas established the fundamental viewpoint on designing such a system. Not being a science-fiction theme any more, in 1973 at the University of California Los Angeles (UCLA), Vidal introduced his online brain-based control game and the term brain computer interface to the research society (Vidal, 1973). Since then numerous BCI systems-under various guises- have been developed employing different signals and signal processing algorithms. They differ in design and performance level and address different groups of users. Building BCIs is an interdisciplinary field combining expertise in medicine, neurology, psychology, machine learning, statistics and signal processing, as well as philosophy (Kennedy et al., 2000) (Scarabino et al., 2003)

(Donoghue, 2007) (Schalk et al., 2007). Figure 1.3 depicts the general overview on the structure of a closed-loop Brain computer interface. Although BCIs were originally developed for assisting or augmenting human cognitive or sensory-motor functions, nowadays they are found in various applications for healthy users such as entertainment and gaming (Mason, 2006) (Lebedev et al., 2006). Nowadays several research groups worldwide try to realize the dream of translating the mere thinking power into mechanical or virtual actions. The result from a review on current trends in BCI shows a technological transition from isolated demonstrations to systematic research and commercial development in this area (Brunner et al., 2011). BCI presents a broad range of applications from a simple binary decision maker to more complex and sophisticated prosthesis control systems (Allison, 2012). Although the development of BCIs shows promising results, in most cases the reported measures of accuracy and information transfer rates come from highly controlled studies in the laboratory with precise indications of which targets the users should pursue (Mason et al., 2005). The use of BCI is still very limited in serious clinical applications (Kuebler et al, 2001) (Allison, 2012).

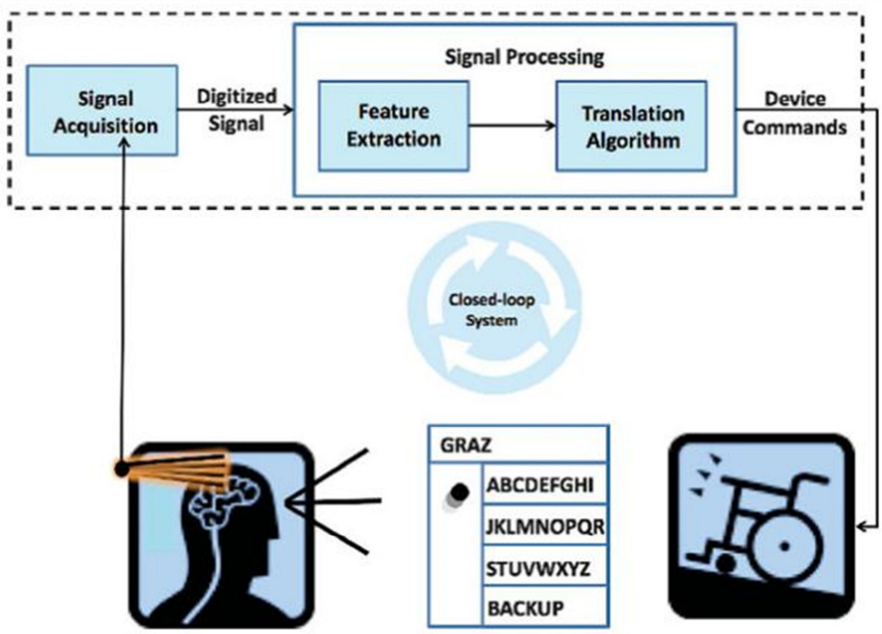


Figure 1.3: A general overview of the structure of a closed-loop brain computer interface for diverse applications such as wheelchair control and virtual keyboard use (adopted from (Graimann et al, 2010))

Applying advanced EEG processing methods to real-time EEG analyses has been used in the polygraph system, also known as the lie detector (Heussen, 2010), trust assessments (BinAb R. et al., 2009), brain fingerprinting (Parasuraman, 2007), cursor control interfaces (Kalcher et al., 1993) and home automation (Farwell, 1986). Virtual keyboards allow users to compose phrases and sentences just by thinking. BCI systems for mouse control can be used in 2D or 3D to facilitate interaction with the computer programs to access the internet (Schalk et al., 2007) (Daly I., 2012).

Before embarking on a tour of technical designs of Brain Computer Interfaces, we will briefly review the human brain anatomy along with the mechanisms underlying the generation of brain activity, the brain signals and their recording history. The rest of this chapter provides a general introduction to the concept of human machine interfaces and presents a background of the available computational tools for analyzing the brain.

1.4. Brain Anatomy

After cardio-centrist historical debates, Hippocrates (460-377 B.C.) was the first major historical figure to suggest the brain as the center of our understandings and all of our feelings (Coyle, 2010). The human brain is still the most astonishing riddle to be unveiled. Each individual obviously has a unique mind affected by his genetics, training and environment, which is developed by the life experiences, yet there are still physiological and logical similarities rendering it suitable for investigation (Attwood, 1989). David H. Hubel (1981 Nobel Prize Winner) describes the brain as a complicated, intricately woven tissue, like nothing else we know of in the universe (Hubel, 1979). This spongy three pound mass of fatty tissue controls all body activities and like every other living organ, is composed of cells. The central nervous system (CNS) embodies two main classes of cells. The majority of them has a protective and regulatory role and is called Glia. Neurons, on the other hand, are the brain's primary computing elements (Hubel, 1979) (Rothwell, 2001) and are highly specialized cells responsible for the brains' electrical functions. There are in total between one billion and one trillion active neurons inside the brain (Dayan et al., 2001). The number of neurons in the brain is very diverse amongst species. One estimate puts the human brain at about 100 billion (10^{11}) neurons and approximately 4-6 quadrillion (10^{15}) synapses (Thomas, 2008). Each neuron is able to "talk" to an average of 10,000 others (Society of Neuroscience, 2008). Neurons have characteristic that are unique remarkable among the cells of the body. Millions of neurons interact together in a complicated fashion which is more nuanced than just ones and zeros being coded. Each neuron seems to be an autonomous, analog computer of its own (Watson, 1997). There are electrical and chemical interactions which enable them to propagate signals rapidly over large distances. This is done by generating distinctive electrical pulses or spikes that can travel through nerve fibers (Attwood, 1989). A neuron simply signals its neighbors when it has information to be sent. Neurons are relatively slow processing elements compared to individual transistors elements in digital computers (Watson, 1997).

Thus, it is even more impressive to know that the brain can compute hundreds times faster than digital computers. The speed comes along with optimum power consumption (Watson, 1997). The outstanding speed and efficiency of the brain is the corollary of the massive parallelism and synchronization of the brain's hundred million neurons. It also reflects the strain of the analog computing that governs the neurons. A neuron has perhaps a hundred internal electrical levels, giving it far more information content than the binary on-off of a digital switch (Attwood, 1989). And instead of treating all inputs in a similar fashion, neurons can weigh the input pulses coming from certain favorite neighbors with simultaneously firing, according to the theory known as Hebb’s rule (Lytton, 2002). The tiny junction where one neuron receives input from its neighbor acts as a little memory element. Thus, each neuron is aware of its previous inputs. This property dramatically increases the efficiency by reducing the need for data to be swapped during a computation. Figure 1.4 demonstrates the structure of a typical neuron (Attwood, 1989).

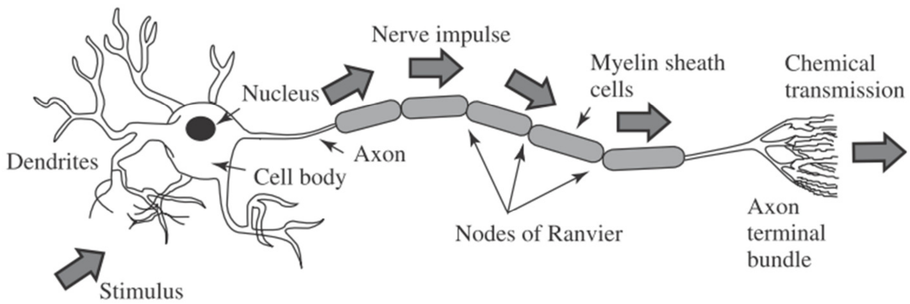


Figure 1.4: Rough structure of a neuron. The stimulus travels from dendrites to the axon terminal bundle and after that is transmitted chemically to the next neuron (adopted from Attwood and MacKay (Attwood, 1989)).

Dendrites are connected to either dendrites or the axons of other cells. They usually receive the impulses from other cells (Attwood, 1989). A stimulus travels along the axon. It is transmitted chemically, by the release of chemical neurotransmitters at the synapse, to send the message further (Attwood, 1989). The entire process of conveying the information is referred to the action potential (AP) (Sanei, 2007). There is a resting electrical potential of around -70mV between the soma (also called body of the neuron) and synapses which changes with variation in synaptic activities. When action potential starts, Na^+ ions rush into the cell and cause the electrical potential to rapidly change from -50 or -90mV to about +30mV (Roche Lexikon Medizin, 2003). This overshoot (depolarization) is then followed by a repolarization phase by closing sodium (Na^+) channels and opening potassium (K^+) channels. Usually after some millisecond, the cell returns to its resting potential (Sanei, 2007) (Dayan et al., 2001). Action potentials and synaptic currents transferred between the synapses contain information and are responsible for the functionality of the CNS (Scarabino et al., 2003). Neurons usually have a fantastic isolating cover, called the Myelin sheath, discovered by Rudolf Virchow (1821-1902), which speeds the transmission of electrical signals along the axon by saltatory conduction (Attwood, 1989). Neurons represent and transmit information by producing sequences of spikes in various temporal patterns. Axons code the information in time and propagate pulses in an "all or none fashion" called action potentials (Rothwell, 2001). Neural coding and measurement of stimulus characteristics, for instance the intensity of light or sound, as well as motor action, such as attributes of a limb movement are studied in the field of computational neuroscience (Chapin, 1999) (Schwartz, 2007). The brain is well structured and highly organized. Each group of neurons typically constitutes a functional section of the brain. A rough lateral and median view of the brain can be seen in Figure 1.5.

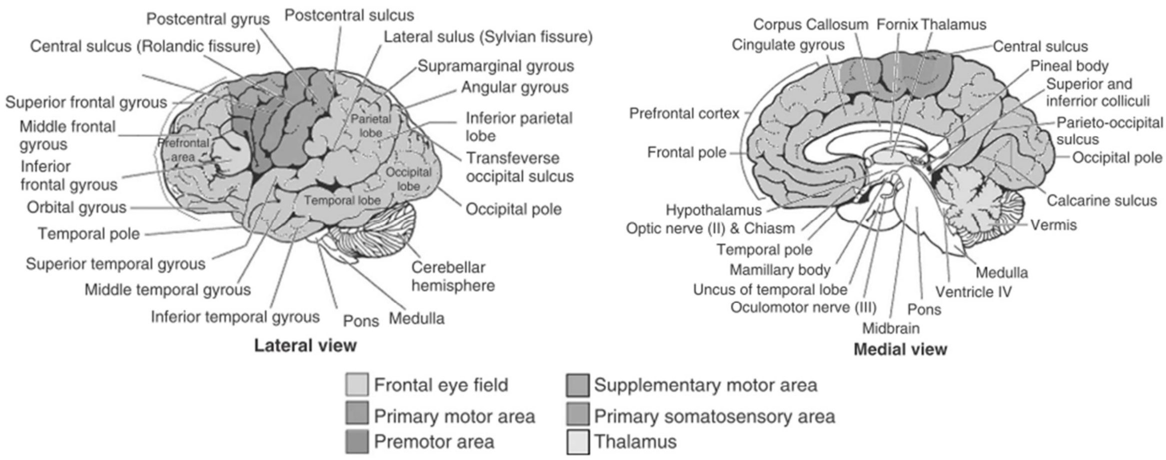


Figure 1.5: Lateral and medial view of the major parts of the brain: Frontal, Parietal, Temporal and Occipital lobes (Sanei, 2007)

Each part is dedicated to a particular task or contributes to a series of functions. The brain cortex serves as the center for thought, perception and memory. The major sections of the cerebral cortex are the frontal lobe, the parietal lobe, the temporal lobe and the occipital lobe. Each lobe is known to be responsible for a set of predefined functions (Scarabino et al., 2003). However, some particular regions could be associated with more than one task or function. In addition, accomplishing more complex functions, like speech or vision, needs cooperation between several regions of the brain (Daly I., 2012). The frontal lobe is credited for the highest intellectual functions such as planning and problem-solving (Daly I., 2012). It is also known that during mental tasks different regions within the brain have interaction or cross-talk (Vuckovic et al., 2008). Although the detailed mechanisms are still not completely understood, the synchronized electrical activities of groups of neurons in the cortex contribute to electrical potentials measurable on the surface of the scalp (Berger, 1930) (Nunez, 1981).

1.5. Brain signals

Brain activity in humans starts between the 4th and 5th month of prenatal development and not only reflects the brain functioning labor but also signifies the status of the entire body all over life (Sanei, 2007). The acquisition and processing of brain activity as a noise-like non-stationary signal requires accurate hardware and highly advanced computational methods. A successful trial was done initially in 1875 by an English scientist, Richard Caton (1842-1926), who used a galvanometer to measure the electrical brain activity from a human for the first time (Sanei, 2007). There has been a huge effort since then to localize and categorize the brain signals regarding the input presented to the brain or the output recorded from the peripheral limbs. Many researchers started to observe cerebral electrical activity over the cortex. Ernst Fleischl von Marxow (1845–1891) for example reported on observed visual cortex activity in animals (Fleischl, 1890). The inception of brain electromagnetic measurement was reported in 1912 by Vladmir Pravdich-Neminsky (1879-1952) who recorded the brain electrical activity called electrocerebrogram from the dura in dogs (Sanei, 2007). In the same years, graphical representations from external electrical stimulation of brain measurements were shown by Napoleon Cybulski (1854-1919) from an epileptic seizure in a dog (Sanei, 2007). In 1903 Willem Einthoven introduced his more accurate device for photographic recording of brain activity (Sanei, 2007). Later in 1928, the discovery of human brain signals under the current name of Electroencephalography by Hans Berger (1873-1941) changed the traditional view of the human mind and opened a new door to a fascinating way of understanding the human brain (Berger, 1933). He reported the first EEG recording using a bipolar electrode mounted on fronto-occipital sites in 1929 (Berger, 1930).

The neologism of Electro-encephalo-graph (EEG) then was used to describe the recording and demonstration (Figure 1.8) of the brain's electrical activities (Electro) emitting from the brain (Encephalo). Understanding the mechanisms underlying the generation of brain activity along with neuro-physiological and electrical aspects of the recording makes the EEG a coarse but valuable view on functionality of the human brain (Sanei, 2007). Discovery of EEG activity opened a window to the topography of brain function. It is now accepted that EEG reflects the integrative activities of synaptic potentials of millions of pyramidal cells in the cerebral cortex (Dayan et al., 2001). It means that extracellular ionic currents caused by dendritic electrical activity make the electrical potentials of the scalp and EEG (Wickelgren, 2003). So the ionic currents involved in the generation of fast action potentials contribute little to the averaged field potentials representing the EEG (Wickelgren, 2003). Although it can only detect large-scale neural dynamics, it is a rich information source and plays an important role in monitoring and controlling of brain function as well as in clinical diagnostic applications. It is a good tool for detecting different physiological and pathological conditions (Sanei, 2007).

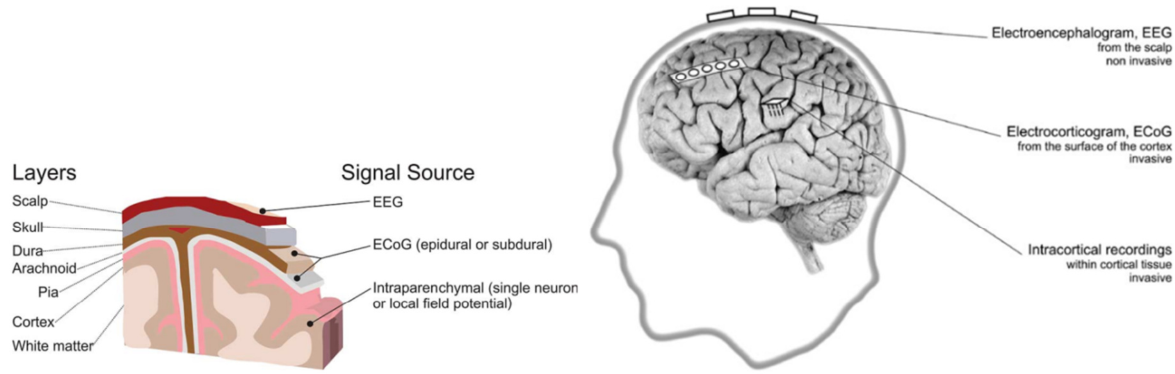
EEG is the main source of brain signals used in BCI research (Dornhege et al., 2007). It can be recorded as a time-varying difference in voltage between an electrode placed over the scalp and a reference electrode attached to the earlobe or other part of the body (Parasuraman, 2007). There are alternative recordings approaches such as Magnetoencephalography (MEG) (Dornhege et al., 2007), functional magnetic resonance imaging (fMRI) interpretation (Dornhege et al., 2007), positron emission tomography (PET) (Sanei, 2007), analysis of near infrared spectra (NIRS) (Karat et al., 2010), and thermography (Graimann et al, 2010). Magnetoencephalography (MEG) is a method for recording the magnetic fields associated with brain activity. fMRI and NIRS are both hemodynamic based techniques which measure small changes in the blood oxygenation level-dependent (BOLD) signals associated with cortical activation (Graimann et al, 2010). There is a clear correlation between absorption spectrum in the near-infrared wavelength and the change of oxy- and deoxy-Hb concentration in the tissue, which is in fact affected by the metabolic activity (Karat et al., 2010).

All mentioned methods are promising modalities for acquiring signals from the brain in a non-invasive way and have been employed for the development of brain–computer interfacing. However, there are drawbacks which make them unsuitable for a long-term solution in most BCI applications: MEG and fMRI are costly and bulky devices. PET and MRI suffer from poor temporal resolution (Coyle, et al., 2004). In the future NIRS may provide both temporal and spatial resolution (Coyle, et al., 2004), but this technique is still in an early stage of development and currently has poor temporal resolution (Graimann et al, 2010).

1.5.1. Invasive vs. noninvasive recording

Brain activity takes place on the molecular scale through electro-chemical processes and produces due to ion fluxes a variety of magneto-electrical signals that can be measured on different levels of resolution using diverse recording technologies (Dornhege et al., 2007). In general, there are three different approaches to detecting the brain's electrical activity: EEG, electrocorticograph (ECoG), and intracortical recordings (Schalk, 2011). EEG is a non-invasive approach, whereas other deeper, yet more precise approaches - ECoG and single neuron electrodes are considered invasive (Blakely et al., 2009). The level of invasiveness depends on the place of the electrode as is shown in Figure 1.6. For EEG recording the electrodes are placed over the surface of the scalp and for ECoG recording they are usually mounted over the cortex, e. g. primary motor cortex (Hochberg et al., 2012). ECoG does not damage any neurons (Schalk, 2010). Intracortical recording electrodes (O'Doherty et al., 2009) penetrate the brain tissue. By using this range of methods, the spiking of individual neurons can be recorded (Dornhege et al., 2007).

Pioneered in the early 1950s by Penfield and Jasper, Electrocorticography (ECoG) or intracranial EEG (iEEG) systems used electrodes placed on the exposed surface of the brain directly under the skull to gather rhythmic neural activity created by a large group of neurons in the cerebral cortex (Donoghue, 2007). The cortical potentials recorded by ECoG led to higher spatial resolution, broader band width, higher signal to noise ratio and less sensitivity to motional artifacts compared to recordings from the scalp (Schalk, 2010). This fascinating technology is usually used to identify epileptogenic zones in patients (Palmini, 2006) but in the future these signals could be targeted as the input to a Brain Computer Interface (BCI) (Jackson, 2012). Achieving higher reliability and accuracy is the agenda for designing and developing these new generations of HMIs. Signals recorded within the cortex would be better candidates to encode more information and might support BMI systems that require shorter training than EEG-based systems (Schalk, 2010), however in practice this issue demands more medical and technical consideration (Graimann et al, 2010). Current research targets their immediate use to augment the control abilities of amputees to steer the movement of a prosthetic hand and its fingers (Hochberg, et al., 2006).



(a) (b)

Figure 1.6: (a) A cut of protective layers of the Meninges: dura, arachnoid and pia matter are common recording domains (Schalk, 2011). (b) Three different approaches to detect the brain's electrical activity: EEG, ECoG, and intracortical recordings (Graimann et al, 2010)

Almost four decades ago, the first BMI experimental endeavor was undertaken by implementing an electrode on the motor cortex of monkeys (Fetz, et al., 1975) and training the brain to be able to control the external limb. Later on, similar research was done on rats (Nicolelis, 2001) to teach rats to access a sip of water by thinking about pushing a lever. This group recently introduced a closed-loop BMI for the first time, in which macaques were able to steer a robotic hand to grasp an object and simultaneously get tactile feedback (O'Doherty et al., 2011). The two dimensional movement of cursors or robotic arms were also successfully achieved by using the electrical signals captured form implanted electrodes in the rats or monkeys cortex (Carmena et al., 2003) (Lebdev et al., 2006). In 2003, as *Science* reported, a grid of 100 electrodes was surgically implanted into the brain of a young quadriplegic man. After establishing the BMI connection and performing the training session, he was able to check his emails or choose a TV channel by controlling his thoughts (Wickelgren, 2003). In 2007, a BCI based on an electrode implanted into the motor cortex of the human brain was successfully developed (Schwartz, 2007). One year later, Pistohl et al. reported the early stage development of neural interface systems (NISs), which was based on an intracortical microelectrode sensor. The initial human test was carried on paralyzed people where their motor cortex derived the control signals (Pistohl et al., 2008).

Invasive approaches without doubt lead to better signal quality, but at the cost of increases surgery risks and long-term safety considerations (Blakely et al., 2009). Ethical concerns, financial considerations and risk of neurosurgery usually prohibit the ECoG implementation for healthy BCI subjects and make it impractical for normal people (Blakely et al., 2009). Most research has been done on subjects who undergo neurosurgery for other purposes, such as treatment of epilepsy. These patients typically wear an ECoG system for a short time before it is removed, hence really efficient training and long-term studies are not possible (Graimann et al, 2010). A recent study from the University of Pittsburg reports a successful implementation of an ECoG BCI on a volunteer paraplegic patient, who achieved a satisfactory result in her hand control actions (Hochberg et al., 2012).

When it comes to potential BCI applications for everyday use, despite the benefits of ECoG recording, EEG has clear advantages. It provides a unique opportunity in terms of wearability and cost with acceptable temporal resolution, compared to other neuroimaging techniques. For these reasons, the field of Brain Computer Interfacing is largely based on EEG (Lebdev et al., 2006). EEG is also used extensively in other research areas such as neuroscience, cognitive science, and cognitive psychology (Society of Neuroscience, 2008). Special attention must be paid to the acquisition, analysis and processing of brain signals, since numerous clinical diagnoses and real-world applications rely on the information provided by the processing system (NaitAli, 2009).

1.5.2. Brain signal processing

Brain signals are believed to contain useful information not only about the brain but also about the entire body. Both EEG and ECoG deliver quite complex signals for processing, which spreads out the cumulative sum of dendritic activity and postsynaptic potentials from millions of neurons working together (Society of Neuroscience, 2008). Within these synchronized sources, ones that modulate movement activities or motor behavior are of most interest to be isolated and applied in an indirect control application (Wolpaw, 2012). However working with these signals is similar to try to perceive the result of a soccer match by just listening to the applause of audiences. In order to extract the hidden information in the brain signal many sophisticated acquisition and processing algorithm are required. This is the main impetuses behind applying advanced digital signal processing algorithms to the electroencephalogram signals measured from the brains of human subjects.

In order to record EEG brain activity, scalp electrodes are often mounted according to the standard 10-20 system. It is an internationally recognized and widely accepted measuring standard for EEG recording which was introduced in (Binnie et al., 1982). The scheme was developed to ensure standardized reproducibility for comparable studies over time and over different subjects (Dornhege et al., 2007). It is based on the proportional distances over the scalp, accounting for subjects head shapes and sizes. So the total distances for front-to-back and the right-to-left of the skull are first measured and the actual distances between neighbor electrodes are set to the segments of either 10% or 20% of those measures in each direction (NaitAli, 2009). Electrodes are usually identified by a letter indicating the recorded site followed by a letter indicating its position on the hemisphere. The first letter of each anatomical lobe, i.e. frontal, temporal, central, parietal, and occipital has been selected for this purpose. The letter "C" is used for identifying the central area and when electrodes are placed on the midline the letter "z" (zero) is added to their names. Electrode positions on the right hemisphere end in even numbers (2, 4 …) while the electrode positions on the left hemisphere end in odd numbers (1, 3 …). Moreover, the letter codes Fp and A identify the frontal polar and earlobes sites, respectively (Binnie et al., 1982). Often the earlobe electrode is called the reference electrode (NaitAli, 2009). Electrodes positioned over the frontal head are usually used to record eye blink or eye movement artifacts (Tatum et al., 2011). The number of electrodes required for signal acquisition depends on the application. Figure 1.7 gives a three and two dimensional view of the most commonly used electrode positions (Sanei, 2007).

In order to render the faint scalp signal suitable for further processing and visualization, it has to first pass through the amplification stage. The digital computing will be applied when the amplified signal passes through an analog to digital (ADC) circuit.

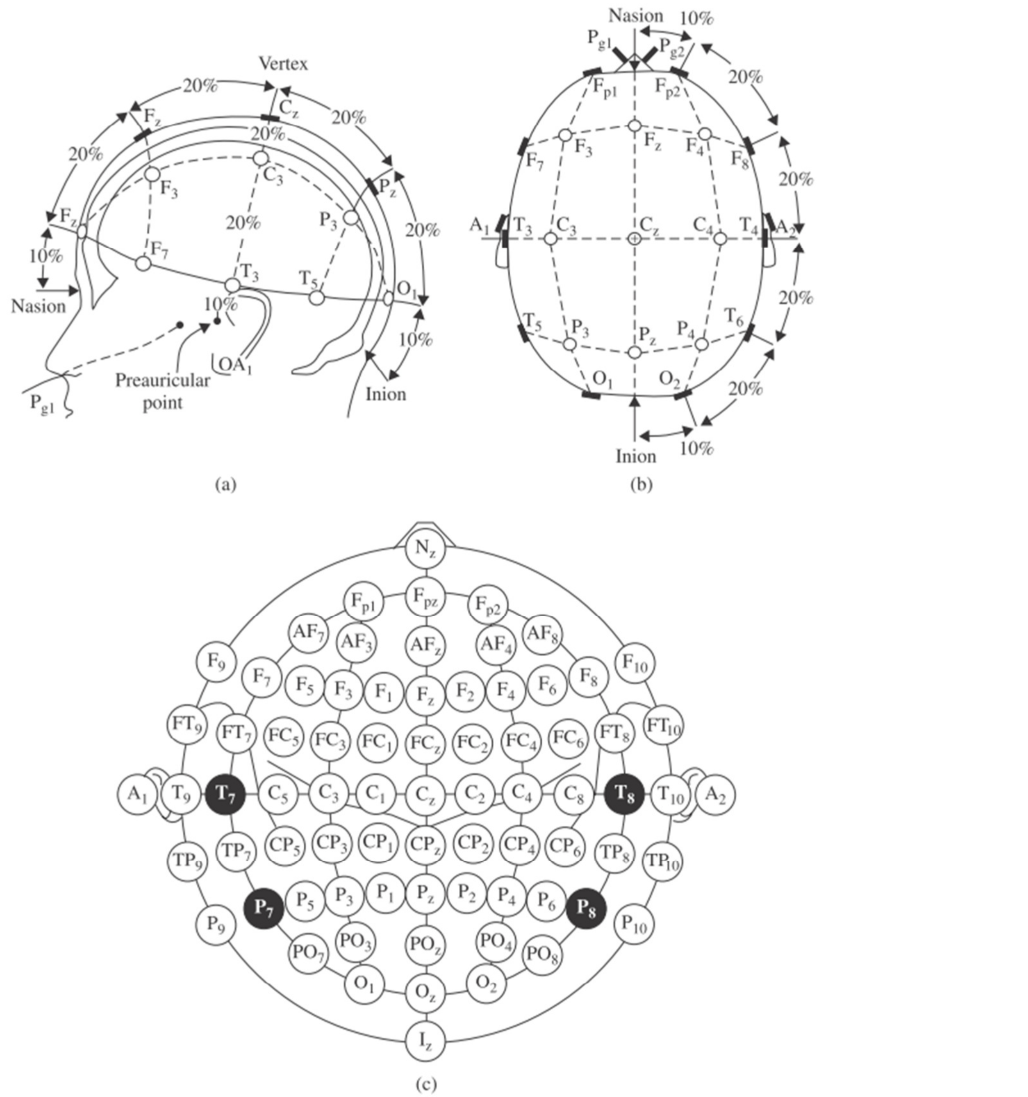


Figure 1.7: A diagrammatic representation of the 10–20 system (representing the three-dimensional measures) for EEG electrode positions including the reference electrodes: (a) shows the position of electrodes on the left side of the head, and the right side, (b) shows the view from above the head, and (c) shows a two-dimensional electrode configuration for 64 electrodes (Sanei, 2007).

Filtering could be applied either before or after this step to enhance the quality of the signal and to improve the signal to noise ratio (SNR). Different kinds of artifacts, either biological e. g. breathing and Electrocardiogram (ECG), electro-occulogram (EOG) or non-biological such as power supply and device noise, could affect the brain signal analysis. Electronic or digital filters, however, are intended to clean up these changes without adding any extra distortion or undesired changes to the signal. A low-pass filter with a cut-off frequency below 70Hz is usually applied to reduce the power of high frequency noise. High-pass filters with a cut-off frequency of approximately less than 0.5 Hz are used to reduce the very low frequency components of DC lines and breathing artifacts. The dominant 50 Hz power supply noise can be perfectly eliminated by using a notch filter with a null frequency of around 50 Hz or 60 Hz depending on the country. Figure 1.8 depicts 5s record of EEG signals in this project during a mental task gathered from 8 different electrode positions bases on10-20 system. A blink artifact is seen in the middle of recording.

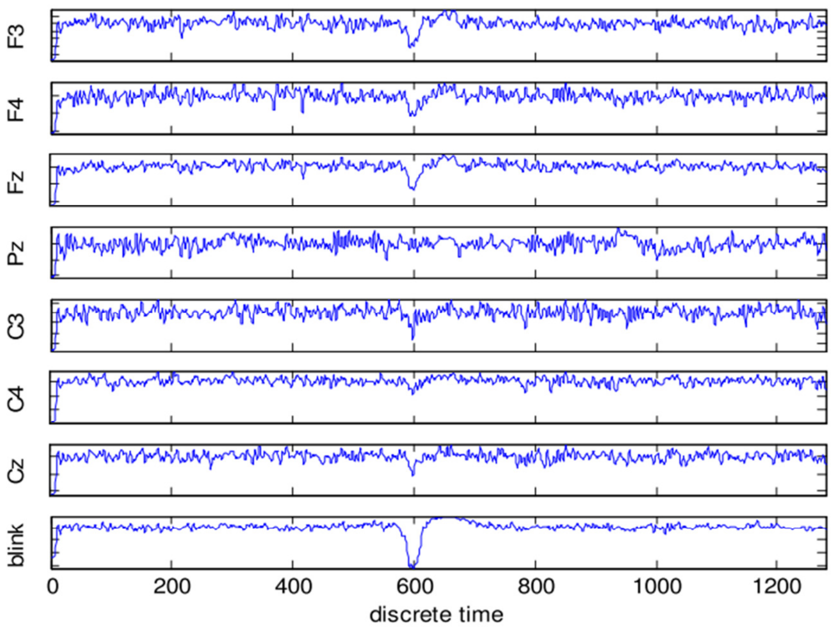


Figure 1.8: A sample of recorded signals in this project- a 5s record of normal EEG signals during a mental task gathered from 8 different electrode positions using the 10-20 system.

1.5.3. EEG Power Spectrum

The power spectral density describes "How much energy is contained at which frequency?" This method, however, assumes the signal does not change with time, i. e. it is stationary (Haykin, et al., 2007). The spectrum can only present the average power distribution in frequency domain but it is blind to the information in time domain. A fast method called Fast Fourier transform (FFT) was introduced in 1965 by Cooley and Tukey to compute power spectral analysis. Since then the FFT algorithm has become the most common approach to quantify the power distribution in EEG (Zhou-Yan, 2003) and to apply its frequency analysis on digital computers (Sanei, 2007). Brainwaves can be thought to consist of sinusoidal waves in specific bands of frequency or ranges. The power of each frequency band is proportional to the number of synchronized neurons working with that particular frequency (Pfurtscheller, 2006). Based on these frequency activities the spectrum of the brain signal can be segmented. This segmentation is not unique and might be described divergently in various documents (Sanei, 2007). By applying the power spectral analysis to the brain signal and filtering the different frequency ranges, the general rhythmic activity of the brain can be observed. Band power features are the most common and successful features used in the BCI field (Townsend, et al., 2006). They are also very common and successful in clinical applications (Niedermeyer et al., 1999). The spectrum of brain waves ranges from 0.5 to 500 Hz (Sanei, 2007), but usually the frequencies above 100Hz are not considered for analysis (Townsend, et al., 2006).

Important frequency bands waveforms are commonly divided in five major groups of frequency ranges (Townsend, et al., 2006) that are more clinically relevant (Niedermeyer et al., 1999). Typical dominant normal rhythms in EEG signals consist of: the delta (0.5-4 Hz), the theta (4-8 Hz), the alpha (8-13 Hz), the beta (13-30 Hz), and the gamma (30-70 Hz). There are other waveforms introduced by researchers based on different research interests but they are usually bonded to a particular task (Sanei, 2007). Among them: phi (less than 4 Hz) (Silbert et al., 1995), chi (11-17 Hz) (Silbert et al., 1995) and sleep spindles (11-15 Hz) (Hubel, 1979) are particularly popular (Sanei, 2007).

The power and amplitude of these bands change over time and are location dependent too. These spatio-temporal changes can be considered signs of brain activity or mental state such as attention, concentration, sensory stimulation, movement action and also some neurological or mental diseases (McFarland et al., 1997) (McFarland et al., 2006). Delta (δ) rhythm (0.5-4 Hz) is a high amplitude brain wave which is normally seen during deep sleep in adults as well as in infants and children (Niedermeyer et al., 1999). A strong presence of this band in awake adults can be a sign of an underlying brain lesion or may be indicative of a mental disorder (Blume et al., 1999). Theta (θ) waves (4-8 Hz) are typically a sign for low levels of alertness and mostly can be seen during waking up/falling asleep states (hypnagogic states). They are generated from the interaction between frontal and temporal lobes (Sanei, 2007). It is assumed that the predominance of these waveforms can be associated with thinking, memory consolidation and image-based and creative-intuitive activities (Niedermeyer et al., 1999) (Society of Neuroscience, 2008). Alpha (α) rhythm amplitude (8-13 Hz) is generated normally during the rest or relaxed state and it is attenuated when eyes are open or the individual becomes alert (Sanei, 2007). This band which has a center frequency around 10Hz is traditionally associated with pure awareness without processing and can be detected primarily in the occipital lobe (Berger, 1933). Beta (β) rhythm is seen in humans during concentration or mental activity (Blume et al., 1999) in the frontal and parietal lobes (Lytton, 2002). There are strong evidences that both alpha and beta rhythms are important for characterizing the imagination of the tongue, feet, left or right hand movements in the motor cortex area (Pfurtscheller et al., 1997). Mu (μ) (8-12 Hz) or beta (15-25 Hz) rhythm amplitude can be easily controlled by trained subjects (Pfurtscheller, et al., 2000). Beta band is divided into two sub bands: Beta1 (15-18 or 16-20 Hz) and Beta2 (20-38 Hz). Beta1 is more associated with the body presence and can be observed during the physical stillness. The upper frequency sub band in the Beta is associated with the active alert state including learning, task completion, and staying focused (Durka PJ., 2006) (Parasuraman, 2007). Gamma (γ) rhythm (also called fast beta wave) has a frequency above 30 Hz and is associated with hyper-vigilance or a highly focused state and to some extent has been correlated with anxious rumination (Graimann et al, 2010). Figure 1.9 (a) shows a typical demonstration of rhythmic activity of EEG. Each band has a specific spectral pattern over the scalp (Figure 1.9 (b)).

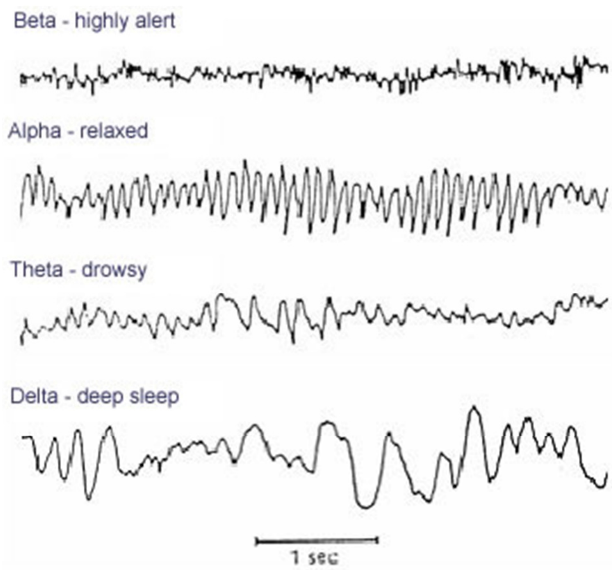


Figure 1.9: (a) A typical demonstration of frequency bands in EEG signals; Delta, Theta, Alpha and Beta.

Aside from these rhythms, there are some transient fluctuations which can be recorded over the scalp during mental activities or external stimulation.

1.5.4. Evoked potentials and event related potentials

Evoked potentials (EP) are elicited by any physical stimulus, consisting of a series of positive and negative deflections, largest at scalp sites overlying the modality-specific cortex: lateral occipital sites for visual stimuli, temporal and central sites for auditory stimuli, centro-parietal sites for somatosensory stimuli. Some researchers suggest that they are a result of the reorganization of the phases in ongoing EEG (Graimann et al, 2010). The term "event-related potential" (ERP) was coined to cover components of the EP that go beyond the mere reflection of physical attributes of a stimulus (and consequently do not have a modality-specific scalp distribution any more). The first published ERP component was the above-mentioned CNV (Walter et al., 1964). The most well-known ERP component is the P300, which is a positive change in EEG signal that occurs around 300 milliseconds after a relevant and/or infrequent stimulus, usually largest at parietal midline sites (Nijboer, 2008). P300 has been used successfully for developing a group of BCIs (Graimann et al, 2010).

Of particular relevance in the present context are movement-related ERPs. The Bereitschaftspotential (BP) (or "readiness potential", RP) has been first described by Kornhuber and Deecke (1965). Usually when no external stimulus is provided and the subjects are asked to repeat movement tasks in their pace for several times, BPs are good evaluation method. The decision making phase can be seen in the first period of BPs, while the period related to the preparation for the movement is joined with the second phase of BPs. Each phase affects different brain areas. In the first phase, mainly a large negativity can be observed over the supplementary motor area with the maximum at the center point Cz, and during the second phase the larger negativity is generated over the lateral motor cortex M1with stronger negativity in contralateral compared to the ipsilateral area of the involved hand (*review*: Shibasaki & Hallett, 2006). In the case that the experimental paradigm determines the pace of the movement through an external trigger, the first phase might be disappeared (Waszak et al., 2005) (Baker, Piriyapunyaporn, & Cunnington, 2012) and the second phase is called lateralized readiness potential (LRP) (Coles, 1989).

1.5.5. Event-Related Synchronization/ Desynchronization (ERS/ERD)

In order to analyze the changes related to sensory stimulation and motor behavior in the electrical activity of the cortex two different approaches are applied. In the first approach, i.e. ERP method, both time-locked and phase-locked changes are extracted from the ongoing EEG activity by applying averaging or other simple linear methods. This method is not proper for the induced changes which are time-locked but not phase-locked. In contrast to the ERP averaging, non-linear algorithms such as power spectral analysis or envelope detection should be applied to extract these changes (Pfurtscheller, 1992) (Graimann et al, 2010).

The fact about the time- and phase-locked responses can be explained in terms of the response of a stationary system and a non-stationary system. It can be assumed that an external stimulus induces existing neuronal networks of the cortex which will be both time- and phase-locked. In contrast, a change in the ongoing activity can be induced by applying the changes in the functional connectivity within the cortex. Consequently a class of BCI systems exists, that is based on the changes in the power of EEG rhythms corresponding to an event

(Karat et al., 2010). These changes, called Event-Related Desynchronization (ERD) and Event-Related Synchronization (ERS), can usually be seen in alpha and beta band during or right before a movement is executed (Pfurtscheller et al., 1993). Desynchronization yields a decrease in the power of the EEG rhythm during or before an event occurs. Conversely, Synchronization causes an increase in amplitude of an EEG pattern corresponding to the event. To assess either a decrease or an increase in brain power at a given frequency, one needs to compare brief reference periods like some few seconds before a movement takes place. ERD and ERS are usually used as features to detect the mental tasks related to the limb movements (McFarland, et al., 2000). Beta ERS occurs immediately after movement (Pfurtscheller et al., 1997). Gamma oscillation (30-45 Hz) in the form of ERS has been seen before movement binding to sensory and motor information (Pfurtscheller, 1992). It is assumed to be a carrier for the alpha and lower beta oscillations (Cheron, et al., 2007). It has been reported that alpha ERD represents the motor-related cortical activity more clearly than beta ERD does (Pfurtscheller, et al., 1999) (Durka PJ., 2006). During imagination of movement, amplitude attenuation happens in μ (8-13Hz) and central β (13-24Hz) oscillations at the contralateral sensorimotor area (Pfurtscheller et al., 2001). Amplitude enhancement also happens within the gamma band at the ipsilateral hemisphere (Pfurtscheller et al., 1997).

1.5.6. Event-related Spectral Perturbation (ERSP)

Makeig introduced a new method based on non phase-locked computation of a series of trials (Makeig, 1993). The aspect of information revealed by ERSP is not attainable while using ERP average of the same response epochs (Makeig, 1993). ERSP is a measure of event–related brain dynamics which is in principle a generalization of the ERD (Makeig, 1993). The ERD analysis is typically applied to a narrow band EEG. The full-spectrum ERSP, in contrast, covers the broad-band EEG frequency and presents the information over time. Consequently, it may yield to the better understanding of brain dynamics. The average relative changes in the dynamic of the EEG signal, can be captured by ERSP locked to a certain time point or the experimental event. The algorithm measures the amplitude of the EEG spectrum regarding an experimental event. It is assumed that similar experimental events induce fluctuations in the EEG spectrum that can be measured in average by ERSP (Cheron, et al., 2007).

The first step in computing an ERSP is calculating the baseline. The EEG signal preceding each event is used to calculate the baseline for each frequency. Then the overlapping windows scan the epoch data and the corresponding amplitude spectra passes through a moving average filter. In the next step, each filtered spectral transform is normalized regards to its corresponding mean baseline spectra. Finally, all normalized responses calculated from different trials are averaged to generate the ERSP, which is usually plotted as a time-frequency plane showing the log amplitude of relative spectral values (Makeig, 1993) (Bradley, et al., 2009).

These characteristics have a wide use in developing some of the recent BCI systems (Karat et al., 2010). Event-related potential (ERP) (Pfurtscheller et al., 2001), steady state visual evoked potential (SSVEP) (Kelly et al., 2005), slow cortical potential (Birbaumer et al., 2000), and P300 (Nijboer, 2008) are some famous and meaningful patterns among others (Allison, 2012).

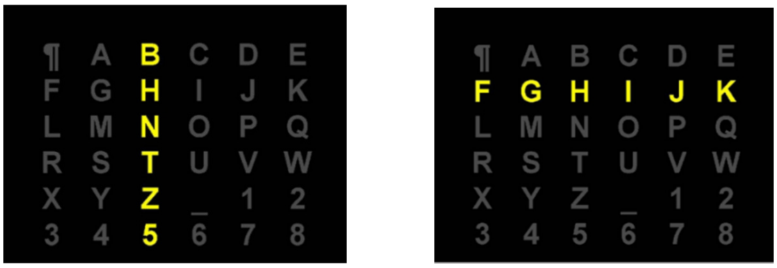
1.6. Mental strategies for BCI

BCIs at the current stage of development are able only to decipher particular patterns in the brain signal which correspond to specific events or a limited number of predefined mental tasks for the subject (Sanei, 2007). From this point of view, each BCI can be categorized under a subgroup according to the mental strategy it employs (McFarland et al., 2006) (Karat et al., 2010). The mental strategy is the foundation for building the other elements of BCIs. It determines the experimental strategy and the design approach. More specifically, the task required to be accomplished by the subject, the amount of training that the subject has had and the design selection of both hardware and software are all affected by this factor (Sepulveda et al., 2004). In order to develop such a system various strategies were employed (Mason et al., 2005) (McFarland et al., 2006). Motor imagery, steady state visual evoked potentials and P300 are common approaches (Nijboer, 2008). Two new branches of BCI, known as affective BCI and cognitive BCI (Daly I., 2012), respectively work based on emotional and cognitive states of the subject. Figure 1.12 demonstrates an overview on available BCI design strategies.

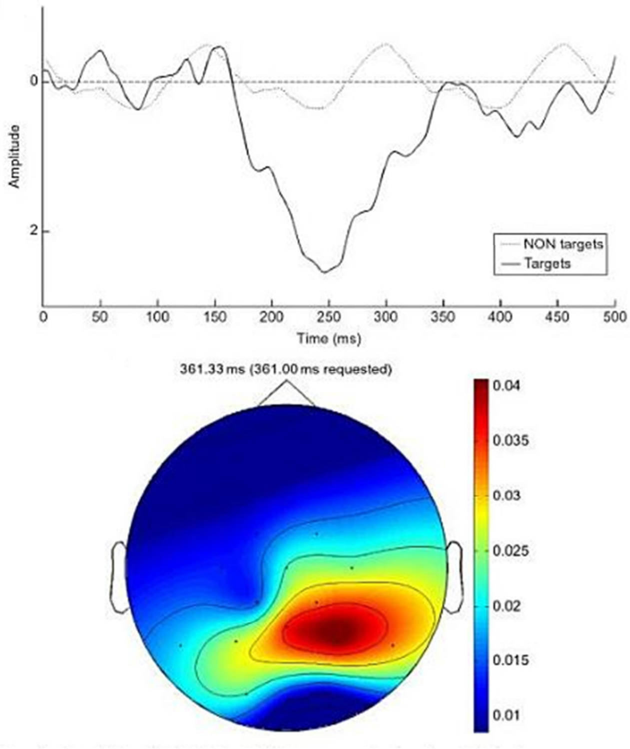
1.6.1. P300-based BCIs

P300-based BCIs form a popular class of BCIs which works based on selective attention in an oddball experiment paradigm. In this group of BCIs, the user should explicitly focus his gaze and pay attention to the external stimuli. When the preselected number or character is shown on the screen, an evoked potential called P300 appears in the brain signal (Mason et al., 2005). A screen shot of a typical P300 speller can be seen in Figure 1.10.

This BCI setup can be exhaustive in long-term usage, since the subject is constantly confronted with stimuli. Furthermore, some patient groups might have problems in focusing their gaze properly and therefore communication based on visual evoked potentials will not be reliable. Virtual keyboards or Yes/No questions are examples of this type of BCIs (Schouenborg, 2011) (Nijboer, 2008). Most of the people are able to control a BCI with this method (Guger, et al., 2009). P300-based BCIs are currently a popular topic in gaming and assistive technology development (Kaplan et al., 2013) (Halder, et al., 2013).



(a)



(b)

Figure 1.10: (a) A screen shot of a typical P300 speller. The user should select the Letter *H*. Each row and column blinks in a random order. In every trial all rows and columns are intensified. (b) P300 amplitude at 300ms and the activation of P300 at centro-parietal regions of brain (Schouenborg, 2011).

1.6.2. Steady State Visual Evoked Potential (SSVEP) based BCIs

SSVEP-based BCI systems work based on an external visual stimulus which has a constant flickering frequency. SSVEP is elicited by repetitive visual stimuli (Dornhege et al., 2007). Using this paradigm, the BCI user needs to focus his gaze for several seconds at the flickering option on the screen. The structure of the test is designed in such a way that usually up to four different frequencies between 8 and 20Hz are selected. The brain signal usually follows the dominant frequency and shows the higher power for the observed frequency in the power spectrum (Kelly et al., 2005). The same frequency and its second and third harmonics can be detected in the power spectrum (Bin G. et al., 2009). Fig. 1.11 shows a screen shot of a typical SSVEP experiment.

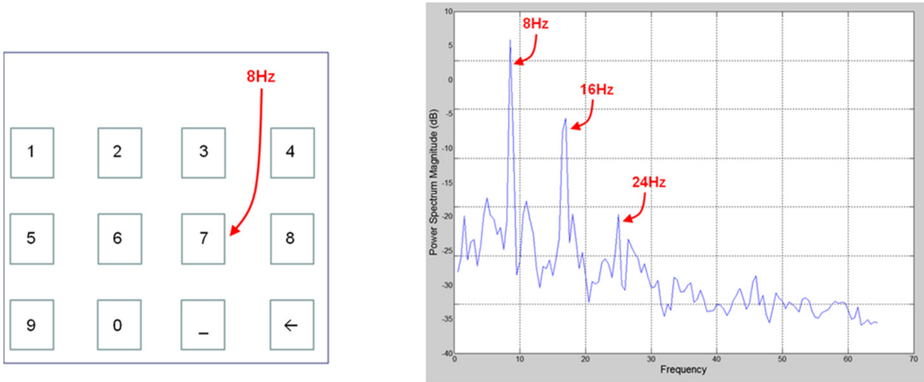


Figure 1.11: (a) A screen shot of a typical SSVEP experiment. (b) Power spectrum of the signal recorded during the experiment, when the subject looked at a button flicking in 8Hz.

1.6.3. Slow Cortical Potential (SCP) based BCI

This class of BCIs relies on changes in cortical potential below 2 Hz. SCPs are the voluntary production of negative and positive potential shifts in EEG (Hinterberger et al., 2003). Through feedback training subjects are able to control their SCP after a long training period of several months (Hinterberger et al., 2003). Positive and Negative SCPs correspond respectively to increase and decrease in cortical excitability. Employed by the BCI research group in Tübingen (Nijboer, 2008) a virtual keyboard called Thought-Translation-Device (TTD) was designed for totally paralyzed locked-in patients (Birbaumer et al., 2000).

1.6.4. Imagination of movement based BCIs

In the 90's some impressive studies discovered the correlations between EEG signals and mental tasks (Pfurtscheller et al., 1997) (McFarland et al., 1997) and the fact that the actual and imagined movements affect the EEG signal in a similar way (Pfurtscheller et al., 2001). In imaginary-based BCI experiments subjects are asked to perform several types of motor tasks, i.e. imagination of moving of the various body limbs, during the experiment (Pfurtscheller et al., 1997) (Graimann et al, 2010). At the contra-lateral and ipsi-lateral hemispheres sensori-motor representation area, imagination of either a left or right hand, tongue or feet movement results in amplitude attenuation or amplification in specific bands (Pfurtscheller et al., 1997). These changes can be applied as characteristics for classification of mental tasks (Vuckovic et al., 2006).

Movement imagery and the selective attention-based BCIs are among the most common mental strategies applied in BCI development (Mason et al., 2005) (Karat et al., 2010) (Vansteensel, et al., 2010) (Kaplan et al., 2013).

1.6.5. Emotional/Cognitive BCI

A relatively new group of BCI strategies are based on emotional or cognitive tasks, such as counting, calculation, mental rotation of an object or recalling a pleasant memory, done by the user during the experiment (Curran E et al., 2004) (Mason et al., 2005). By recognizing emotions a BCI will be able to provide a more natural way for a user to exert control and potentially lead to a higher information transfer rate (Molina, et al., 2009). If the user's emotional state is detected through brain activity patterns, it presents a more robust BCI to noise and signal deviations (Vansteensel, et al., 2010). In the following, we focus on BCI systems which work based on intentional control in a synchronized structure. Imagination of movement is the essential part of our EEG-based communication system.

1.7. Research groups

The number of active laboratories involved in BCI research is growing (Graimann et al, 2010). Many of the active BCI research groups worldwide are focused on "their" signal paradigm. Pioneers in developing BCIs among them are:

Prof. Gert Pfurtscheller (1999) and his Graz BCI team utilized the power of imagination to develop a cursor-control system. Using the same approach, they demonstrated the control of a hand prosthesis through the imagination of right and left hand movements by a totally paralyzed subject. Left or right hand movements will cause, for instance, ERD in the contralateral hemisphere of the brain over the motor cortex (Pfurtscheller et al., 2001) (Pfurtscheller, 2006). They found out that the brain cannot differentiate between reality and imagination. The neural activity during the imagination inside the brain is more or less similar to the one induced by the outside world (Pfurtscheller et al., 2001) (Graimann et al, 2010).

In 2003, Prof. Niels Birbaumer and his team at University of Tübingen, Germany developed the Thought Translation Device (TTD). The device is based on slow cortical potentials. The TTD users succeeded to compose their own phrases or sentences and also to search the web purely by thinking about it (Donoghue, 2007). Biofeedback is a core feature of some of these systems. It gives the information about the biological brain state back to the subject in an online mode. Using it patients can learn to control their state of mind to control and observe the output of the system (Birbaumer et al., 2000).

Prof. John Wolpaw's research group in Albany focuses on classification of EEG signals of real or imaginary movements and usually extracts features related to event-related desynchronization (ERD) of the mu rhythm (8-12 Hz) in their system (Wolpaw et al., 2000) (Wolpaw et al., 2002).

The Berlin BCI (BBCI) group supervised by Prof. Klaus-Robert Müller and Prof Curio is an active group in the area of minimizing the level of required subject training in BCI designs. Since 2000, they developed sophisticated algorithms based on machine learning to transfer the effort of training from the human subject to the machine and hence to reduce the inter-subject variability of BCI (Blankertz et al., 2006) (Blankertz et al., 2007).

1.8. BCI literature

During the last three decades, noninvasive recording of brain activities has been widely applied as a useful source of data for medical diagnosis and clinical applications (LeVan et al., 2006) (Barbati et al., 2004), as well as recent research on brain computer interfacing (Wolpaw et al., 2002). Past studies have demonstrated the ability of using EEG signals to control a computer or restore hand orthosis (Birbaumer et al., 2000) (Pfurtscheller, et al., 2000) (Pfurtscheler, 2003). For example, the 2000 article by Pfurtscheller et al. demonstrated the use of an EEG-based BCI to control an orthotic device to restore hand function in a tetraplegic patient (Pfurtscheller, et al., 2000). The prosthetic devices were controlled with the same motor imagery as the computer cursor, and did not require additional training (Pfurtscheller, et al., 2000). The patient, who could only move his upper left arm, restored some left hand functions with that technology. The orthotic hand device was controlled by motor imagery of the right hand and both feet. Motor imagery of the left hand was not used to avoid classification errors due to physical movement of the upper left arm. The BCI device used signals recorded from pairs of electrodes at the C4, C3 and Cz locations (Pfurtscheller, et al., 2000). Pfurtscheller et al. continued their work with this patient, and in 2003 reported the successful restoration of the patient's hand grasp using functional electrical stimulation (FES) controlled by the BCI system (Pfurtscheler, 2003). FES surface electrodes were used to stimulate the appropriate muscles to control the patient's hand without requiring an orthotic device. Only two pairs of EEG electrodes, located at the C3 and Cz locations, were necessary to control the BCI system using foot motor imagery. Closed, open, and relaxed hand states were cycled through by repetitions of foot motor imagery. The foot movement imaginations caused large beta oscillations that could potentially be detected by simple thresholding, without requiring a classifier. The major limitations of this work are the limited degrees of freedom, which may be improved by multi-channel EEG recordings, and lack of sensory feedback (beyond visual feedback), which can most likely only be provided by implantable systems (Pfurtscheler, 2003). Recently a hybrid BCI design was introduced by the same group (Pfurtscheller, et al., 2010).

Research with EEG-based BCI systems has since expanded to other applications such as robotics. A 2008 article describes the high-level control by BCI of a small robot, which could potentially function as a “helper robot” for disabled persons. Subjects controlled the robot by focusing on particular choices presented on a computer screen as they flashed, causing a strong P300 response. This type of response requires little training, but it is coupled to stimulus presentation, and therefore is most likely not suitable for prosthetic control. This study focused on automating most actions of the robot, and only requiring high-level decision commands from the user. This allows the system to work well despite the low bandwidth of EEG (Bell, et al., 2008).

Based on prior research, it is clear that a protocol for controlling a prosthetic hand with an EEG-based BCI system should start with only a few simple motor imagery-triggered BCI commands (Obermaier et al., 2001) (Vuckovic et al., 2006). The system should be able to autonomously complete motions, only requiring general high-level commands from the user. For example, particular movement imaginations could command the system to open or close the hand, without requiring the user to continuously control the precise state of the hand or separate fingers. After a simple system such as this is developed, experiments with multi-channel EEG may lead to finer control and more degrees of freedom. Such an improvement will likely require more user training and may lead to an increase in classification errors, which will also need to be addressed. There is only a minute variation in spatial distribution between ERD/ERS of the same limb doing different tasks (Vuckovic et al., 2008), so movement imaginations should also be selected carefully. This research is a good example of noninvasive BCI which is based on a two-level classifier in order to distinguish between the wrist flexion and extension movements of the same limb and also between movements of different limbs (Vuckovic et al., 2006). When the discrimination between movements of different limbs is desired, differences in spatial and temporal distribution between significant ERD/ERS of two different movements can be more obvious

1.9. Our Brain Machine Interface

Typically the first step in developing a BCI system is to explore the measurable characteristics in the brain signal which are controllable by the subject’s mind (Karat et al., 2010). Different brain strategies have been used to date for BCI design, e.g. motor imagery, steady state visual evoked potentials and P300. Essentially it is much more complicated than a typical ML setup, because in BCI both human brain and the computer are subjects to learning (Coates, 2008). In the computer side a learning phase is mandatory and it is achieved when the subject is instructed to perform a series of predefined mental activities (MAs) (Karat et al., 2010) and meanwhile the computer algorithm learns to extract the EEG patterns associated with mental states. These brain states hereafter are termed “classes”.

Our proposed BCI system relied on imagined movement. It is known that the brain activity associated with imagined movement produces reliable changes in the EEG which are similar to the brain activities produces with real movement. EEG signals recorded in an imagination-based BCI usually are not phase-locked to the onset of stimuli and averaging across trials tends to cancel out the changes. We improved and applied time-frequency domain approaches to extract features. Essentially the main role of BCI systems starts after the training phase is fulfilled. At this point the computer should be able to differentiate between mental states (classes) received from the subject’s thoughts using the ongoing brain signals. Before delving into the details of the project, we define the essential elements of the system.

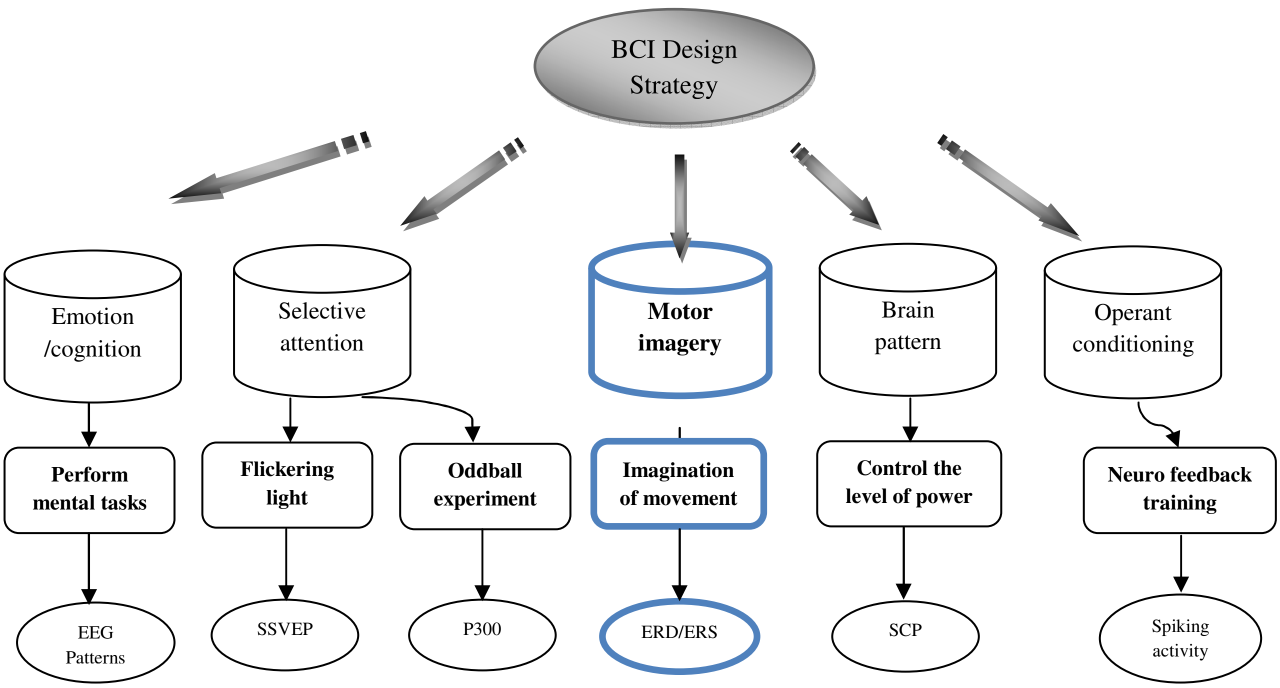


Figure 1.12: BCI design strategies according to the type of brain activity. In the current project the motor imagery is selected as the foundation of the BCI design.

1.10. Ergonomic design

Computer interfaces in general include any physical or software interface between humans and computers. At first this meant simple mouse, keyboard and joystick devices which emerged several decades ago. These were followed in the recent years by more natural and less restricted solutions such as bionics and BCIs (Allison, 2012). Ergonomics, or human engineering, is an expression in applied science which refers to the concept of maximizing the productivity and minimizing the operator fatigue and discomfort in these interfaces (Tangermann, et al., 2011). Ergonomic design as defined in (Carrol, 1997) as any interface that meets certain conditions such as being easy to use, easy to remember, effective to use, efficient to use, safe to use and finally enjoyable to use. The new interaction methods with higher quality and productivity should be designed to lessen the risk of error, discomfort fatigue, as well as injury (Coates, 2008). BCIs in long-term should be considered ergonomic devices and be designed according to ergonomic standards. In our primary design we tried to follow some important aspects of ergonomic design by making it easy, safe and effective to use in a rather fun environment (Tangermann, et al., 2011).

1.11. Online vs. Offline strategy

Typically a BCI consists of two major steps. The first step, offline mode, can be done when for each subject a sufficient amount of data (trials) is collected through experiments. The system then tries to extract a rule in order to determine the unseen signal in online mode. In online mode, as opposed to offline mode, the biofeedback is provided to the subject. Figure 1.14 displays and compares these two strategies.

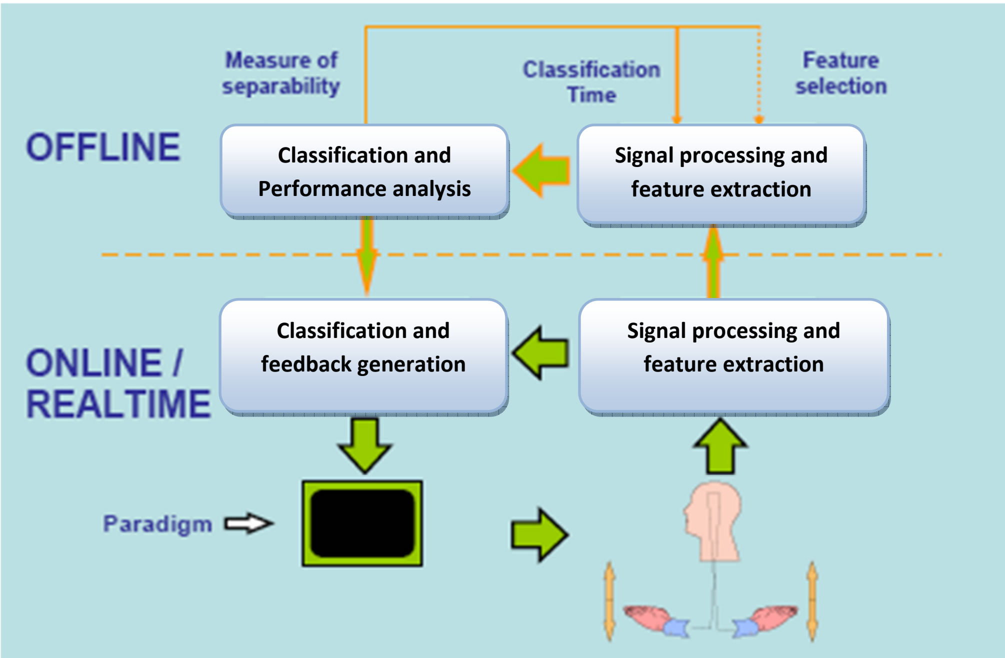


Figure 1.14: Offline and Online strategies in developing a brain computer interface

During the online experiment, both the brain and computer adapt themselves for better performance. The problem of interaction between two adaptive systems has been attracted several researchers in this area (Hinterberger et al., 2003). Repetition of experiments in different sessions may help to improve the functionality of the system and to enhance the subject's brain in governing the control with his brain signals (Blakely et al., 2009). We refer to this concept in Chapter 5 to explain our proposed BCI system.

1.12. Cue-paced (Synchronous) vs. self-paced (Asynchronous) strategy

In an ideal BCI system, the subject is able to control an artifact with the BCI system whenever he wishes. This type of BCI is called "self-paced" or "asynchronous" and until recently was not possible to implement (Mason, 2006). The emerging class of self-paced BCI systems introduced in (Mason et al., 2000) is still in its infancy and the related experiments are usually performed in intensely restricted conditions (Sanei, 2007). In practice and due to the sources of signals, most BCI paradigms added timing restrictions to the system. Hence processing the signal and consequently the decision was done in predefined time intervals. This is known as synchronized BCI (Allison, 2012). Figure 1.15 is a rough demonstration of a self-paced paradigm (Mason, 2006). The aim of a BCI system is to detect the brain activities associated with physical or cognitive events and to translate them to the commands of interest. The detection will be easier when the timing information of those events is known. The conventional synchronous BCIs are based on the evoked responses which occur when the neuronal activity changes due to the presentation of a stimulus (Karat et al., 2010).

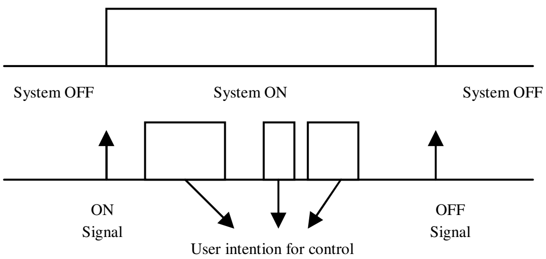


Figure 1.15: In a self-paced BCI, as long as the BCI system is in the ON state, the user's signal is processed and the system presents the corresponding feedback. In synchronized BCI, this can be done just in predefined intervals.

The expression endogenous BCI system is used when a spontaneous brain activity affected by thinking about the particular mental task runs BCI. When the BCI system depends on external stimuli or evoked activity is called exogenous system (Graimann et al, 2010).

1.13. Issues

Although the last several years of research have produced numerous kinds of BCIs with as many patients and promising results, several fundamental challenges still remain in this field (Allison, 2012). In order to leave the state of laboratory systems, the accuracy, information transfer rate, usability, and reliability of BCIs have to be improved significantly (Graimann et al, 2010).

Reported success stories usually are limited to specific subjects under controlled studies in laboratories with expensive and time consuming experimental setups (Allison, 2012) (Mason, 2006). Each research group seems to follow its own strategy, paradigm and instrumentation to collect and analyze the data, although some cutting-edge researchers have tried to design general purpose and extendable frameworks to encourage others to follow a predefined structure. BCI2000 and Open VIBE are the most famous examples of these frameworks (Schalk, 2010) (Renard, et al., 2010).

Another important issue in designing a BCI system is coping with significant subject to subject and day to day variations which leads to instability in the results (Blankertz et al., 2006). Physiological aspects of brain activities are not completely understood. The nature of the original sources and the characteristics of the medium affect the process in an unknown manner and consequently cause different results even under comparable restricted conditions. Communication rates and accuracy might also vary with mood, emotional state and the degree of fatigue or concentration of the user (Thomas, 2008). Sought for control signals are usually merged with the background EEG and separating them from each other is another important step in systematic design (Sanei, 2007). Meanwhile, removing artifacts in the EEG is a continuous challenge and a consensus method for fully efficient computational is still missing (Tatum et al., 2011). There are even more issues when dealing with the EEG-based BCIs online, because of the limitation in time and storage space. From the computational and processing point of view, advanced machine learning algorithms and well-designed interfaces could improve the acceptance and functionality of the available systems.

“BCI illiteracy” is a recent concept proposed by Blankertz et al. as a phenomenon in which approximately one fifth of individuals cannot learn to gain BCI control independent of their training or concentration level (Blankertz et al., 2006) (Vidaurre et al., 2010). Thus not only experience and learned adaptation but psychological and neurophysiological factors may modulate the ability to use a BCI. Hence it is important to be selective to prevent potentially unproductive or unnecessary training sessions (Allison, 2012).

1.14. Overview of the dissertation

Each BCI is tailored for a specific application regarding the hardware employed and the computational techniques and algorithmic approaches used to form its elements. In this dissertation novel mathematical methods as well as design considerations are combined to shape the imaginary-based BCI for control applications. The current chapter of this research essay presented a discussion on the significance and capability of EEG signals as a diagnostic tool for application in modern healthcare. It also illustrated a general definition of brain-computer interfacing for users not familiar with this relatively novel field of research. An overview of all required components is presented in the next chapter and then in each following chapter a standalone BCI application is described and evaluated.

1.14.1. The main structure of BCI

In order to gain insight into details, chapter 2 discusses the elements of a BCI system. Technically, BCI can be considered as a machine-learning system since the computer should learn how to decipher the associated EEG patterns. The relationships between given EEG patterns and the execution of the desired actions by the computer construct a pattern recognition or machine learning problem. Next chapter introduces the framework of the brain computer interface and addresses the theoretical design and machine learning aspects. The rest of this chapter covers the preprocessing techniques, features extraction approaches and classification methods proposed or employed in this dissertation. The first section discusses issues and challenges in processing brain signals and presents some solutions. The first section starts with an introduction to this concept. The functionality of the entire BCI system mostly depends on the choice of appropriate features and classifiers. I categorized each feature set based on the mathematical background and concepts. The first group consists of features based on power bands and spatio-temporal characteristics of the EEG signal. The next groups of features introduced in this chapter are statistical moments based on fractal components analysis, higher order statistics, and matching pursuit, which hold rather physical interpretations. In the following methods for classification of mental states are discussed. Popular classification techniques such as LDA, QDA and SVMs are explained and the idea of ensemble classification in introduced. Chapter 2 covers their background. This build the foundation of classifier employed in the next chapters.

1.14.2. BCI Applications

The next four chapters investigate Chapters 3 and 4 deal with BCI applications for hand grasp. These chapters use the previously recorded data to evaluate new signal processing algorithms for BCI. Chapter 3 presents an offline study to investigate fractal features over 6 subjects. On the preprocessing section of this chapter I propose an online ocular artifact reduction based on independent component analysis (ICA) and higher order statistic in BCI.

The proposed method were evaluated its performance using the datasets in the chapters 3 and 5. Results demonstrate that the proposed structure is a suitable alternative for online rejection of ocular artifacts from EEG signals without using an extra channel. ICA has, however, inherently some disadvantages. In chapter 4, I implement and evaluate an information-based method for online recognition and removal of ocular artifacts. This is a new kernel-based approach. Correntropy, a localized similarity measure, was used to evaluate the performance of the second method. We use the same method in chapter 6.

In chapter 4 a simulated hand and a custom built robotic arm (LAnDRoH) are controlled by receiving commands from our BCI system. Employing scientific game-based therapies and technologies also constructs another aspect of new BCI generation. In chapter 5 we report our designed game based BCIs for robot control and chapter 6 describes a BCI for navigating an avatar in virtual world interfaces. The last chapter summarizes the prevalent foundation of this research field, looks to the future, and considers possible challenges. These systems were evaluated using different criteria such as sensitivity, accuracy and information transfer rate. In each chapter we apply common statistical tests to evaluate our proposed methods.

Chapter 2

BCI Design and the main structure of its elements

2.1. BCI design

Over the last four decades, progress in the field of brain computer interfacing research was increasingly reported and several research groups and organizations conducted new and advanced strategies to setup BCI studies. Advances in artificial intelligence techniques and development of robotics and bioengineering paved the way for realizing these kinds of systems in everyday life. From the computational side of view, a typical BCI structure is depicted in Figure 2.1. In the design of the BCI system the following modules are required: (i) data acquisition, (ii) preprocessing and artifact reduction, (iii) data analysis, (iv) decision making, and (v) performance evaluation of the system (McFarland et al., 2006).

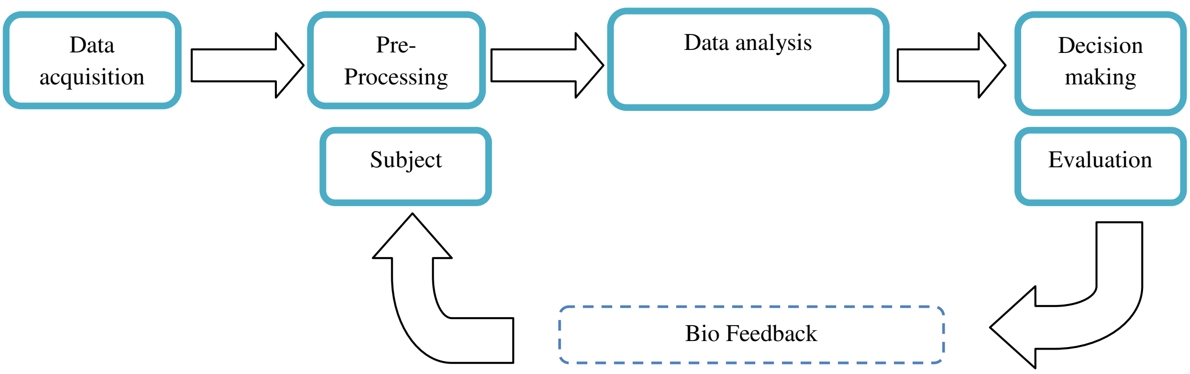


Figure 2.1: BCI chain: Typical structure for a brain computer interface. For the design of a BCI system the following modules are required: data acquisition, preprocessing and artifact reduction, data analysis, decision making, and performance evaluation of the system.

2.2. BCI : A Machine learning problem

The relationships between given EEG patterns and the execution of the desired actions by the computer construct a pattern recognition or Machine Learning (ML) problem

(Vuckovic et al., 2006) (Zhang, 2010). In the following we will have a closer look at the BCI design and its elements. By innovating and improving data analysis algorithms, EEG-based BCIs could be made a promising tool in many real world applications (Allison, 2012).

2.2.1. Machine learning

Learning is a general and widely used concept in the scientific literature (Zhang, 2010). According to the Britannica Concise Encyclopedia, it can be defined as a "process of acquiring modifications in existing knowledge, skills, habits, or tendencies through experience, practice, or exercise." The formulation of a learning metaphor plays an important role in machine learning and specifies the type or the degree of intelligence (Zhang, 2010). Chris Argyris, the Harvard Business school psychologist, delineates the concept of learning as "detection and correction of error", which in that "error" means "any mismatch between our intention and what actually happens" (Sepulveda et al., 2004). As Simon (1983), the winner of the Nobel Prize in Economic Sciences in 1978, states: "Learning denotes changes in the system that are adaptive in the sense that they enable the system to do the same task or tasks drawn from the same population more efficiently and more effectively the next time" (Zhang, 2010). The majority of learning techniques deal with static systems, whose characteristics do not change over time (Lotte et al., 2007). However, a few advanced methods handle time varying systems (McFarland et al., 2006). In order to establish a rational framework of biomedical signal and image analysis, we consider this factor in a wide sense: Techniques developed to acquire knowledge artificially and to apply this knowledge in the future. Based on this approach a computer is taught to simulate a child's brain instead of being programmed as an adult brain. This underpins the concept of machine learning.

Statistical learning theory or machine learning is a computational science which deals with establishing a logical-functional relation between the finite set of inputs and corresponding outputs from an unknown but assumed system. Machine learning, which originated from a conference under that name held at Carnegie-Mellon in 1980, is nowadays a rich and progressing area of research (Vuckovic et al., 2006). In essence, a machine learning algorithm takes the input data and finds a fitted model or structure, statistically or probabilistically, which has some minimized predefined cost function over the trained data (Zhang, 2010). In this model intelligence or knowledge is extracted from the training data. Machine learning has a wide scope of applications including predicting outcomes or simply mining knowledge from the data. It employs a variety of techniques spanning from feature extraction and feature selection techniques to regression and classification approaches for a wide set of algorithms such as support vector machines (Zhang, 2010), neural networks (Lotte et al., 2007), and genetic algorithms (Corralejo, et al., 2011). The topic is so rapidly growing that covering all literature available and all aspects in detail is not possible in this essay.

2.2.2. State of the art

Technically, BCI can be considered as a machine-learning system since the computer should learn how to decipher the associated EEG patterns. Several reviews and survey papers have been published to cover the widespread scope of machine learning in BCI (McFarland et al., 2006) (Graimann et al, 2010). Figure 2.2 shows four important steps of designing a BCI as a ML scenario.

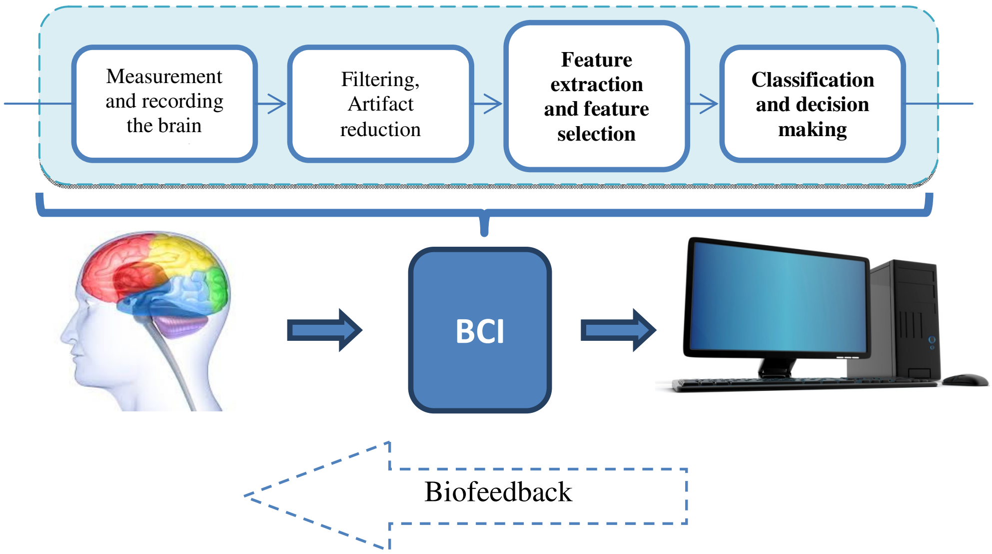


Figure 2.2: BCI as a machine learning problem consists of four steps. Signal processing starts with filtering and signal enhancement methods like artifact reduction, followed by the machine learning algorithm proper that includes 1) feature extraction, 2) feature selection, 3) classification, and 4) decision making techniques. The decision making step can be more advanced than just applying the output of the classifier directly to the external device.

Some studies showed that the performance of EEG-based BCIs can be higher than it was expected at the beginning. However the achievements highly depend on the extensive subject training and it is not stable over time. Therefore similar to other ML problems a training phase has to be accomplished to estimate the required parameters. In this phase the subject is asked to follow the instructions and to perform the prescribed mental activities (Mason et al., 2005). Data should be collected in such a sufficient amount that the main assumption of machine learning is fulfilled (Steinwart et al., 2008). The available input and output sets should be able to describe the structure of relations in a sufficient way (Vapnik, 1999). It is worth noting that in many other fields the challenge has shifted from this to gaining insight from the bulky amount of data which has been already collected (Zhang, 2010). The main goal, however, is not to decipher the system itself but to infer a valid mathematical relation or meaningful pattern which can be used for the prediction of future or unseen data (Steinwart et al., 2008).

Research has shown that new neural circuitry might be created when the brain encounters new problems and challenges (Jones et al., 2008). In some BCI designs the output of the system is presented to the user as biofeedback in order to improve the functionality of the entire system (Velliste et al., 2008). This makes the BCI model more complicated, since with existence of the biofeedback, the entire system should be scrutinized mathematically as two adaptive interactive systems (Vidaurre, et al., 2011). Here we take the simplified mode into consideration.

2.3. Main elements of a BCI

In the current chapter the state-of-the-art on structure of a BCI system is introduced. After the preprocessing step, feature extraction, feature selection and classification techniques for BCI will be presented. Techniques are presented in details in the next chapters when they are applied to each application. In each chapter I will go through one BCI application. It will be noticed that although the façade of each system is different the main structure is still similar.

2.3.1. Preprocessing

Like other biomedical signals, EEG signals are susceptible to a large range of artifacts and noise such as muscle movements, eye movement and blinking, cardiac activity, breathing movements and equipment interference (Ramoser et al., 2000) (McFarland et al., 2006). Artifacts are defined as undesired changes in measured EEG signals that originate from other sources than the subject's Central Nervous System (CNS) (Barlow, 1986). In general, artifacts can alter the shape of a neurological phenomenon or mimic any kind of electro cerebral activity (Verleger, 1991) that may drive an EEG-based system or make decision making or diagnostic interpretation difficult. Artifacts are the main source of quantitative miscalculation and misinterpretation during EEG analysis. Therefore, one of the first steps in analyzing EEG signals in context of BCI and mental state monitoring applications is to clean up the contaminated EEG signal by applying filtering, de-noising and artifact reduction methods categorized under preprocessing. In Figure 2.3 a normal EEG signal (a) and some common noise and artifactual distortions are depicted (McFarland et al., 2006).

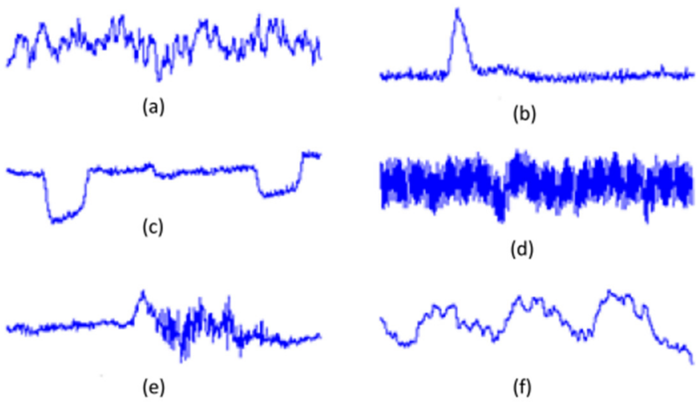
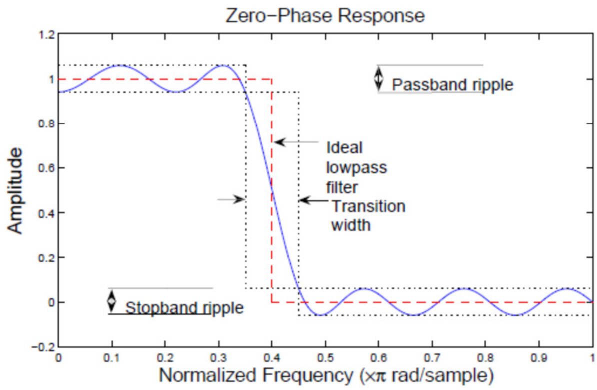


Figure 2.3: A typical demonstration of (a) normal EEG signals, (b) EEG affected by eye-blinking artifact, (c) EEG affected by eye movement (d) EEG affected by the 50Hz power line noise, (e) EEG affected by EMG, (f) EEG affected by ECG noise.

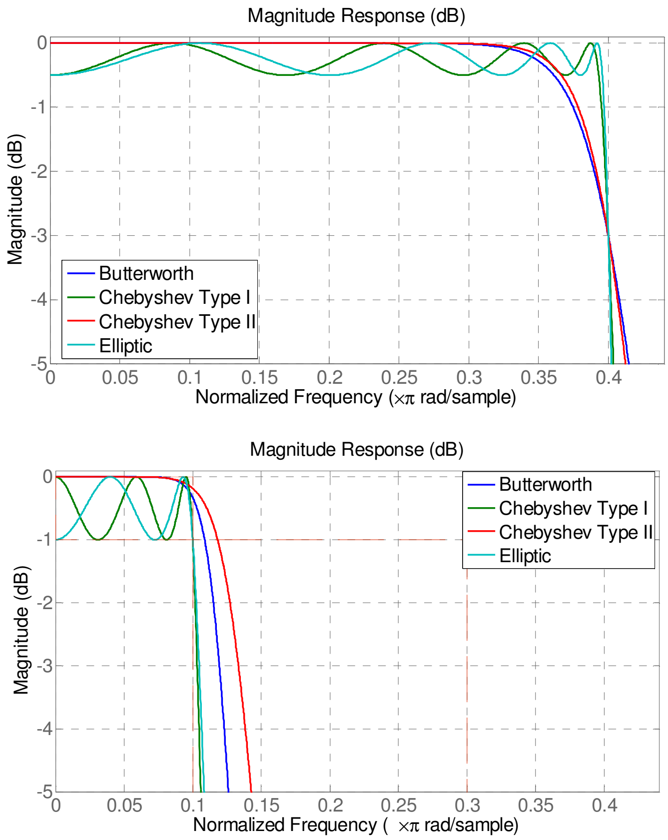
Generally artifacts are divided into the two groups of physiologic and extra-physiologic artifacts (Sethi et al., 2007). Physiologic artifacts arise from sources other than the brain but they are sill generated from the human body. They may be caused by eye blinks, vertical or horizontal eye movements, and muscle activity of different part of the body and in particular of vicinity of the head (e.g. tongue, jaws, face muscles) (Goncharova et al., 2003), respiration, rhythmic cardiac activity, and transpiration. Extra-physiologic artifacts originate from sources outside the body such as equipment and environment (Sethi et al., 2007) (e.g., power-line noise (50/60 Hz), changes in electrode impedances, line humming or equipment interferences) (Barlow, 1986) (Ramoser et al., 2000).

2.3.1.1. Filtering

In those lucky cases, when noise originates from sources with distinct, but recognizable frequency contents, it may be suppressed by appropriately filtered attenuation. The desired signal thus becomes more prominent above the noise level. From the signal processing point of view, a filter is a mathematical operation applied to the original signal in order to extract the desired (selected) information from it (Diniz, 2008). Notch filters are common to reduce power line noise (50 Hz or 60 Hz). We used both Butterworth and Chebyshev filters, as classical IIR filter designs, in our BCI systems. Butterworth has maximum flatness in the pass-band and thus it restricts the other parameters in designing the filter. Chebyshev filters are optimum in the sense that their error between the ideal and the actual filter characteristics is minimized over the range of the filter; it has steeper roll-off at the price of more ripples. Elliptic filter design allows for ripples in both the pass-band and the stop-band, therefore has maximum degrees of freedom. The degree of freedom is higher in Chebyshev filters, since the ripples might happen in the pass-band (for Chebyshev type I) or the stop-band (for Chebyshev type II). Indeed, both Butterworth and Chebyshev filters are special cases of the elliptic filter. For the purpose of comparison with other linear filters. Figure 2.4 (b) shows the general frequency responses of the four common filters using the same filter order (number of coefficients).



(a)



(b)

Figure 2.4: a) Ideal low-pass filter (red hashed line) and a typical filter and its specifications, b) General frequency responses of four common filters. Filter type determines the general form of the frequency response. Other parameters like order of the filter, cut-off frequency, pass band and stop band ripples also affect the frequency response.

In EEG signal processing high-pass filter is usually used to remove drift due to sweating and low-pass filtering is the common approach for eliminating the interference from the power line. In all of our BCI application we applied band-pass Butterworth and Chebyshev filtering between a low frequency value (under 0.5) and a high frequency value less than 45 Hz. These filters don't have perfectly linear phase, but they are robust and have relatively small group delay.

2.3.1.2. EMG artifact filtering

In contrast to the non-physiological artifacts, physiological ones such as Electrooculogram (EOG) and Electromyogram (EMG) have generally undetermined shapes and therefore, they are far more difficult to deal with. Moreover, controlling them during signal acquisition is not easy (Goncharova et al., 2003). The two physiological artifacts that have been examined most are ocular (EOG) and muscle (EMG) artifacts (Ramoser et al., 2000) (McFarland et al., 2006). Fortunately, the latter type of artifacts can usually be minimized by proper filtering, shielding, etc. Any body movement including head, jaw or even tongue movements can cause EMG disturbances in the brain signal. EMG artifacts have a broad frequency range, with maximum frequency higher than 30 Hz (Anderer et al., 1999). In a BCI system a direct relationship exists between the task difficulty and the amount of EMG artifacts produced by movement of facial muscles (Waterink et al., 1994).

2.3.1.3. Eye blink artifact reduction

EOG artifacts are patterns with high amplitude in the theta range mainly caused by blinking or floating movements of the eyeball in closed eyes (Anderer et al., 1999), or high-frequency patterns in the gamma range of large or small amplitudes when eyes are open caused by goal-directed eye movements ("saccades") (Keren et al., 2010) (Plöchl et al., 2012). They in particular are inevitable in long-term EEG recordings and can alter the electrical field around the eyes and consequently propagate over the scalp and mislead the diagnosis in clinical applications (Woestenburg et al., 1983) (Tatum et al., 2011). Since the artifacts can mimic any kind of electro cerebral activity, they may easily deflect the functionality of systems based on EEG analysis (Waterink et al., 1994). They can also erroneously yield to an unintentional control of the device (Wolpaw et al., 2002). From this point of view, artifacts seriously impede the concept of applying EEG as a control source for BCI. In online BCI systems, artifacts can impact the performance, precision, and reliability of the system either during Intentionally control (IC) periods by wiping out the shape of the desired neurological event or during the Nonintentional control (NC) by mocking the properties of the neurological aspect (Erfanian et al., 2005) (McFarland et al., 2006). There are various methods of handling artifacts in offline analysis published, ranging from simple avoidance to automatic rejection (Ramoser et al., 2000), auto-regression (Anderer et al., 1999) and Independent Component Analysis (ICA) (Barbati et al., 2004) (Makeig et al., 1996). But dealing with real world online BCI scenarios demands more efforts. Finding a robust and fully reliable solution, especially in online and long term recordings, is still a challenging problem (McFarland et al., 2006) (Sethi et al., 2007) (Tatum et al., 2011). The targeted artifacts in signals have to be removed completely and the related neurological phenomenon should not be distorted by the artifact-removal algorithm (Fitzgibbon et al., 2007)

2.3.1.3.1. Conventional methods of dealing with ocular artifacts

Artifact avoidance is the primary method in that during the experiments users have to avoid moving their bodies or blinking (Anderer et al., 1999). It has obviously the advantage of creating the minimum number of artifacts during data collection and consequently demanding the least computational load among all the artifact handling methods. However, it has two drawbacks. First, because of the long period of experiments in an online BCI system, it is usually very difficult to collect the sufficient amount of data without occurrence of any artifacts. Second, it is possible that extra attention to avoid artifacts causes an additional cognitive task for the individual. For instance, there are studies that show the amplitude of some evoked potentials are affected when the subjects intentionally refrained from eye blinking (Verleger, 1991) (Ochoa & Polich, 2000).

Manual artifact rejection could be considered as the simplest way of handling brain signals contaminated with artifacts. It refers to the process of rejecting that part of signals which is affected by artifacts. Compared to the previous method, subjects can take part in the experiments and perform the required tasks with more convenience. As the artifacts have distinguishable patterns carried in the EEG signal, thus, they can detected visually based on their topographic and spectral properties by experts (Barlow, 1986). Manual rejection of artifact-contaminated epochs has been very popular for offline analysis in EEG signal processing and also in the BCI field (Vaughan et al., 2003). Each normal person usually blinks around 20 times per minute and each blink has lasts between 300 and 500 milliseconds (Stern et al., 1984) (Iwasaki et al., 2005). This method is not applicable in online BCI recording at all and comes at the cost expense of exhaustive human labor and the selection of contaminated trials can be in some extent subjective.

Using an automatic rejection algorithm, the BCI system automatically discards the EEG trials contaminated with predefined types of artifact. For instance, by setting a predetermined threshold, the system considers the epoch as the artifact-contaminated one, when its amplitude exceeds the threshold and automatically rejects the epoch (Barlow, 1986). Since the rejection algorithms are usually based on a simple criterion, they can recognize and exclude the epochs only with a strong presence of artifacts. Consequently the artifacts may still sustain in the so-called clean data. This fact poses a huge drawback for online applications of a BCI system (Ramoser et al., 2000) (Romero et al., 2008).

The second disadvantage of the artifact rejection is the extensive data loss it causes. Extracting and removing artifacteous signals puts quite a computational load on the system and in addition because of eliminating the contaminated signals will render it unusable for real-time and online BCI control (Romero et al., 2008).

In order to overcome the problem, artifact correction has been introduced (Berg et al., 1994). This method is applicable in online BCI recording and can also be determined due to extracted information either from EOG or directly from EEG channels.

In most reported BCI systems in literature the employed artifact reduction techniques are not reported or are limited to calculating a predefined threshold (Fatourechi et al., 2007). However, there have been suggestions for online correction of contaminated data in order to not ignore a considerable part of recorded data during the experiment due to artifacts (Ramoser et al., 2000) (Erfanian et al., 2005). Some studies are focused on automatic artifact reduction by exploring the effect of feature extraction techniques (Winkler et al., 2011), blind source separation methods (Fitzgibbon et al., 2007), regression methods (Ng et al., 2008) and adaptive techniques (Romero et al., 2008). In this dissertation two different methods will be applied. In the first section I will propose a method based on independent component analysis (ICA) and higher order statistic and evaluate its performance in chapters 3 and 5. In the second approach I will introduce and evaluate an information-based method for online recognition and removal of ocular artifacts and apply them for BCI application in chapters 4 and 6.

2.3.1.3.2. Automatic Artifact correction with ICA

Approaches based on independent component analysis (ICA) have proved to be useful tools for artifact reduction in EEG signals (Makeig et al., 1996) (Delorme et al., 2001) (LeVan et al., 2006) (Winkler et al., 2011). Makeig introduced the applicability of independent component analysis for EEG signals with the assumption that artifact sources are independent from EEG sources (Makeig et al., 1996). ICA is a statistical technique for transforming an observed multidimensional signal into its components (Makeig et al., 1996), which can be categorized under blind source separation (BSS) group of algorithms. ICA tried to decompose a multivariate signal into all its additive subcomponents, which are as much as possible statistically independent from each other (Makeig et al., 1996). The "Cocktail party problem" is a classic example of ICA (Haykin, et al., 2007), where data is recorded form microphones mounted inside a room and components are speech signals of people talking simultaneously in the room. In the typical ICA scheme for artifact removal the original signal is decomposed to its Independent Components (IC), among these components those which better represent the dynamics of artifacts are visually or automatically distinguished and rejected (Makeig et al., 1996). Finally the remaining ICs are remixed to form an artifact free signal (Ng et al., 2008) or used for classification purposes (Graimann et al, 2010) (Winkler et al., 2011).

Assume $x = (x_1, x_2, \dots, x_m)^T$ is an observed zero-mean m-dimensional random variable and y is the result of its projection by $f(.)$: $y = f([W.x])$. Among all possible functions we consider a linear transform. Assume $u = (u_1, u_2, \dots, u_n)^T$ is an n-dimensional matrix including the independent components of x.

$$x = A.u = \sum_{k=1}^{n} u_k \, . \, a_k \tag{2.1}$$

The mixing matrix A includes basic vectors $\mathrm{a_k}$. Mathematical calculation is done by adaptively estimating the optimum ω_k vectors and sources simultaneously, where $W = A^{-1}$.

Onton and Makeig (2006) described a method based on ICA to minimize eye blinking artifact (Onton et al., 2006). They suggested that the spatial localization of the components is important for recognizing the component containing EOG artifacts. By observing each component they selected the IC which includes the strongest blinking artifacts. This component is then set to zero before inverting the data to the EEG domain. Figure 2.5 describes the process. They also reported that the EOG artifact in lower frequencies (0–2 Hz) may not be corrected completely after this process and significant statistically differences can be observed in signals before and after artifact reduction. The cost function is defined either based on maximizing the nongaussianity of s components or based on minimizing their mutual information (Sanei, 2007). Mutual information is minimized when the entropy of the elements of u are maximized making them as independent as possible (Graimann et al, 2010).

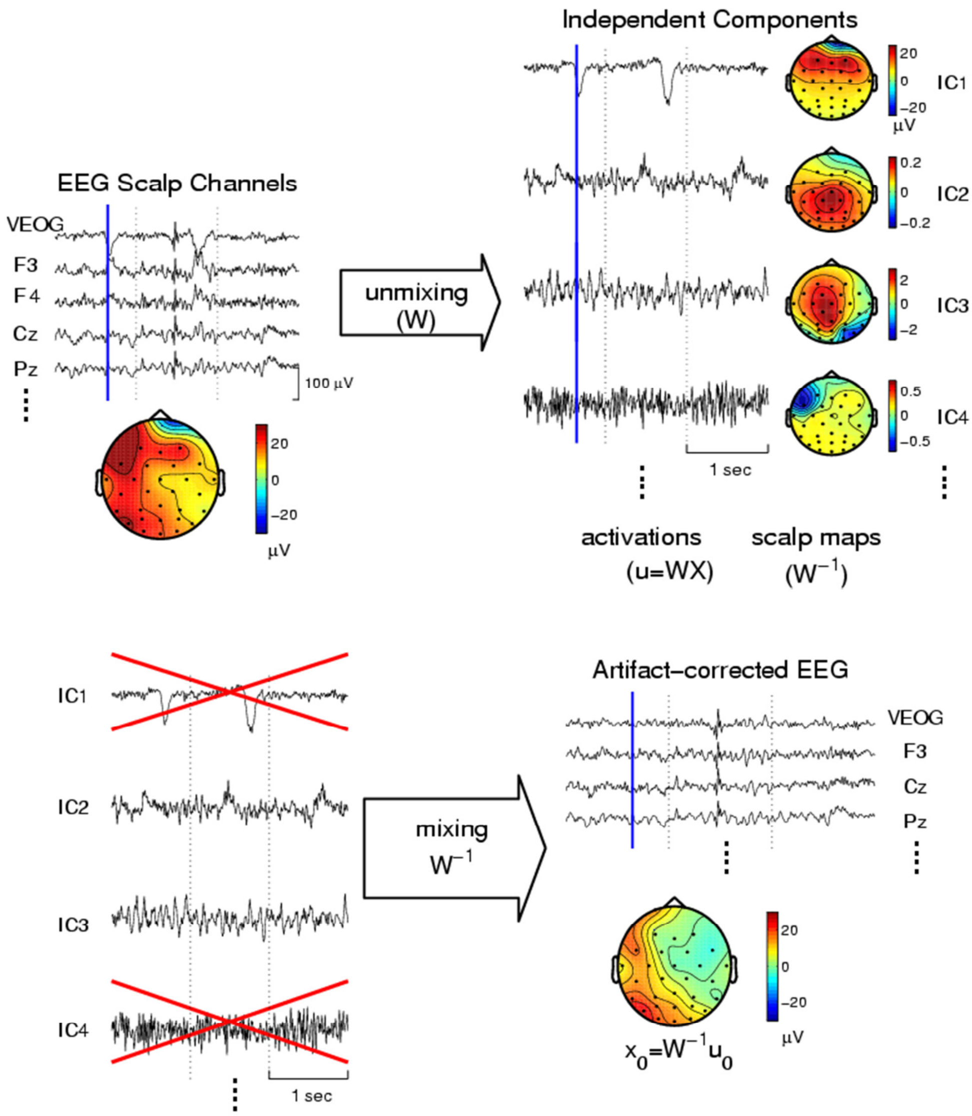


Figure 2.5: ICA decomposition: The procedure of artifact reduction using independent component analysis. First the multichannel EEG signal is transformed to components which are statistically as independent as possible. The component that shows artifact waveform is set to zero. The components are then mixed again to build artifact-free signals (Reprinted from (Makeig et al., 1996)).

The sources are then recovered by $u_k = \omega^T * x$. In the over complete case that $n > m$ there is still a solution which can be calculated by pseudo inverse techniques (Makeig et al., 1996). Blind source separation by Independent Component Analysis is thought to give good results on demixing signals, as long as the independence assumption holds true. Principally ICA performance depends on the length of the data (Makeig et al., 1996). ICA has inherently some disadvantages. Despite all attempts on this realm, in most cases the algorithm is not suitable for a robust online and adaptive scheme. The method cannot identify the actual number of source signals, nor a uniquely correct ordering of the source signals, nor the proper scaling or sign of the source signals. As an alternative we implemented an adaptive information filter and its performance was compared to a LMS adaptive filter and the ICA-based approach.

2.3.1.3.3. Automatic Artifact correction using Adaptive filter

Originally suggested by Wiener an adaptive filter optimally recovers the original message from a signal contaminated by noise for control and prediction applications (Haykin, 1996). An adaptive filter (AF) is a supervised learning system which requires two inputs, the desired signal and the reference input (Haykin, 1996). An adaptive noise cancellation has a very similar structure. Based on a reference signal, it adaptively eliminates the noisy part of the input signal, which is assumed to be highly correlated with the noise signal (Diniz, 2008). Based on adaptive filtering theory, conventional adaptive noise cancelling systems exploit a criterion called mean square error (MSE) in order to adapt the filter weights. Remove the noise component is achieved when the weights converge to an optimum solution in the available data space (Haykin, 1996). Figure 2.6 demonstrates the structure of a typical noise canceller. The error signal is the difference between the outputs and the desired signal (Diniz, 2008).

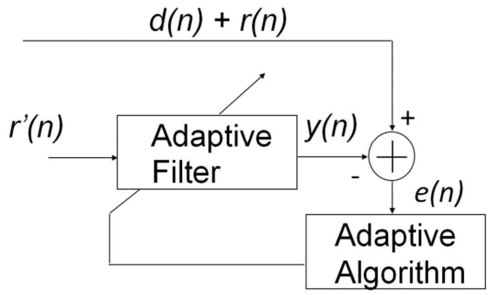


Figure 2.6: Adaptive filter with the application of noise cancellation is a well-known method in communication to reduce noise from the signal.

For many practical applications, the efficiency of MSE as the optimum criteria can be argued (Principe, 2010), especially, where the target signal is non-stationary and non-Gaussian signal such as EEG (Arndt, 2001). Indeed, the MSE minimization is just optimum for Gaussian distributed errors, since it only can explore the second-order moment of the error distribution. In cases where the assumption of Gaussianity is not valid for the error distribution, study of alternate cost functions for adaptation makes complete sense (Principe, 2010). In this dissertation we consider an approach based on information-theoretical concepts, and consider the error entropy criterion (EEC). Error entropy is an analytical expression that involves the probability density function (PDF) of the error. The function is able to capture all the underlying dynamics of the adaptive system which we are coping with. In order to eliminate as much artifact information as possible from the EEG signal, we also consider an adaptive scheme for this problem (LeVan et al., 2006). In chapter 4 we will describe the implementation of an entropy-based adaptive noise canceller for eye blink artifact reduction and present the results obtained while testing the filter on pre-recorded data in a simulated online paradigm. The later method is up to 5 times faster than an ICA-based approach. Since the technique used is based on information theoretic learning, we borrowed the metric correntropy from information theory (Weifeng et al., 2006) to measure the performance of our proposed system. Correntropy is defined based on kernel methods and information theoretic learning and is a measure for the localized similarity.

2.3.2. Feature extraction

Selecting the appropriate features and classifiers affects recognition rate and the efficiency of BCI systems (Lotte et al., 2007). This section aims to cover the major ideas in this important era in modern science and nails it down in more advanced issues of feature extraction and selection in BCI. In order to map the brain activity to motor behavior, the main key is to extract the appropriate spatiotemporal and spectral features. In the following we present a comprehensive overview of each method employed in our work. Literally each brain-computer interaction is a unique problem to be solved. Here our attempt is to extract an appropriate feature set and classification approach to solve this marvel. We exploited several feature extraction schemes for the current BCI project. Fractal components, higher order statistics, and time-frequency analysis have been applied.

Features describe the relevant information embedded in the EEG signals. The problem of unknown parameters or possible features that could be extracted from the brain signal to show the exact brain state is one of the main difficulties in design of BCI systems. Despite the large number of features that were sought previously in the field of BCI design (Koprinska, 2009), current EEG-based BCI systems are still not in at a satisfactory level. Thus the BCI community has invited research groups to explore quantification methods and to investigate alternative and meaningful features to improve the performance and reliability of BCIs (McFarland et al., 2006). Due to the nature of brain physiology, each person has a unique brain mapping affected by his/her genetics, environment and training during life. For example, folding of cortex differs between any two persons (even identical twins) (Sanei, 2007).

In addition relevant functional mapping differs in each individual across different experimental sessions and even sensor locations differ across recording sessions. The fact that the features extracted from brain signals are person-specific, task-specific, and hardware-specific (Lotte et al., 2007), makes the finding of a unique solution of best features for BCI rather complicated and cumbersome. Hitherto several different feature groups have been exploited in the context of BCI. According to the type and strategy of BCI design different feature sets are extracted and employed (Boostani, et al., 2004) (Koprinska, 2009) (Corralejo, et al., 2011). The method for constructing the feature space is highly tentative (Schouenborg, 2011). The final goal is to improve the classification results (Alpaydin E., 2004). Table 2.1 summarizes popular feature extraction methods proposed by different research groups for each specific application.

Table 2.1: A summary of feature sets applied in BCI research

Research group	Strategy	Features	Application	Reference
Berlin BCI	Imagination of movement, P300	Time-frequency	Gaming, Car control	(Blankertz et al., 2002) (Dornhege et al., 2007) (Haufe et al., 2011)
Graz BCI	Imagination of movement	Power band, AAR	Prosthesis, wheelchair control	(Pfurtscheller et al., 2001) (Vidaurre, et al., 2005) (Graimann et al, 2010)
Wadsworth	SSVEP, Cognitive	Power band	Cursor control	(Schalk et al., 2007) (Allison et al., 2008)
Tübingen	SCP, P300	Time-frequency	Virtual keyboard, TTD	(Birbaumer et al., 2000) (Hinterberger et al., 2003)
Freiburg BCCN	Limb movement ECoG	Time-frequency	Arm and finger movement prediction	(Mehring, et al., 2003) (Ball, et al., 2004)
British Columbia	Self-paced	Time-frequency	Cursor control	(Mason, 2006)
Warsaw	Cognitive	Wavelet, HOS	Control	(Daly I., 2012)
Amsterdam	Emotions	Time-frequency	Control	(Molina, et al., 2009) (Curran E et al., 2004)

Here I categorize each feature set which I suggested in my PhD study, based on the mathematical background and concepts and in each following chapter I will combine these sets of features with different classifiers to evaluate the entire system in an online scheme. It is worth adding that some of these feature sets have been used for offline studies by other groups but the innovation in my work is to implement fast algorithms in order to extract them from ongoing brain signals in an online BCI.

The first group consists of features based on power bands (Townsend, et al., 2006) and spatio-temporal characteristics of the EEG signal (Ramoser et al., 2000). The next groups of features introduced in this dissertation are statistical moments based on higher order statistics (Johansen, et al., 2000) (Kołodziej, et al., 2011). Fractal components analysis (Pereda, et al., 1998) (Boostani, et al., 2004) and matching pursuit (Mallat, et al., 1993) will also be investigated, which hold rather physical interpretations.

2.3.2.1. Feature extraction based on joint time-frequency analysis

EEG waveforms can be categorized according to their amplitude, frequency and shape. Several features based on time, frequency or space domains have been examined in BCI research (McFarland et al., 2006) (Koprinska, 2009) (Corralejo, et al., 2011). Recently some review papers presented a comprehensive description and detailed mathematical formulation for each method applied in the corresponding BCI literature (Huan, et al., 2008) (Koprinska, 2009). These methods usually consider one or two dimensions of the whole space. The common feature groups in time and frequency domains can be listed as: Amplitude values of EEG, band power (Vidaurre, et al., 2005), autoregressive and adaptive autoregressive (AAR) parameters estimated with recursive least square (RLS) (Dornhege et al., 2007) (Sanei, 2007), inverse model-based features and energy density maps (McFarland et al., 2006).

The spatio-temporal oscillations in EEG waves are indicative of brain functional activity (Society of Neuroscience, 2008). Both rhythmical and transient features exist in the EEG signal (Sanei, 2007). Time-frequency features such as power spectral density (PSD) values show promising results in BCI applications (Pfurtscheller, et al., 2000). Furthermore, the topography of the electrodes on the scalp from where the data is recorded contains spatial information (Binnie et al., 1982). Most atomic decompositions applied on the EEG signals consider only two out of the three inherent dimensions available in the data, e.g. space–time decompositions by principal component analysis (PCA)(Lagerlund, et al., 1997), independent component analysis (ICA) (Koprinska, 2009), and spatial filters (Townsend, et al., 2006), time–frequency analysis with the use of windowed Fourier transform (Nicolas-Alonso, et al., 2012), wavelet transformation (Kołodziej, et al., 2011) and matching pursuit algorithm, just to name a few (Mallat, et al., 1993). Recently, new attempts at finding a multidimensional (space–time–frequency) atomic decomposition of the EEG have been made (Cheron, et al., 2007) in the way of having a complete description of the electrical activity of the underlying neural masses (Qian, et al., 1996). In any time-frequency signal analysis, estimating the energy density suffers from a trade-off between the time and frequency resolution, which is also known as uncertainty principle (NaitAli, 2009). A promising alternative for time-frequency analysis of EEG signal is matching pursuit (MP). In the MP algorithm, in contrast to conventional approaches such as the short time Fourier transform (STFT) (Cheron, et al., 2007) or wavelet transform. (Kołodziej, et al., 2011), local features are sought out of a dictionary to be locally adapted a time-frequency trade-off to the structure of analyzed signal. MP can provide better time-frequency resolution thanks to its local adaptation to transient patterns in the signal (Sanei, 2007).

2.3.2.2. Feature extraction based on higher order statistics

The conventional techniques of statistics rely on the first and second moments of the sample, e.g. arithmetic mean and variance (Fukunaga, 1990). The idea of applying higher order statistics to analyze EEG signal was originally proposed by Kim and Power (1979). The impetus behind this is that the traditional second-order measures are not sufficient to model most of the natural processes such as biomedical signals. For instance the autocorrelation function and its Fourier Transform (FT), known as power spectrum (Wiener-Khintchine theorem), are not able to capture all information hidden in the signal. Unlike power spectrum, higher order spectra contain information about random processes.

The brain activity in different biological and physiological situations may combine the local sources in a different scheme and cause the entire black box to act as a nonlinear system. The EEG signal is a result or unnatural output of this mixing system. Although the assumptions of stationarity, Gaussianity and linearity have been applied successfully on some practical problems regarding EEG signals (Allison, 2012), it may lead to better or more accurate estimations when nonlinearity concerns or higher order statistics are considered in the solution (Greb, et al., 1988) (Johansen, et al., 2000). EEGs are quasistationarity signals. It means, they can be considered stationary only within short periods of the data. In addition, the Gaussian assumption is valid only during the normal brain states and it may not be valid anymore when the person has intensive mental and physical activities (Sanei, 2007).

Greb and Rusbridge presented a detailed report on the non-linear interactions between different modes of oscillation using bispectrum and bicoherence (Greb, et al., 1988) which can be applied on visually evoked potentials (Johansen, et al., 2000) or used as features for BCI (Kołodziej, et al., 2011). Bispectrum analysis was applied on focal ischemic cerebral

EEG signal using third-order recursion method (Zhang, et al., 2000). In chapter 3 we will propose and evaluate a set of features based on these definitions.

2.3.2.3. Feature extraction based on fractal components

In the absence of the external stimulation, brain activity is typically unpredictable and aperiodic. However, the question of whether or not the brain activity is truly chaotic cannot be answered based on a mathematically stringent proof (Krasner, 1990). Yet, it eminently has been claimed that EEG signals are examples of fractal geometry (Boostani, et al., 2004) (Phothisonothai, et al., 2007). If a variable as a function of time undergoes characteristic changes that are similar regardless of the time interval over which the observations are made the underlying process is defined as a fractal (Mandelbrot, et al., 1968). The existence of natural self-similarity in different scales of systems is referred as chaotic. Many definitions including geometric and dynamical measures have been developed to describe fractal systems, e.g., Mandelbrot and van Ness set (Mandelbrot, et al., 1968), Hausdorf dimension, Higuchi's fractal dimension (Higuchi, 1988), Hurst exponent, largest Lyapunov exponent (LLD), and correlation dimension (CD) (Mandelbrot, 1982). The fractal dimension is a popular and informative parameter to characterize a chaotic system. Finding the fractal dimension from data observations is usually done by analyzing of time series (Elbert et al., 1994) (Boostani, et al., 2004). Fractal dimension is a common mean to estimate the scale independent complexity and also irregularity in the biological system over either space or time (Phothisonothai, et al., 2007) (Georgiev, et al., 2009). Esteller et al. reported a comprehensive comparison amongst several well-known approaches using both intracranial EEG and synthetic data (Esteller, et al., 2001).

Chaos theory or chaotic computation is an advanced and effective method especially to describe and predict the complex self-organizing systems and natural phenomena as encountered in astronomy, weather forecast and biomedical signal processing. There are some measures defined based on chaos theory which can represent the nonlinear behavior of EEG signals. One of the most important dynamical invariants is the Lyapunov exponent (Sanei, 2007). Multifractal cumulants and predictive complexity of the EEG time series are introduced and compared to the power band features as the most common feature set in for EEG based BCIs (Brodu, et al., 2012). The first new feature, multifractal cumulants measures the signal regularity and is a statistical measure for inter-frequency band relations. On the other hand, predictive complexity determines how complex the signal is by measuring the difficulty in predicting the future of the signal knowing its past (Brodu, et al., 2012). A new method for extracting the fractal components in the frequency domain was used in chapter 4 as a feature set for BCI.

2.3.3. Feature selection techniques

The goal of this element in BCI design is to find a set of features or attributes out of the EEG signal which contain the most useful information about the brain state. The combination of different feature sets with different classifiers may lead to better results in BCIs (Lotte et al., 2007). In the following we will present relevant feature selection tools. Applying the feature selection method significantly shrinks the feature space and may improve the classification results.

Any classification method uses a set of attributes or features relevant to the task to categorize the input data. The dimension of this feature set can rapidly increase depending on the number of attributes extracted from the data (Dash, 1997). It is recommended that the number of training samples per class exceeds the dimension of the dataset by ten (Alpaydin E., 2004) (Zhang, 2010). This is however not applicable in real-time problems. The "curse of dimensionality" is a serious problem in machine learning (Cortes, et al., 1995). Thus, feature selection plays a critical role in pattern recognition problems (Dash, 1997). In BCI applications the feature space grows very fast by adding the recorded data from multichannel recording. Feature selection or variable subset selection is a technique to determine the optimum subset of features for constructing the most robust and efficient learning model. The approach is more effective when the number of features or attributes is much higher than the number of data samples. If the dataset is large, it is reasonable to apply the feature selection algorithm and remove the irrelevant and redundant features from the data. It not only helps to mitigate the effect of the curse of dimensionality and to speed up the learning, but also it may boost the generalization capability and improve the performance of learning models.

Several methods have been proposed to estimate the usefulness of a feature set for predicting the target variable in data mining and machine learning (Dash, 1997) (Huan, et al., 2008). Among them are correlation-based feature selection (CFS) (Huan, et al., 2008), Relief (Huan, et al., 2008), information gain ranking (IG) (Dash, 1997), genetic algorithm (Corralejo, et al., 2011) and Maximum relevance minimum redundancy (mRMR) (Peng et al., 2005) all of which have been applied successfully on different feature spaces extracted from EEG signals (Koprinska, 2009). There are two general schemes for the feature selection. In the *Filter* evaluation method, features or feature subsets are evaluated independently of the classification algorithm. In the *Wrapper* evaluation technique, feature subsets are evaluated based on a particular target classification algorithm (Koprinska, 2009). So the filter and wrapper schemes differs exclusively their relation with the classification algorithm. To identify the most prominent features, the state-of–the-art method, mRMR, will be investigated over our datasets. By altering the number of features, we selected several feature subsets. The performance of each subset was then evaluated. We applied the introduced feature selection method to reduce the feature space and meanwhile to improve the classification results. Figure 2.7 shows an interesting example of how irrelevant features may lead to misclassification and why feature selection is critical to minimize the classification error. Shapes with a small circle inside belong to one class and the ones without a circle belong to another class. Existence of other irrelevant or redundant information can cause misinterpretation by the classifier (Koprinska, 2009) (Peng et al., 2005).

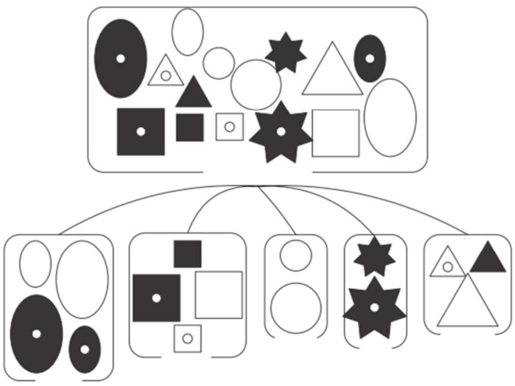


Figure 2.7: An example of impurity based feature evaluation (Adopted from (Huan, et al., 2008))

2.3.4. Classification

The techniques related to the feature space and their properties as they pertain to BCI application were covered in the previous section for a subset of potential feature extraction methods. In the current section, we draw our attention to the next important step in the machine learning problem. The performance of machine learning systems depends on both the feature extraction and the classification algorithm employed (Lotte et al., 2007).

Many algorithms of varying complexity have been reported for the purpose of classification in BCI literature (Sanei, 2007). Lotte et al. present a comprehensive overview on theoretical aspects of classifiers employed in the BCI field until 2006 (Lotte et al., 2007). Based on the No-Free-Lunch theorem there is always a tradeoff between bias and variance, hence the generalization of a classifier cannot be determined without considering the entire feature space. It means, in practice, there is no such best classifier and the relative better classification accuracy is calculated only on a predefined set of data, which can be examined experimentally (Soria Frisch, 2012) (Zhang, 2010). In our BCI applications we considered stable classifiers which in theory have a high bias and low variance (Lotte et al., 2007).

2.3.4.1. Bias-variance tradeoff

The major impetus behind the classification section of each BCI is to estimate a certain aspect of cognitive state from the brain signal (McFarland et al., 2006). In order to recognize and classify the different cognitive states, it is important to define each state carefully and determine if it is compatible with the task during the EEG measurements. This can pave the way for finding the best solution in a certain situation. The question is what method works best and what accuracies can be achieved. Let's assume that a mathematical model with function $y(x,t)$ maps an observation x in D dimensional feature space to a discrete label $\in \{1,\cdots,C\}$, where C represents the label of classes. The values for parameters w of the model are estimated using N observations along with their respective label $\{x_i,t_i\}, i = 1,\ldots,N$ called the training set. The inferred model should possess predictive, descriptive and generalizing properties. If $\hat{y}$ is the output of the classifier and $f(x)$ is the label, the Mean Square Error (MSE) of classification outcomes can be decomposed in three terms (Breiman, 1998) (Friedman, 1997):

$$
\begin{aligned}
MSE &= E[(f(x)-\hat{y})^2] \qquad (2.2)\\
&= E\left[\left(\hat{y}-\hat{f}(x)+\hat{f}(x)-E[f(x)]+E[f(x)]-f(x)\right)^2\right]\\
&= E\left[(\hat{y}-\hat{f}(x))^2\right]+E\left[(\hat{f}(x)-E[f(x)])^2\right]+E[(E[f(x)]-f(x))^2]\\
&= Noise^2+Bias^2(f(x))+var(f(x))
\end{aligned}
$$

MSE can be broken down into three components that are familiar: squared noise, variance and squared bias. Such decomposition is always possible (Hastie et al., 2008) and the equation shows three possible sources of classification error:

- Noise: The first term represents the noise within the system.
- Bias: The second term is known as the divergence between the estimated and the best mapping.
- Variance: The last term reflects the sensitivity to the training set used.

In the above equation, the noise term is an irreducible error, because it arises from the inherent structure of the system. Therefore, both the Bias and the variance have to be kept low in order to minimize the overall classification error. Nevertheless, it is logically not attainable because of the natural bias-variance trade-off in each classifier structure (Hastie et al., 2008)

(Zhang, 2010). The bias component is the squared difference between the true mean and the expected value of the estimate, $E\left[(\hat{f}(x) - E[f(x)])^2\right]$, and depends on the type of the function f (linear, quadratic . . .). The expectation demonstrates the averages of randomness in the training data (Alpaydin E., 2004). The fact is that high variance classifiers tend to be unstable, while high bias and low variance classifiers are usually the stable ones (Soria Frisch, 2012). When the model complexity increases, the training error tends to decrease and it means we try harder to fit the data to the model. The error in the test set should always be comparable to the error in the training set (Principe, 2010). However, over-fitting will happen when the model adapts itself too closely to the training data. In this case, the test error gets larger and generalization decreases (Figure 2.8). Under-fitting, in contrast, happens when the model is too simple to extract the structure of the data (Friedman, 1997).

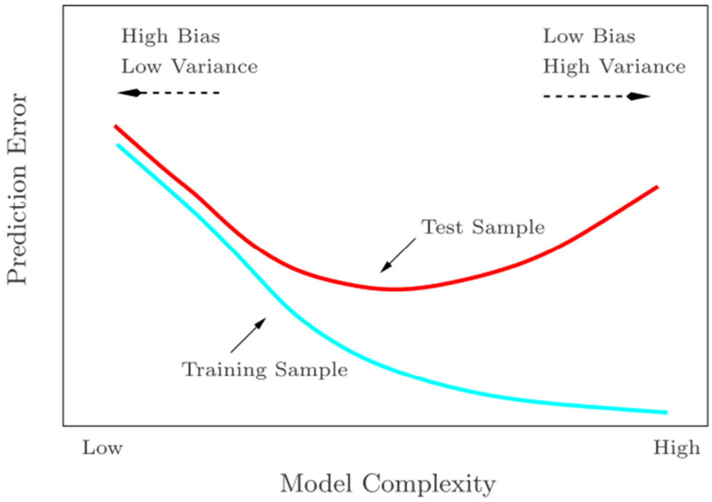


Figure 2.8: In order to minimize the classification error on the test samples the bias-variance trade-off should be considered. Models with higher complexity tend to have lower bias and higher variance (Adopted from (Hastie et al., 2008)).

For the same reason, in some cases reported in literature, simple classifiers often outperform the more complex ones for BCI application (Lotte et al., 2007). Training sets collected in different sessions are plausible to be rather different. Therefore, for BCI systems a low variance classifier can be an appropriate solution to cope with the variability of input data. Combination of classifiers (Breiman, 1998) and regularization are two techniques to improve stabilization and decrease the variance (Lotte et al., 2007). Recently, ensemble classifiers have attracted the attention of machine learning researchers and in the BCI field a chapter in a very recently published book from (Allison, 2012) introduces an extensive overview on this concept (Soria Frisch, 2012). Taking inspiration from this overview, I applied some selected methods successfully in my BCI designs. In the following, I will initially focus on some materials that are commonly applied in brain signal analysis for BCI classification purposes. The concept of the combination of classifiers known as ensemble classifier is explored in the last section of the chapter. My contribution is then described, where the ensemble classification is used to enhance the performance of the BCI system

2.3.4.2. Supervised and unsupervised learning methods

Generally there are two types of learning for classifiers: with or without supervision. Supervision is an important factor that influences learning. The existence of a tutor or supervisor enables direct feed-back about correct answers usually in each step of training to improve the performance (Steinwart et al., 2008). In supervised learning, the supervised discrimination or classification is based on past observations with specific class information known as a label. Supervised classifiers such as Fisher's linear discriminant analysis (Müller et al., 2003), regularized discriminant analysis (Blankertz et al., 2002), and support vector machines (SVMs) (Cortes, et al., 1995) have been employed widely in developing BCI systems. In the BCI structure, a supervised learning paradigm includes a cue from the program side to give the user hints. So the subject is not intentionally free to think about the desired command. A classifier that does not use any class label is considered an unsupervised learner. This type of classifier clusters the feature space based on the common attributes they share. The unsupervised classification algorithms used in BCI research are self-organizing feature maps (SOFM) (Liu, et al., 2005), and principal component analysis (PCA) (Jung T-P. et al., 1998), the hidden Markov model (HMM) (Argunsah, et al., 2010), logistic regression (Soria Frisch, 2012), and hierarchical clustering (Soria Frisch, 2012). Information theoretic principles play an important role in many of the unsupervised learning methods (Principe, 2010). Other popular classification techniques have variations in both supervised and unsupervised fashions; K-nearest-neighbors (KNN) (Manocha et al., 2007), artificial neural networks (Breiman, 1998) (Rakotomamonjy et al., 2005) and Bayesian classifiers (Pfurtscheller et al., 1993) (Lotte et al., 2007) to name a few (Fukunaga, 1990).

Real-time classification and pattern recognition is an issue in biomedical signal processing, and many research groups attempt to apply adaptive approaches. Neural networks (Bishop, 1995) and adaptive Bayesian classifiers (Tipping M., 2001) (Lawrence et al., 2001) are very common. In this dissertation, five classification algorithms, representing different learning paradigms, will be employed. These algorithms are SVM and LDA (2.3.4.3), QDA, and RFD.

2.3.4.3. Well-known classifiers: SVM and LDA

As earlier discussed, minimizing the error of training alone (also called empirical risk) does not guarantee the design of an optimum classifier. Regularization techniques were used since the year 1960 to establish a compromise between the complexity of the model and its aptitude for generalization (Cortes, et al., 1995). The theory was initially formalized by Vladimir Vapnik and Cortes in the first introduction of the support vector machine (SVM) idea in 1995. In the 1990s, many researchers greatly developed the idea (Friedman, 1997) (Vapnik, 1999) (Chang et al., 2001). SVM is a mathematical technique with attractive features and seems to have promising empirical performance in a wide variety of applications (Sanei, 2007) (Liang, et al., 2011). Its success due to its delicate mathematical foundation which is based on four factors: feature space, separating hyper-plane, kernel function, and optimization problem (Liang, et al., 2011). Support Vector Machines (SVMs) are generalizations of linear decision boundaries. When the data is not linearly separable or overlaps, linear separating hyperplanes cannot provide an optimal solution for the problem. The optimal decision function is the one that minimizes the test error (Lutz, 2009). SVM constructs nonlinear boundaries by making linear ones in an extended transformed feature space (Steinwart et al., 2008). The main drawback, however, is finding the optimum parameters which should be discovered by trial and error (Rakotomamonjy, 2003).

The literature available for choosing the kernel function is still in the primary level and is based on trial and error for most practical works (Hastie et al., 2008). Kernel-based techniques represent an efficient and attractive field in machine learning (Alpaydin E., 2004). Other kernel–based classifiers such as kernel fisher discriminant (KFD) were also suggested to improve the classification accuracy in many applications (Mika, et al., 2001). A nonlinear version of support vector machines was used successfully for classification of the EEG data for a BCI speller program (Krusienski, et al., 2006).

SVM in general is a very powerful method for classification. It is appropriate for practical purposes, where high transfer rates are required along with least amount of data (Meinicke et al., 2003). Its strong mathematical background, good generalization performance and the existence of unique solutions triggered researchers to use SVM in many applications (Mika, et al., 2001) (Liang, et al., 2011). It has been applied successfully to EEG signals for the detection of epileptic seizures (Gonzalez, et al., 2003), and the detection of evoked potentials (EPs) (Sanei, 2007), the removal of the eye-blinking artifact (Shoker, et al., 2005), and classification of left and right finger movements in BCI (Shoker, et al., 2005) (Lotte et al., 2007). SVM performs in most applications with satisfactory results, but still suffers from some practical drawbacks. To cope with the real-time application, we should consider the training time to find the optimum parameters (Steinwart et al., 2008) (Lutz, 2009).

From the mathematical point of view the aim of each classification algorithm is to define a function for the decision boundary between two or more classes. In a multidimensional feature space the decision boundary is indeed a separating hyper-plane (Hastie et al., 2008). The main purpose of classification is finding the optimum hyper-plane with maximum distance from all classes (Sanei, 2007). When these decision boundaries are linear; methods for classification are linear. Linear classifiers were initially developed before the advent of the computer era. But because of their simplicity and performance they are still employed extensively in classification and prediction problems, especially when the amount of training data is limited or the input signal is very noisy (Hastie et al., 2008) (Zhang, 2010). Linear discriminant analysis (LDA) is a classic classifier, as its name suggests, with a linear decision surface (Hastie et al., 2008). In 1990, Fukunaga presented a mathematical solution for LDA problems (Fukunaga, 1990).

LDA is based on the assumption that the classes have a common covariance matrix. It is likely in practice, though, that this assumption does not hold (Zhang, 2010). Decision boundaries in LDA can be easily extended to quadratic boundaries in QDA. By including the square values and cross products of the variable set $X_1^2, X_2^2, \ldots, X_1 X_2, \ldots$ linear functions in the augmented space can be mapped to quadratic ones in the original space. Pattern recognition theory states that Fisher's discriminant is able to classify the data with minimum probability of misclassification, when the data has known normal distributions and a Gaussian covariance matrix (Alpaydin E., 2004). This assumption is usually made, when LDA is used for classification of brain signals (Müller et al., 2003).

Both LDA and QDA classifiers are parameter-free and easy to compute because of their closed form solutions. This feature along with their inherently multi-class structure makes them attractive and practical for plenty of machine learning problems (Blankertz et al., 2002) (Zhang, 2010).

In both methods the class conditional distribution of the data, $P(X|y = k)$, have to be modeled for each class k. The predictions can be calculated according to Bayesian rule.

$$P(y = k|X) = P(X|y = k).P_k(y)/P(X) = P(X|y).P(Y) \Big/ (\sum_{y'} P(X|y').p(y')) \qquad (2.3)$$

Where $P_k(y)$ is the prior probability of class k that satisfies $\sum_{y'} p(y')=1$. Usually in discriminant analysis, either the linear or the quadratic case, a Gaussian distribution model is selected to model $P(X|y)$. In the LDA algorithm, we assume that the Gaussian distributions for each class have the same covariance matrix. In this case, a linear decision surface suffices for discrimination and the log-probability ratios are constant (Hastie et al., 2008). Nonlinear decision boundaries can be achieved by using mixtures of Gaussians (Hastie et al., 2008) if we assume that each class density is a multivariate Gaussian function:

$$P(X = x|y = k) = \frac{1}{(2\pi)^{p/2}|\Sigma_k|^{p/2}} e^{-\frac{1}{2}(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k)} \tag{2.4}$$

In order to derive the mathematical criteria, a log-ratio between two classes (k=1, 2) conditional distribution is calculated:

$$\log\frac{P(y = 1|X = x)}{P(y = 2|X = x)} = log\frac{P(X|y = 1)}{P(X|y = 2)} + log\frac{P_1(y)}{P_2(y)} \tag{2.5}$$

$$= -\frac{1}{2}(\mu_1 + \mu_2)^T\Sigma^{-1}(\mu_1 - \mu_2) + x^T\Sigma^{-1}(\mu_1 - \mu_2) + log\frac{P_1(y)}{P_2(y)}$$

In the case of LDA the normalization factors, as well as the quadratic part in the exponents is canceled out by covariance matrices being equal (Mika, et al., 2001), which implies that the decision boundary between two classes is a linear function of *x*. In practice the parameters of Gaussian distributions should be estimated from the training data (Hastie et al., 2008): N_k is the number of observations from class *k* with the mean of $\hat{\mu}_k$.

$$\hat{\Sigma} = \sum_{k=1}^{K}\sum_{i=k}(x_i - \hat{\mu}_k)(x_i - \hat{\mu}_k)^T/(N - K) \tag{2.6}$$

$$P_k(y) = N_k/N\ , \quad \hat{\mu}_k = \textstyle\sum_{i=k} x_i/N_k$$

Sample *x* belongs to class *k* when its linear discriminant functions have a greater value compared to the other class.

$$\xrightarrow{yields} \delta_k(x) = -\frac{1}{2}\mu_k{}^T\Sigma^{-1}\mu_k + x^T\Sigma^{-1}\mu_k + logP_k(y) \tag{2.7}$$

Classification is fulfilled according to the largest probability of occurrence of "sample *x*" among all classes, $k \in G$ (Hastie et al., 2008).

$$G(x) = \ argmax_{k\in G}\ \delta_k(x) \tag{2.8}$$

There is a subtle note that might come up here: How can we describe the situation that a student outperforms his teacher based of information theory? Is it related to the other information sources or mentors? In other words, is it not more rational to use more than one expert or supervisor in the machine learning problems? The answer to these questions will be discussed in the following section. The idea of ensemble classification will be introduced. This will build the foundation of the classifier employed in the next chapters.

2.3.5. Theoretical aspects of ensemble classification

There are several methods in the literature for improving weak classifiers (Alpaydin E., 2004). Ensemble learning can provide a solution or modification for these problems (Soria Frisch, 2012). It uses multiple models to obtain higher accuracy in classification than could be obtained from any of the constituent models. The general idea of ensemble (also known as jury or committee) learning has an intrinsic connection to our daily life experience (Hastie et al., 2008) (Liang, et al., 2011). In decision making situations we ask experts about their opinions and try to make the best decision based on the degree of the deftness of a particular expert. In machine learning, this concept refers to a finite and preselected collection of alternative models (Opitz, et al., 1999) which has so far shown to improve several practical classification problems (Liang, et al., 2011). It seems that the combination of similar classifiers is plausible to outperform each single classifier. In fact, the combining classifiers can reduce the variance (Zhang, 2010) and accordingly will lessen the classification error (Alpaydin E., 2004) (Hastie et al., 2008). The concept of combining several classifiers has received other names in the literature (Soria Frisch, 2012) such as mixture of experts, classifier fusion, collective recognition method, “divide and conquer” classifier, and voting pool of classifiers, to name some.

In an ensemble classification, the fusion level is the stage at which processing chains are combined to yield of a group of classifiers. In the next step they are evaluated and weighted. Classifier ensembles in particular can be discussed in four regimes (Soria Frisch, 2012). Possible structures are shown in Figure 2.9. Soria-Frisch discusses different design principles as a guideline to ensemble classification in BCIs which can help a new user to identify the optimum solution when encountering a new paradigm (Soria Frisch, 2012). It means the decision making step in figure 2.9 can be more advanced than just applying the output of the classifier directly to the external device. Different types of ensemble classifiers exist in the literature including boosting, “Adaboost”, “bagging”, “stacking”, and “random forest”, which form four powerful and popular representatives (Breiman L., 2001) (Sun, et al., 2007). The ensemble strategy can be based on different classifiers like KNN, linear SVM and decision trees (Soria Frisch, 2012). Decision trees such as random forest (RF), along with classification and regression trees (CARTs) are some of the most efficient methods (Alpaydin E., 2004).

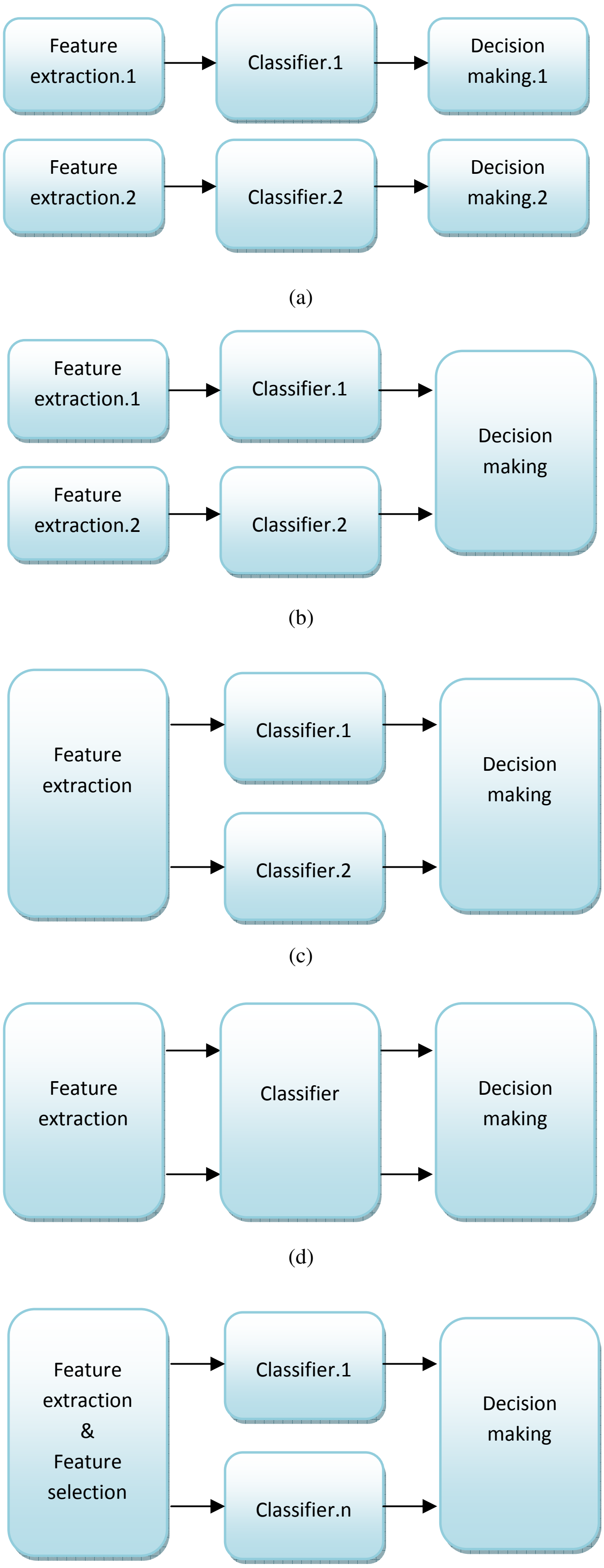


Figure 2.9: Different approaches for integration through (a) Separate operation (b) Fusion at classification level (c) Concatenation (d) Feature concatenation

The increasing interest in using ensemble classifiers among the machine learning community sparked the idea of employing this principle for BCI applications (Shoaie Shirehjini et al., 2009). In (Lotte et al., 2007) authors characterized the group of classifier ensembles as one of the best alternatives for designing BCI systems which are not biased by a particular database or application. Using the BCI competition data, a P300 BCI system was used to exemplify this concept (Rakotomamonjy et al., 2005). The idea was used in the early stage of BCI research (Pfurtscheller et al., 1993) but was not completely analyzed and mathematically scrutinized (Soria Frisch, 2012). Ensemble classification may cope with variability in EEG signals and leads to a more reliable and robust classification accuracy (Rakotomamonjy et al., 2005). Inspired by these works and after some investigation on the classifier techniques in BCI applications, we have considered some ensemble algorithms and investigate their advantages over the traditional single classifiers for our BCI systems.

2.4. Our contributions

In this chapter several aspects of EEG signal processing and the concepts of learning and machine learning were introduced. Algorithms developed in literature for this purpose are enormously diverse and covering all available methods is not possible in a single chapter. The methods introduced in this chapter provide a fundamental step in building the BCI structure for the following applications. Table 2.2 presents a summary of BCI applications described in the following chapters.

Table 2.2: A summary of BCI application contributions in the dissertation. Two types of applications are investigated: Movement control and gaming

BCI Application	Features Extraction	Feature selection	Classification Method	Hardware	Software
First group of Applications					
Virtual hand control	Fractal Components	--	SVM	g.tec-simulation	MATLAB GUI
Hand Prosthesis	Higher order Statistics	mRMR	Ensemble SVM	g.tec simulation	MATLAB GUI
Second group of Applications					
EPOC Robot Navigation	Power band	mRMR	Ensemble SVM	g.tec	MATLAB GUI
Avatar Navigation	Matching pursuit	mRMR	Ensemble LDA	Emotiv	Second Life-C++

Chapter 3

An offline study on classification of hand movement[1]

3.1. Introduction

In the last two chapters we have mainly focused on the theoretical aspects and the basic knowledge of elements of brain computer interfacing. After building the theoretical foundation, one would like to investigate the experimental advantage and to evaluate the performance of the entire system, while dealing with the real data and environments. Therefore the main theme of this chapter and the next four chapters is to evaluate the designed BCI system over different datasets in movement control applications.

In the current chapter an offline study on the classification of hand movement is fulfilled. The expression "offline" is used when the dataset has been already collected during the associated online experiments. The primary purpose and focus of this study is to investigate whether a new feature set can provide higher classification accuracy while using the similar paradigm and dataset. Control of sequence of hand grasping and holding is a critical issue in developing an EEG-based hand control (Carmena et al., 2003) (Hochberg et al., 2012). In addition, for the purpose of man-machine interface consistency, there should be some coherence between the intended movement and the task which the subject literally imagines. To this end, I designed and implemented an online synchronous BCI paradigm based on the imagination of hand grasping and resting states which had one-to-one consistency with the provided feedback (Hazrati et al., 2008). The goal of that research was to develop an interaction technique that allows the BCI user effectively to control hand grasp in real-world scenarios. Using data collected in these series of experiments, I will study now

---

[1] Part of this work was published in : **M. Kh. Hazrati**, A. Erfanian and U. G. Hofmann," Fractal Components From Electroencephalogram Provide Features For Brain Computer Interface", *20th Biennial International EURASIP, Biosignal, June 2010.*

whether an alternative feature selection method can improve the accuracy of the system. So an analogous online paradigm was developed and the data recorded from the previous work was applied to evaluate the proposed scheme.

3.2. Experimental Setup

Experiments were conducted in the BCI Laboratory, Neural Technology Center (NTC), Departments of Electronic Engineering at Iran University of science and technology (IUST) in Tehran. The experiments were carried out with six healthy university students (2 male, 4 female, mean age: 24.3) who volunteered for the study. The participants had no previous experience in working with BCI systems. When questioned after the experiments, subjects asserted that they enjoyed the experiment session. However, it was strenuous for them to concentrate and considerable efforts were needed to perform the mental task. We have designed a special scenario which contains two states of brain activity. The data was collected and classified to two classes of the relaxation and the imagination of hand movement. During the experiments, the users had to interact with a virtual reality environment. In this BCI system subjects tried to control their attention according to a predefined paradigm in imagination of hand movement and relaxation periods. At the start of trial, a blank screen was shown to the subject for 1 s. Then the subject observed an opened hand on the screen which indicates the onset of the relaxation phase. In this period the subject should not perform any specific mental task but to try keeping the hand open for 5 s. Following the relaxation phase, as displayed in Figure 3.1, the hand re-opened again (in case it was partially closed before) a ball began to fall and by reaching the virtual palm the second phase of the trial started.

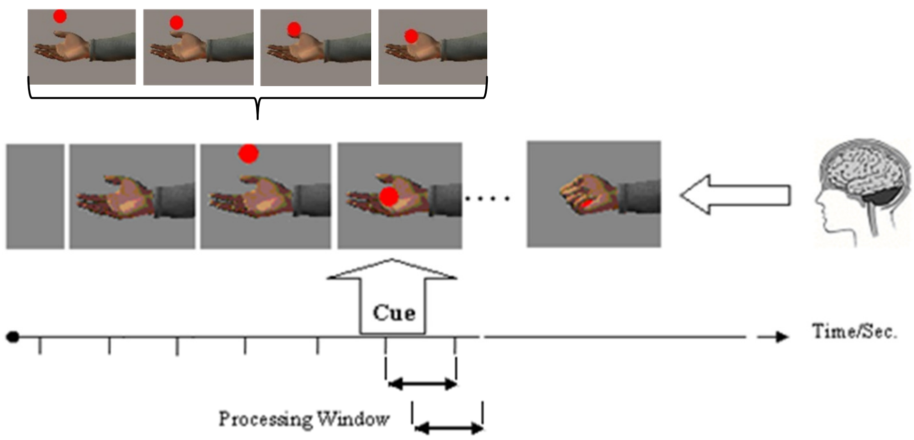


Figure 3.1: The appearance of a red ball on the palm cues the onset of imagination for the hold and grasps mental states which each lasts 5 s. Transition phase between relaxation and imagination during a trial is shown in details. The cue (ball touching the hand) marks the start of the second 5 s phase. The bottom part of the figure shows two consequence processing window (with 75% overlap), lines on x-axis mark 2 s intervals.

This transient phase was fixed and consisted of four images displaying every 0.5 second (total of 2 s). This phase was considered as a cue for conveying the message that the imagination section is about to start. At 7 s after onset of the trial, an active feedback phase was started. The entire feedback phase lasted 5 s. Users were asked to imagine of hand-grasp action by grasping the ball (i.e., closing phase). In the online experiment biofeedback was provided every 0.5 second in both holding and closing phases based on the results of a binary neural-network based classifier (Hazrati et al., 2008). The closing sequence was controlled by the output of a decision maker which was calculated as a simple logical majority vote function between outputs of classifiers trained over different channels (except Fp1). Two schemes of classification process were used for virtual hand grasp control: adaptive and static classification. In contrast to the static classifier, the adaptive classifier was continuously updated during the experiment. Adaptive scheme was used to train the classifier on-line during the first sessions. In some sessions, the adaptive classifier was used during the first five runs and static for the next subsequent runs (Hazrati et al., 2008).

Thus, during the experiment each subject observed the sequence of hand grasp uniquely based on his/her mental activity and the result of the classifier. The output of the binary classifier determined whether the next image should be displayed or the previous one should remain on the screen. Upon detection of motor imagery by the classifier during the trial, the hand was being closed gradually. So during the relaxed phase, the detection of movement-type EEG activity by the classifier led to the biofeedback corresponding to the closing hand movement. Similar to the imagination phase, it consisted of ten separate steps of closing the hand (Fig. 4.2). If for instance the classifier detected three movements during the relaxed phase then the third image was shown on the screen and 70% accuracy for this phase was recorded.

Figure 3.2: Subject observed an empty hand for 5 s (1280 points of data) in each trial. During this period, the result of the online classification was displayed to the subject as bio-feedback. It consisted of ten separate steps of closing the hand. If the classifier recognized the relaxed class all over 5 s, the first image of an open hand is kept till the end of the trial.

In the cases that the subject was successful during the entire trial, i.e., the classifier recognized a holding phase for the entire first 5 s (0 to 5 s) and a closing phase for the entire second 5 s (7 to 12 s) 100% classification accuracy was recorded. Thus, experimental control and bio-feedback were identical in the first and second 5 s period, except that the hand was holding a ball in the second period and participants' goal was opposite in the two phases: Keeping the virtual hand open for the first 5 s and making it close completely at the end of the second 5 s. The first run of each session was recorded without feedback. The data of this run was used to initialize the classifier and to calculate the normalization and artifact reduction parameters. Figure 3.2 shows some steps of virtual hand grasp control.

3.3. Implementation

Optimized computer software was required to implement the virtual reality environment for hand grasp control on a PC within the BCI. In our case, I used MATLAB Simulink (THE MATHWORKS, 1998–2000), Real-Time Workshop (THE MATHWORKS, 1999–2000), and Real-Time Windows Target for Windows XP for on-line data acquisition, filtering and real-time ocular artifact suppression, feature extraction, classification within an interactive virtual reality environment. Each block was written in C-code using the S-function technique which is supported by Real-time workshop in MATLAB environment. So the entire system is able to read the EEG data from the hardware and to process and provide feedback within some millisecond delay (Figure 3.3).

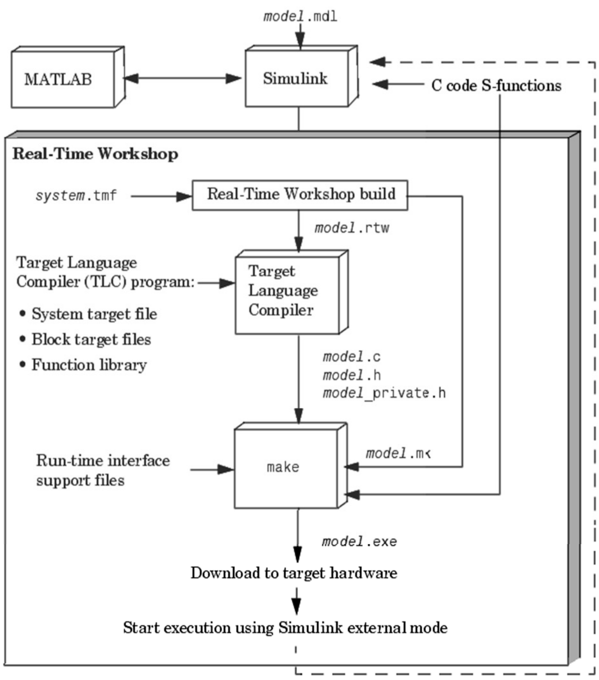


Figure 3.3: The general structure of real-time Workshop in MATLAB. Each block was written in C-code using the S-function technique.

Data recorded during each run of experiment was saved automatically at the end of the experiment in separate MATLAB files in format (*.mat) containing variables:

OnlineData =

- time: [30721x1 double]
- signals: [1x8 struct]
- blockName: [1x58 char]

The EEG data signals were recorded using a g.tec amplifier (Guger Technologies, Graz, Austria) at a sampling rate of 256 samples/s from positions Fp1, F3, C3, F4, C4, Fz, Cz and Pz by Ag/AgCl scalp electrodes mounted according to the 10-20 international standard and were band-pass filtered with a cutoff of 0.1-45 Hz using Chebychev II filter order ten.

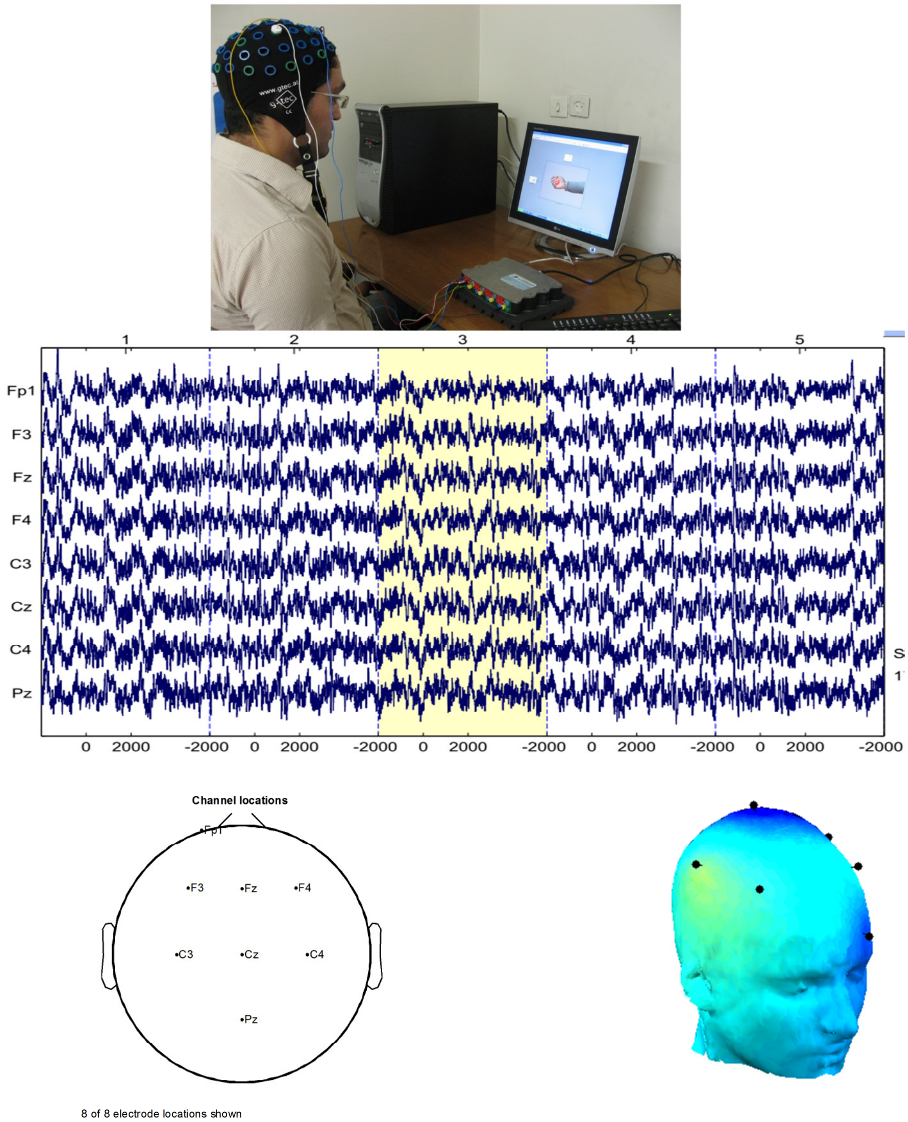


Figure 3.4: (a) Experimental setup for hand grasp application. Six subjects participated in online sessions for control of hand grasp and hold. Each subject participated in at least ten sessions of recording. Each session was recorded on a different day. (b) Epoched data for the last 7 s of each trial (preparation for movement + imagination section) recorded from 8 channels. Each segment shows 1792 (7 × 256) data points, units on x-axis are ms, (c) A 2D and 3D image of channel location used for the recording of EEG during the experiments.

All recording channels were referenced to the right earlobe and a ground electrode at the left earlobe. Eye blinks and vertical eye movements were recorded from Fp1 electrode site located on the forehead exactly above the left eyebrow. The subjects sat on a relaxing chair with armrests. In the preparation phase and before the start of the experiment, the impedance of each electrode path was measured. The experiment started when satisfactory values were obtained. The value of impedance measured in the connection of electrode tip and scalp should be set below 10 kΩ. Nu Prep™ Electrode prepping gel and Ten20™ conductive EEG paste were used to prepare the scalp skin and to decrease the impedance value. During the experiments EEG signals were continuously collected and processed. Subjects were free to move slightly their eyes or blink. The experiment for each individual consisted of different sessions and each session was conducted on a different day. Each session consisted of at least 10 runs and each run consisted of 10 trials with online feedback. A resting period of about 2-5 minutes was enforced between every two runs.

3.4. Evaluation of the recorded data

3.4.1 Evoked potentials

In order to evaluate the recorded data, some further offline analyses have been accomplished. The grand average ERPs for this dataset are presented in Figure 3.6 and Figure 3.7 for the imagination and relaxed phases, respectively. In Figure 3.6 units on x-axis are ms which represent the last 7 s of the entire trial starting from the point when the red ball appears on the screen till the end of imagination trial. The first deflections happen at about -1750 ms, consisting of a negative peak followed by a positive peak all over recorded sites. These peaks generally decrease from parietal sites to the frontal sites and decay at Fp1. In Cz negative peak maximizes with -2.3 µV at -1750 ms and positive peak reaches to +2.4 µV at -1550 ms. In general the deflection reaches its extremes a few millisecond earlier at Pz. These values are measured at Pz with the maximum -3.3 µV negative peak at -1770 ms and +5.1 µV positive peak at -1580 ms. Evidently this is the potential evoked by the event at -2000 ms, which is the time that the red ball appears on the screen. At Fp1, F3, F4, and Fz, this bipolar evoked potential is followed by a 2.0 µV peak at -1300 ms and a -1.2 µV peak at -1000 ms. Being maximum at Fp1 with 2.2 µV peak, the cerebral origin of this potential is not certain: It might be due to systematic upward eye movements induced by the stimulus. These are followed by a potential evoked by the start of the imagination trial (the ball touching the hand), consisting of a -2 µV negative peak at +200 ms maximum at Pz and a 4.5 µV peak at +400 ms.

The duration varies in each channel but usually it lasts till 500ms, since it is broader than the previous corresponding peak evoked by the -2000 event, it may include a P300. The maximum amplitude decreases from parietal to frontal sites and reaches 2.2 µV in F4 location. In the last second of trial, there is a similar waveform with a 3.5 µV peak at +4300 ms maximum in Pz.

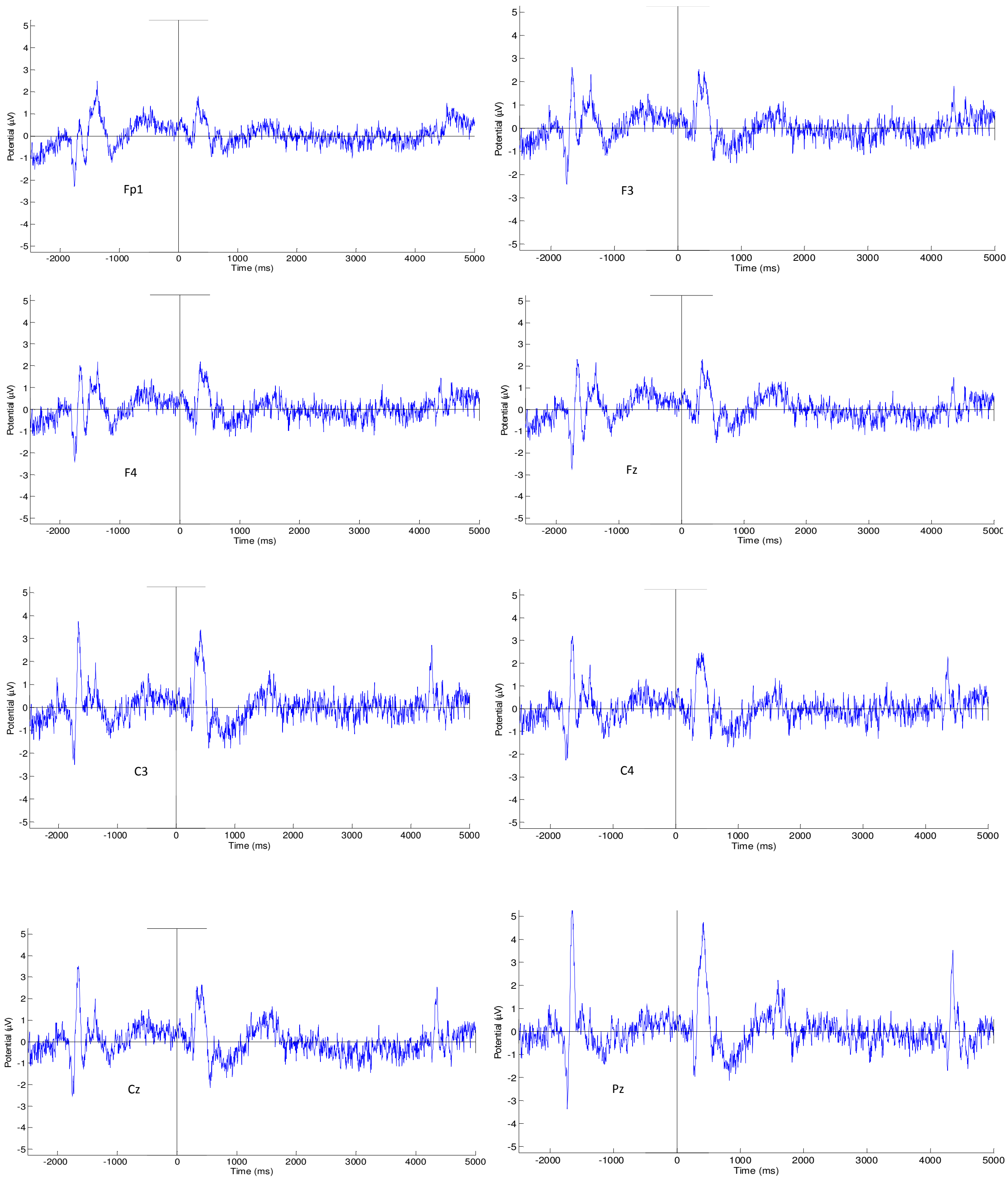
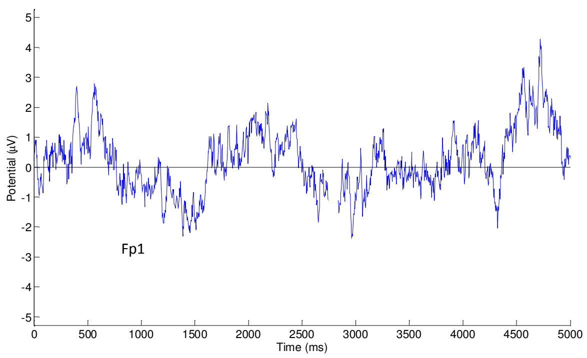


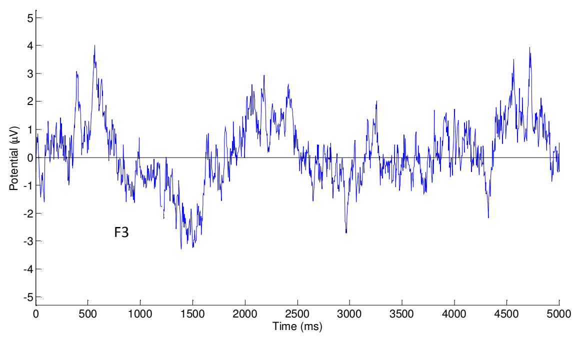
Figure 3.6: The grand average over 8 recorded channels (positions Fp1, F3, F4, Fz, C3, C4, Cz and Pz) in this dataset. Units on x-axis are ms which represent the last 7 s of the entire trial starting from the point when the red ball appears on the screen till the end of imagination trial. Time zero shows the start of imagination trials.

It can be also seen at the same time at Cz, C3, C4 and Fz with 2.1 μV, 1.8 μV, 2.5 μV and 1.5 μV respectively. Apparently, this is another evoked potential, which is quite surprising since no fixed event was scheduled to happen at +4000 ms. However, considering the experimental paradigm, the highest concentration was usually achieved in the last two seconds of the experiment, such that a positive biofeedback became most probable in the final second of the experiment. This can be the reason for the late appearance of the third evoked potential in this dataset.

Figure 3.7 shows the grand average in the relaxation phase. The signal/noise ratio appears to be worse than in the imagination phase. The negative peak of the potential evoked by presentation of the opened hand is hardly seen. Rather, the first deflection is the positive peak of the evoked potential, consisting of a 4.1 μV peak at about 300 ms, maximum at Pz and C4. The maximum amplitude decreases from parietal to frontal sites and reaches 2.2 μV in F4 location. Especially at Pz and Cz electrode locations, the grand average shows another positive peak after 500 ms, most probably a P300 component according to its timing and topography. This is followed by a -5.2 μV negative peak 1500 ms after the start of the relaxation phase, maximum at Pz. This may be the result of observing some movement in the displayed hand (biofeedback) which happened usually in the beginning of the relaxation phase for the majority of subjects. This might have been the start of some slow potentials that might have accompanied movement observation or imagination. But note that such slow potentials were here largely filtered out, due to the lower frequency limit of the amplifier at 0.1 Hz.

The appearance of biofeedback varied widely in this interval across trials and participants. Some subjects where totally successful in keeping the hand open all during the trial and for the others it was not a trivial task. Thus it is likely that the happening of the movement biofeedback was not time-locked to the cue for all subjects and over all trials. Still there are some noticeable changes in the EEG signal in this phase.





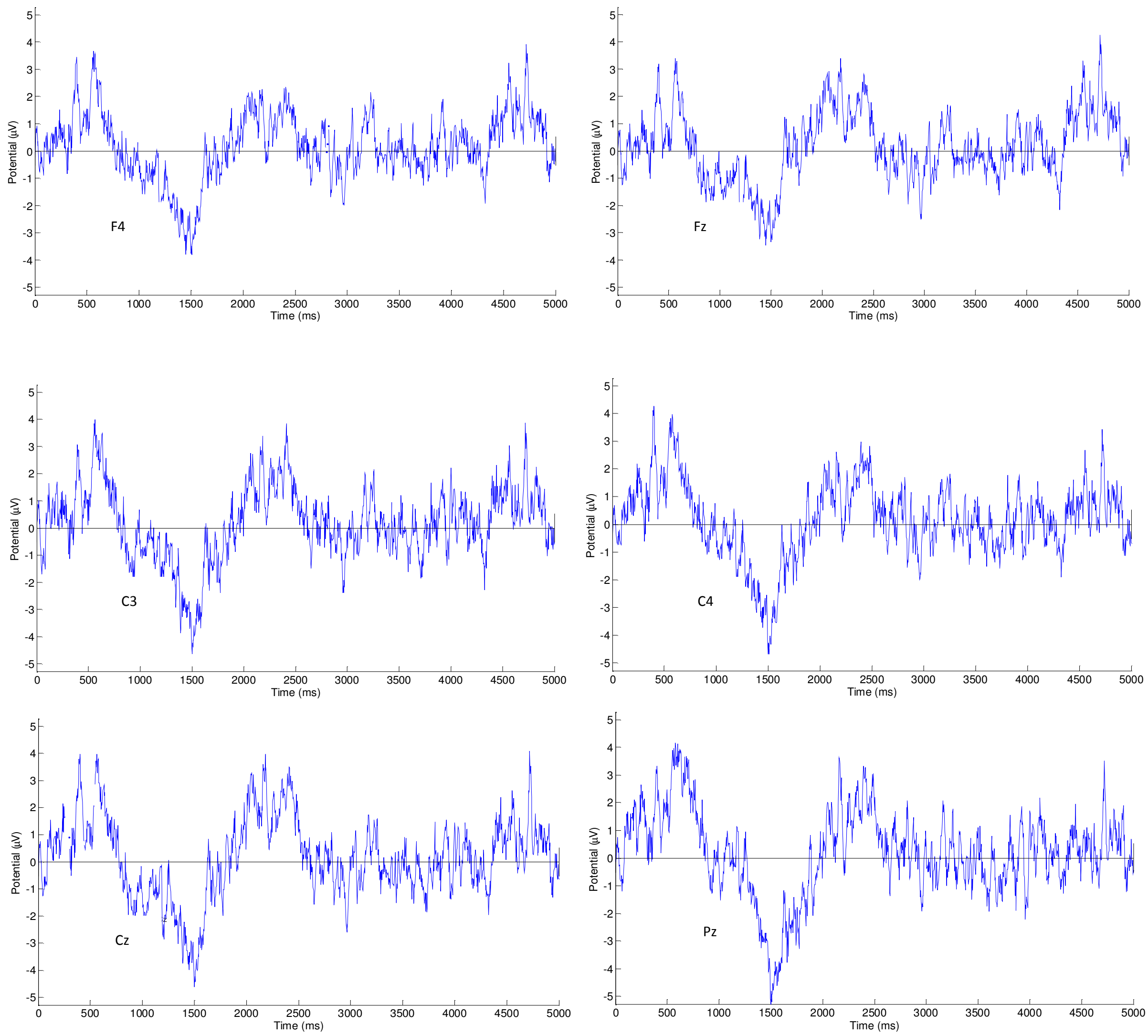
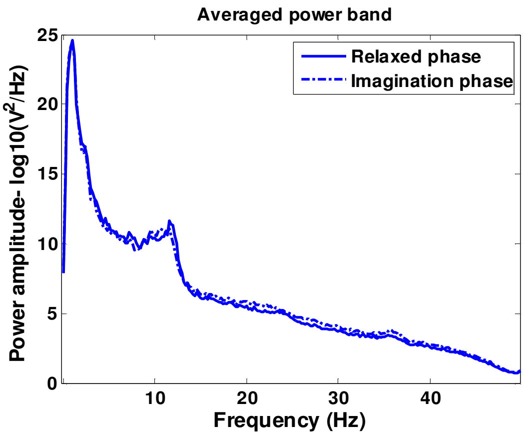


Figure 3.7: The grand average over 8 recorded channels (positions Fp1, F3, F4, Fz, C3, C4, Cz and Pz) in this dataset. Time zero shows the start of relaxed phase.

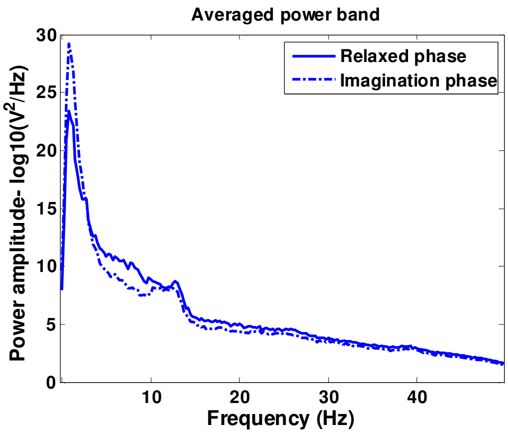
3.4.2 Power band analysis

Figure 3.8 illustrates the power band spectrum averaged over all trials and subjects. The EEG power was computed over 8 channels. The amplitude is calculated in log10 (V^2/Hz) unit. Band power features were depicted for the imagination and relaxation signals acquired from each individual channel. The discrete Fourier transform of the imagination and the relaxation signals were computed for each trial and positive integer frequencies. The signals were first padded with zeros up from 1280 samples to 12800 samples to increase the resolution of the Fourier transform. Magnitudes of the Fourier transform were squared and the 10-base logarithm was obtained.

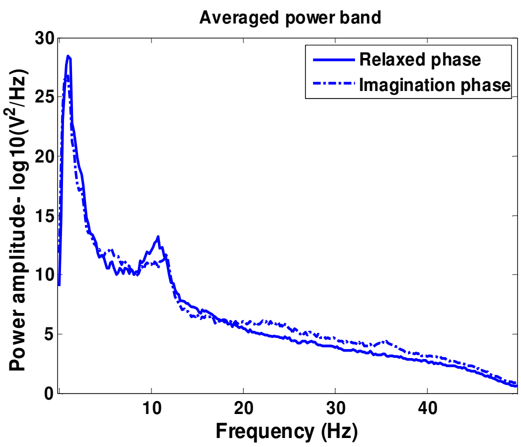
Values of the EEG power in delta band (1-3 Hz) of imagination phase were found to be noticeably higher than those in relaxation phase at Fz, F3 and C3. The Delta ERS topography can be correlated to the brain activity in SMA (supplementary motor area) and M1 (Primary motor cortex) which is represented by C3-F3 and Fz, respectively. Due to its low frequency it might represent some kind of slow motor potential similar to Bereitschaftspotential (Pfurtscheller et al., 2001). In C4 and F4 the difference is smaller. There is a slight increase in the power of theta band in relaxation phase over all recording sites except Fp1 and F4.



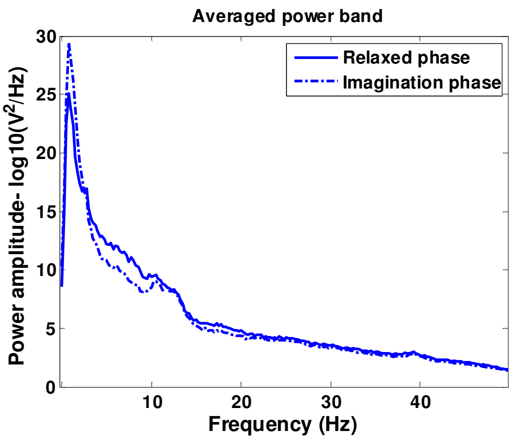
Fp1



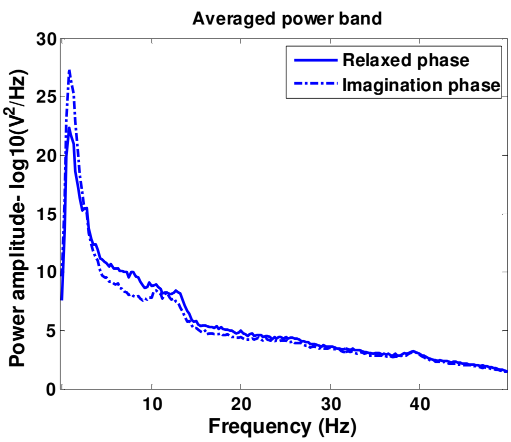
F3



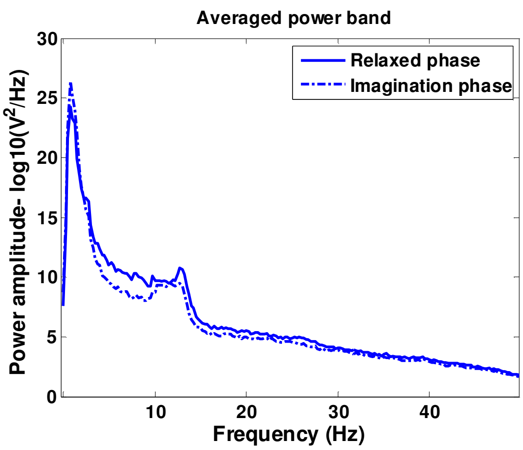
F4



Fz



C3



C4

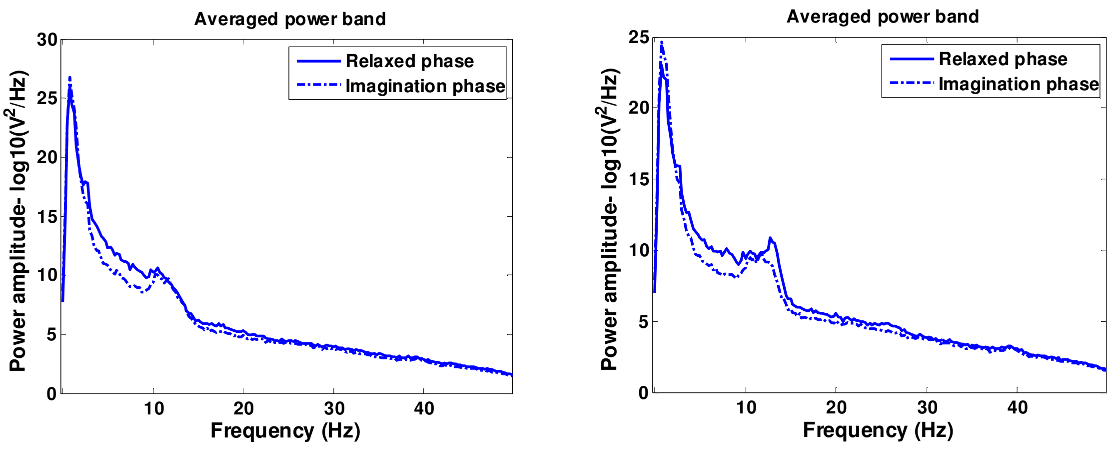


Figure 3.8: Average power band over the entire dataset for two different brain states demonstrated over Fp1, F3, F4, Fz, C3, C4, Cz and Pz - Unit on x-axis is Hz, unit on y-axis is V^2/Hz.

In general, a selective enhancement of alpha (8-13 Hz) power is seen during rest and imagination and the average power in alpha band is higher in the relaxation phase in comparison to the imagination phase at C4, F4 and Pz. Also another increase in gamma band (30-45 Hz) can be seen in C3, F4, Fz and Pz. Also at Fp1 a small Gamma peak exists. This can be the effect of a degraded movement of subjects in some trials.

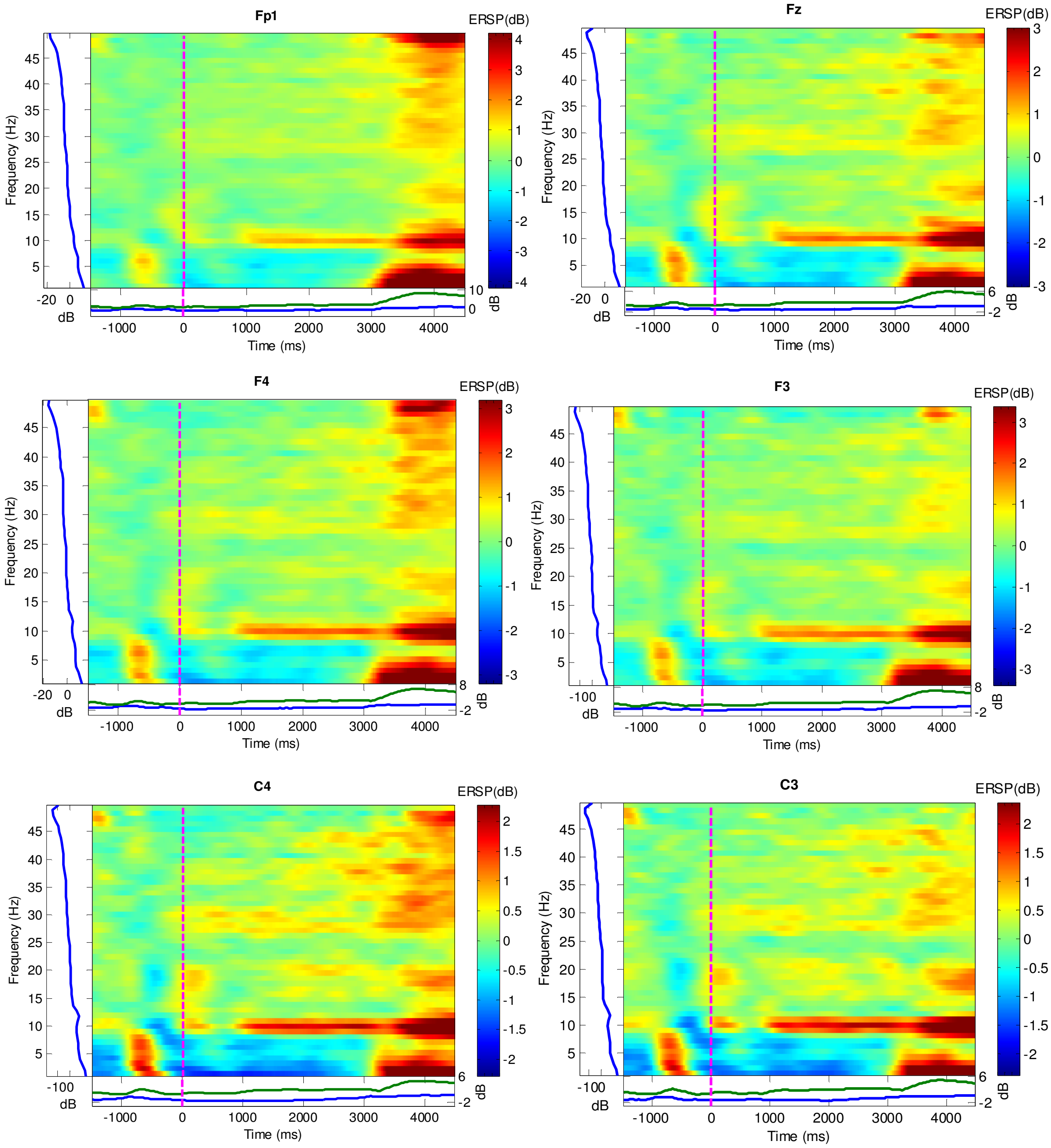
3.4.3 Time-Frequency analysis and event-related oscillations

While calculating frequency spectrum using FFT, information is integrated across time. To overcome this shortage, an offline time-frequency analysis was applied to investigate the ERD/ERS changes in this dataset. For calculating the ERD in a particular band the following steps are used (Makeig, 1993):

- Band-pass filtering of all trials
- Obtaining power samples by squaring the amplitude of each sample
- Averaging the above values (power samples) across the entire trials
- Smoothing the data by averaging over time

By repeating the above calculations for each particular frequency band ERSP changes over the entire dataset is calculated. Figure 3.9 demonstrates the grand average of the results of ERSP analysis for different channels before and during the imagination phase. Unit on x-axis is seconds, unit on y-axis is dB. The first 2 s of data is considered as baseline for the rest of the trial. It means the average power between -2 s and zero point are averaged and subtracted from the entire processing data. The baselines are calculated for discrete values of frequencies over the entire spectrum. Using this technique, time-frequency changes can be seen more accurately in a wide range of frequencies and time intervals.

The first time-frequency fluctuation in figure 3.9 is an ERD in delta and theta bands that happens between -1.5 s to -1 s. At the same time a short lasting alpha ERS is observed which reaches its maximum at Pz. Following this an ERS in delta and theta bands starts at -1 s and lasts for 0.5 s. Again the maximum power is seen at Pz. The power decreases when moving from Pz towards frontal sites. Before the start of imagination phase another ERD in Theta and Beta bands is observed clearly around -0.5 s in all recording sites. There is also an alpha-band ERD visible before the start of the imagination phase between -0.5 s to -0.2 s on Pz and in some extent in C4, C3 and Cz.



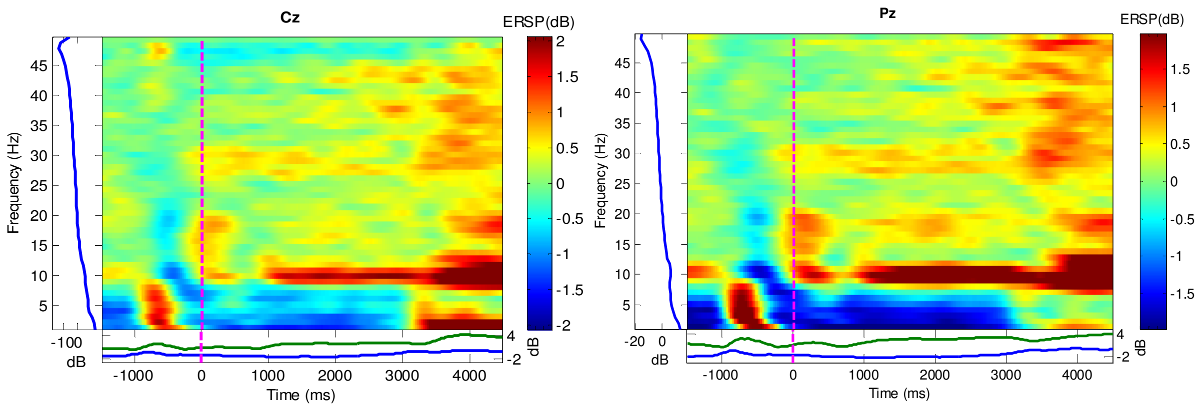


Figure 3.9: Time-Frequency analysis based on ERSP during and before the imagination phase. Event related desynchronization (ERD) and event-related synchronization (ERS) in different frequency bands over all subjects depicted for all recording sites. These fluctuations can be seen in either alpha and beta bands during or right before a movement is executed. (Negative values represent ERD and positive values represent ERS)- Unit on x-axis is milliseconds, unit on y-axis is Hertz. The color bar represents the absolute values of ERSP in dB. Zero point indicates the start of hand movement imagination by the subject.

Following this baseline phase, delta and theta ERD are seen from the onset of the imagination phase till 3 s all over channels, particularly at Pz, Cz, C3 and C4. From 3 s till the end of the trial an ERS can be seen in these bands. Alpha ERS starts at 1 s and lasts till the end of the trial over all recording sites and reaches its maximum at Pz. All over channels an ERS in Gamma is also seen after 3 s of the onset of the imagination phase. The increase was most prominent in the central region towards right hemisphere at Pz and C4. The Theta and Gamma ERS at the end of the imagination phase is also strong at Fp1 and can be the results of a movement artifact.

Figure 3.10 shows the similar analysis for the relaxation phase. No baseline is defined in the current analysis, so the ERSP powers are referred to the average of the entire trial. Delta and theta ERS exists in the first one and half second of this phase. Simultaneously an alpha-band ERD is visible over all channels. Staring from the onset of the trial, this Alpha ERD diminishes at 2 s. At C3 and C4 it lasts longer till 3 s. In the first 1.5 s, an ERD in Beta band is also seen clearly at C3, C4, Cz and Pz, when possibly the subjects were struggling to keep themselves in the relaxed state. It is followed by a Beta ERS which can be observed between 1.5 s to 2 s and also 3 s to 4.5 s in Cz, Fz and to some extent in C3 and C4 and frontal sites. Around 3 s after the onset of the relaxed state at F3, C3, Cz, and Pz an strong alpha-band ERS starts. All over channels sparse Gamma band ERD can be seen. The decrease is most prominent in the central region at Cz and Fz.

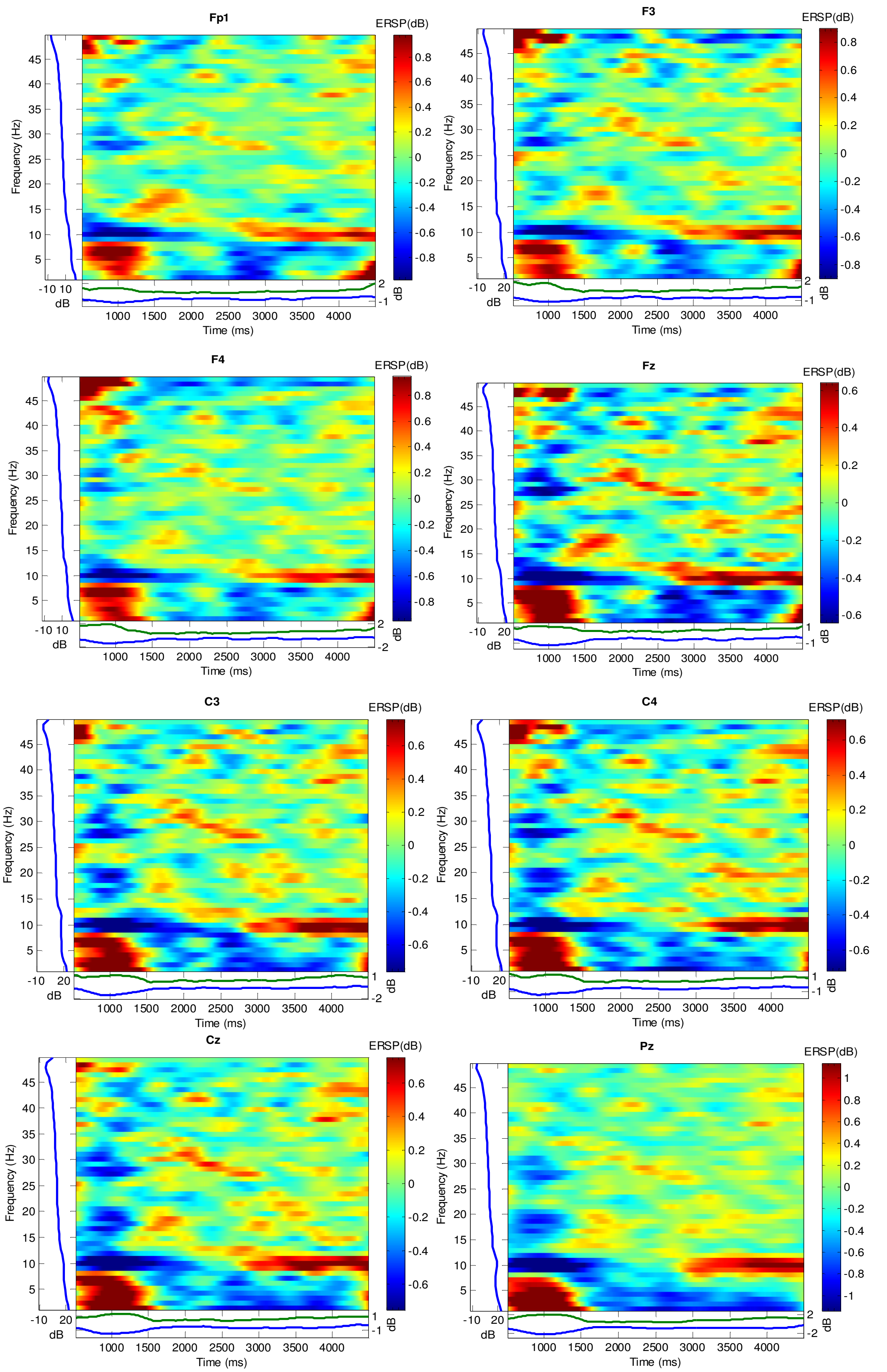


Figure 3.10: Time-Frequency analysis based on ERSP in relaxation phase. Unit on x-axis is milliseconds, unit on y-axis is Hertz. The color bar represents the absolute values of ERSP in dB. Zero point indicates the onset of the relaxed trial.

In this section we applied some offline analysis, the ERP grand average, ERD/ERS and the ERSP analysis, to evaluate the recorded data. These data show that the imagination of moving one's hand elicited ERD and ERS in different EEG power bands which can be used as a feature for detecting this mental task in BCI. Having the main structure of the BCI in mind, in the following I will go through all BCI components individually in order to construct the entire system. These elements are listed as: Preprocessing (filtering and artifact reduction), feature extraction, classification and decision making.

3.5. Materials and Methods

In the following I will explain each element of BCI thoroughly. The recorded EEG signal was treated as a simulated online scenario. The ongoing EEG in each trial of experiment is processed in two-second windows, overlapping with 75% overlap. In the first step an artifact correction and filtering technique was applied on this window. Then normalized power band features as well as proposed fractal features were calculated from each window and classified. The classification accuracy was measured based on the number of correct binary outputs of an ensemble classifier.

3.5.1 Automatic Online Ocular Artifact Removal based on FastICA

For automatic and online EOG artifact correction in this dataset a technique based on ICA was applied. The proposed method is based on the combination of *FastICA* and higher order statistics in an online scheme without using an extra channel for artifact recording. Although ICA has been applied broadly for artifact reduction and there were several studies done on automatic detection of artifact component (Frank & Frishkoff, 2007), (Viola et al., 2009) two major challenges remain. These methods are often not suitable for automatic detection and recognition of the artifact component in an online application. Furthermore there is also a need to use the information of the eye blink signal.

In fact there are common independent components between each two EEG channels. That nourishes the idea that the artifact information of EOG channels is already available in other recorded EEG channels. The correction methodology using EEG signals is similar to the one using artifact channels, however in this case the EEG signal is used instead of an EOG signal to extract the artifact components (Winkler et al., 2011). This approach has the advantage of being independent of an additional EOG channel, and is useful if the EOG signal is not recorded during data collection (Ng et al., 2008). The idea of not using the reference channel for artifact correction was used in offline studies (Delorme et al., 2001) (Barbati et al., 2004). However, it is completely innovative to apply this method for an online application. For this dataset, Fp1 channel was already recorded. But in our current study artifacts are extracted and rejected automatically by using an online technique based on fastICA without using the reference channel.

3.5.1.1 FastICA algorithm

FastICA algorithm developed by Hyvärinen et al. is a version of ICA with optimized programming to shorten the time of calculating the components (Hyvärinen, 1999). The number of components is usually selected equal to number of recording channels (Fatourechi et al., 2007). FastICA is based on fixed-point iterative method that uses maximizing non Gaussianity as a measure of independence. The data should be centered and whitened before the main algorithm can be applied. Centering the data means removing the mean from each segment of the signal $x \leftarrow x - E\{x\}$ and whitening is a linear transforming scheme that makes the new components uncorrelated with variance one $\{\tilde{x}.\tilde{x}^T\} = I$. PCA (Principal component analysis) is used as preprocessing step for the purpose of whitening the signal in FastICA algorithm. After whitening the data, the next step in ICA algorithm is the axes rotation. The idea is to minimize the Gaussianity of each projection on its corresponding axes. Here, unlike PCA approach, axes do not need to be orthogonal to each other. The FastICA algorithm iteratively searches for the direction of weights ω which maximizes the non-Gaussianity of the projection $w^T x$ (Hyvärinen, 1999). Table I contains the pseudo code for the single unit case. $f(.)$ is a nonquadratic nonlinear function and $g(.)$ and $g'(u)$ are the first and second derivatives of $f(.)$, respectively.

Table I: FastICA pseudo code- Single unit
1. Randomize the initial weight vector w
2. Let $w^+ \leftarrow E\ \{xg(w^T x)\} - E\{g'(w^T x)\}w$
3. Let $w \leftarrow w^+/\|w^+\|$
4. If not converged, go back to 2

This algorithm only estimates one of the independent components. Hyvärinen proposed several methods for the construction of multi units weight vector (Hyvärinen, 1999). Table II presents a simple pseudo code for it, where C is the number of the components, $X \in R^{N\,X\,M}$ represents the input samples.

Table II: FastICA pseudo code- Multiple units
For p= 1: C repeat the following:
Randomize the initial weight vector w_p
While w_p changes:
$w_p \leftarrow \frac{1}{M}Xg(w_p{}^T X) - \frac{1}{M}g'(w_p{}^T X)1w_p$
$w_p \leftarrow w_p - \sum_{j=1}^{p-1} w_p{}^T w_j w_j$
$w_p \leftarrow w_p / \lVert w_p \rVert$

The output is calculated as $S = WX$, where $W \in R^{C\,X\,N}$ is the un-mixing matrix. Each row of this matrix projects X into independent components. It makes the independent components matrix $S \in R^{C\,X\,M}$. Some general purpose and robust values for $\mathrm{f}(.)$ are (Hyvärinen, 1999):

$$f(u) = \log\cosh(u);\ g(u) = \tanh(u);\ g'(u) = {}^{1}\!/_{\cosh^2(u)} \tag{3.2}$$

$$f(u) = -e^{-u^2/2};\ \ g(u) = ue^{-u^2/2};\ \ g'(u) = (1-u^2)e^{-u^2/2} \tag{3.3}$$

Modified ICA algorithms have been developed rapidly over the past years in order to handle the biological signals where a linear combination of independent non-Gaussian sources form the signal. Some researchers suggested that there is a consistent and clear-cut relationship between ICA sources and some physiological or behavior signals (Barbati et al., 2004). It is suggested that recording the biomedical phenomenon simultaneously from multiple sensors or different locations might facilitate the computation and increase the accuracy (Ng et al., 2008). The dimension of the input signal, from the mathematical viewpoint, will not affect the functionality of the ICA algorithm. Indeed, the functionality is related to the statistical distributions of the temporary independent and spatially fixed concurrent electromagnetic activities. In general, it can be assumed that ICA components have (one or more) distributed sources in brain networks (Makeig, 1993).

3.5.1.2 Implementation

In order to implement my suggested method, I investigated the changing in the value of kurtosis of each independent component (IC) and it seems that this value can be an appropriate criterion for artifact recognition. Furthermore information pertaining to artifacts was directly extracted from the EEG data. Results demonstrate that the proposed structure is a suitable alternative for online rejection of ocular artifacts from EEG signals without using an extra channel.

We used the modified *FastICA* toolbox and *kurtosis* function in MATLAB and executed the computational processing in a simulated online scenario. When the mixing matrix adapts slowly over time, the true sources can be calculated almost online and at the time of recording. This characteristic of artifacts yields a contingency for correcting EEG signals. It should be considered that the artifact sources differ from artifact channels. Especially in the case of EOG, a combination of EOG and EEG sources are recorded over the same channel. It is necessary to extract the independent artifact sources in order to eliminate their imposed changes in EEG channels. Figure 3.11 summarizes the proposed scheme. Delorme et al. proposed the kurtosis and entropy as markers calculated based on the distribution of the signals, to measure unexpected and also transient events in the EEG signal (Delorme et al., 2001).

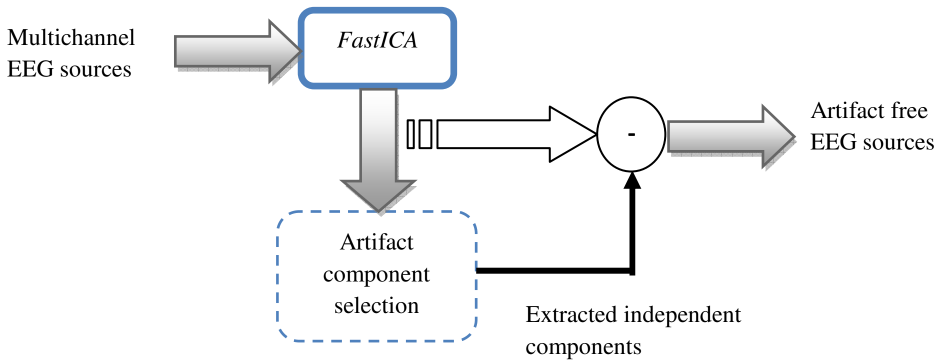
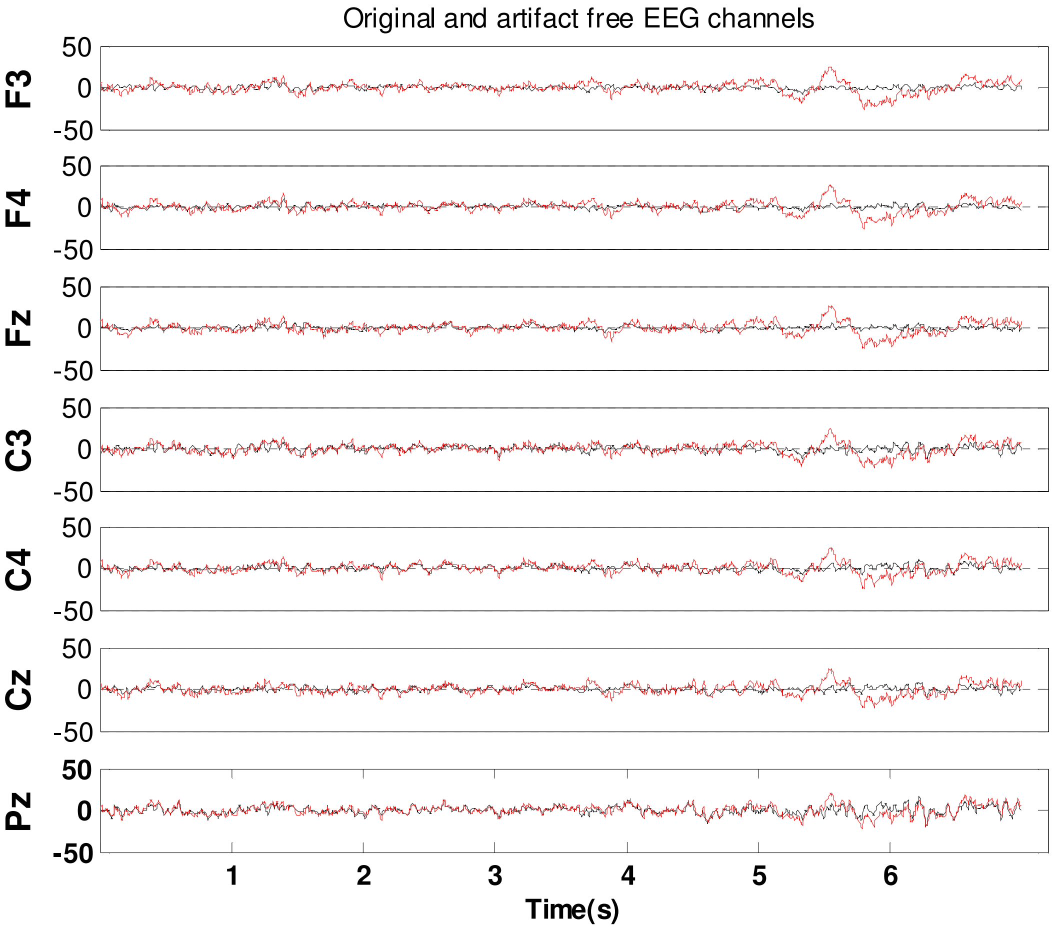


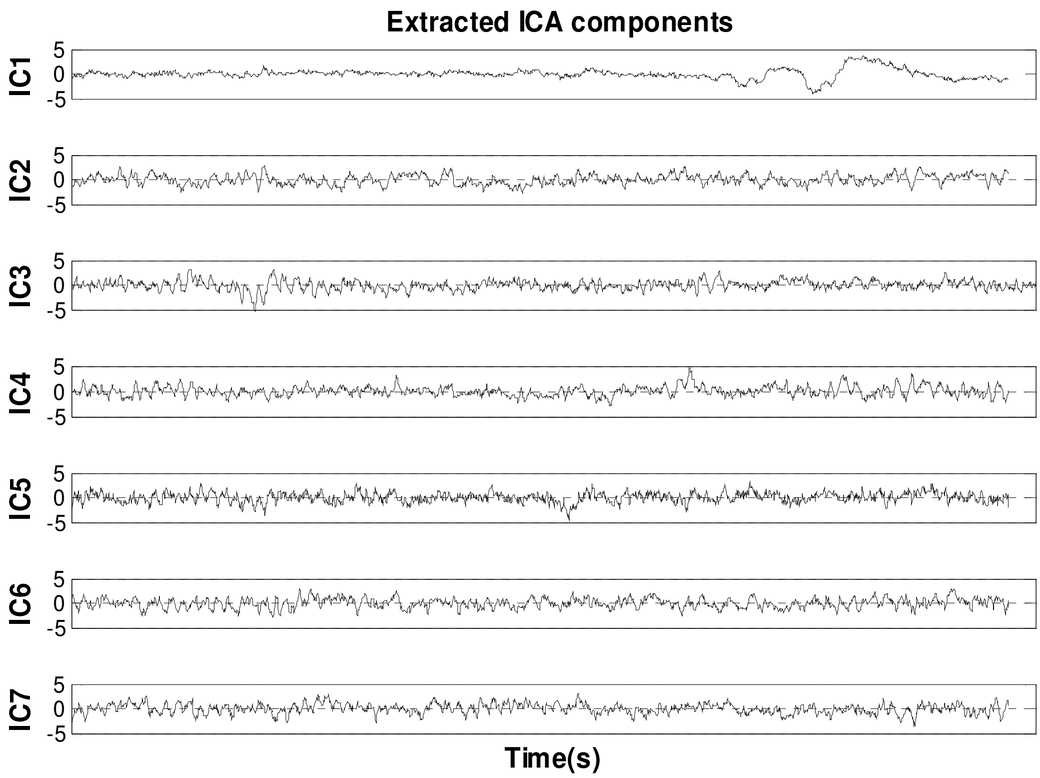
Figure 3.11: The structure of the proposed method for automatic artifact. The artifact component is selected automatically.

Considering artifact component we make the selection based on measured Kurtosis value of all independent components (ICs) using *FastICA* algorithm in an online implementation. In a similar work this value has been used for offline artifact rejection (Greco et al., 2005). Independent Components are extracted from the sliding 2 s windowed EEG signal with the sampling rate 256 samples/s over each recording channel. In order to automatically extract artifact components, we calculate the value of kurtosis for all extracted components in each window. The combination of ICA and higher order statistics has been already employed for artifact correction (Delorme et al., 2001) (Greco et al., 2005), however, it is limited to offline applications. Goal of ICA is to find components which have maximum kurtosis. The higher kurtosis, the sharper is the data distribution compared to the normal distribution. It is mathematically defined as the value of 4^{th} cumulant divided by the square of the 2nd cumulant. Fourth and second cumulants are the fourth moment around the mean and the square of the variance, respectively. It is summarized by the following equation:

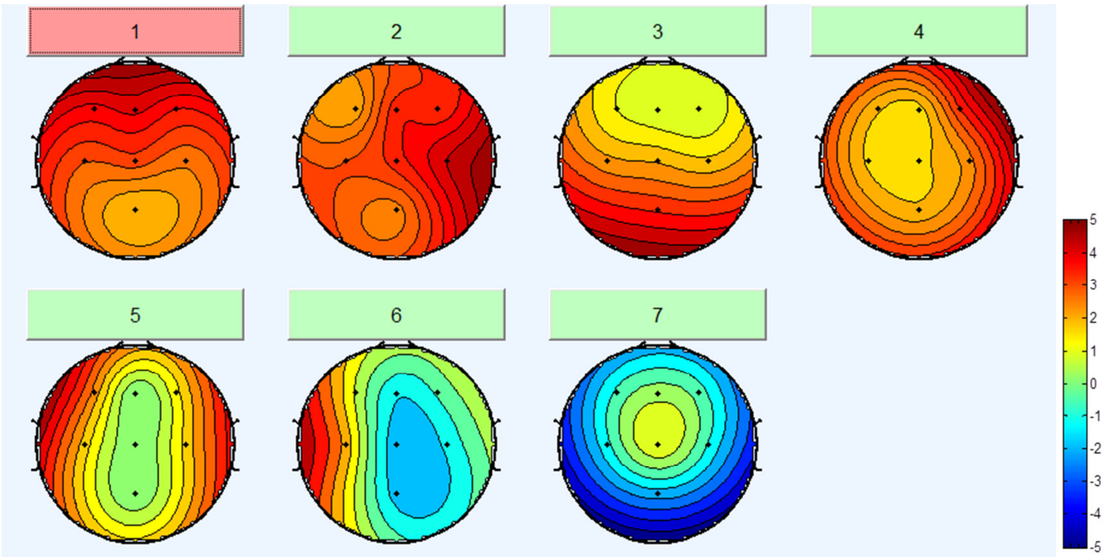
$$k = \frac{m_4}{m_2^2} - 3 = \frac{E(x-\mu)^4}{\sigma^4} \tag{3.4}$$

Where x is the recorded signal, μ is the mean of x, σ is the standard deviation of x, and *E(x)* represents the expected value of the quantity x. Kurtosis is a measure of how much a distribution is outlier-prone. In some literature the kurtosis of the normal distribution is defined to be 3 instead of zero (Delorme et al., 2001). Consequently, outlier-prone distributions (compared to the normal distribution) have kurtosis greater than 3 and, on the contrary, the kurtosis of distributions with less outlier is less than 3. Among all extracted components, we consider the one with the highest kurtosis as the artifact component, as it is known that the ocular artifact has peakedness compared to the ordinary EEG signals. We used the *K=kurtosis (X)* function in MATLAB for this purpose. It returns the sample kurtosis of values in *X*. In this definition, for vector input *X, K* is the fourth central moment of *X*, divided by fourth power of its standard deviation. With the present dataset, the average value of kurtosis for artifact free signals was measured less than 2.5 averaged over all subjects, however this increased in contaminated EEG signal up to 7.38. By applying the proposed method, the independent component related to artifact is selected automatically by setting a threshold of 5.5, which was selected by trial and error. In practice the optimum value for threshold should be selected individually for each subject. The data recorded from the probe run in the beginning of each experiment has been used to predict the threshold value. The data was normalized for each subject before applying the method. Over 92% contaminated data could be cleaned using this method. Figure 3.12 demonstrates the recorded signal and its corresponding independent components and the Artifact-free EEG. In Figure 3.12 units on x-axis are ms and units on y-axis are µV. In the bottom figure red and black legends represents the original signals and the signal after artifact rejection. This particular artifact is probably a blink, as may be seen from the behavior of the uncorrected EEG; It has its maximum at F3 and F4 and diminishes noticeably at Pz. 2D plot of ICs over the scalp is provided to locate the artifact component in this interval.


Original and artifact free EEG channels
F3
F4
Fz
C3
C4
Cz
Pz
50
0
-50
1
2
3
4
5
6
Time(s)


Extracted ICA components
IC1
IC2
IC3
IC4
IC5
IC6
IC7
5
0
-5
Time(s)

(a)



(b)

Figure 3.12: (a) Artifact rejection in a 7s window - contaminated EEG signal (red line) and artifact free signal (black line). Unit on y-axis is μV and unit on x-axis is s. Artifact Components for the same interval (IC1 is recognized as artifact component $K=6.9$), (b) 2D plot of independent components over the scalp. The color bar represents the power distribution.

The correctness of automatic selection was first validated offline for six subjects in the current. There are some practical concerns applying the current method. First of all there would be a delay of around 2 s between the corrected version of data and the raw data which is inevitable due to the inherent structure of the ICA method. In some cases the algorithm does not converge after a predefined iteration (n=1000). No artifact correction will take place if the calculated components are less than the number of channels. Figure 3.13 illustrates 10 seconds of data recorded from C3 and C4 of Subject 3 in this dataset, where the artifact free signal has similar dynamics all over the interval except the contaminated intervals (1 s - 2 s) and (7 s - 8 s).

The assumption of independency of artifactual signals from the EEG signals is compulsory for employing ICA (Makeig et al., 1996). Other critical assumptions consist of linear mixture of independent components, prior knowledge about the number of components, and stationarity of the sources and the mixture (LeVan P et al., 2006) (Park, et al., 2002). In practice these assumption are debatable. Computational timing varies for each segment and depends upon the number of independent components extracted from the signal. In average it takes around 85 ms to apply the algorithm and to generate the artifact free signals. So I have proposed and implemented another online ocular correction method which is considerably faster than current approach. It has been applied to the other BCI setups in my dissertation.

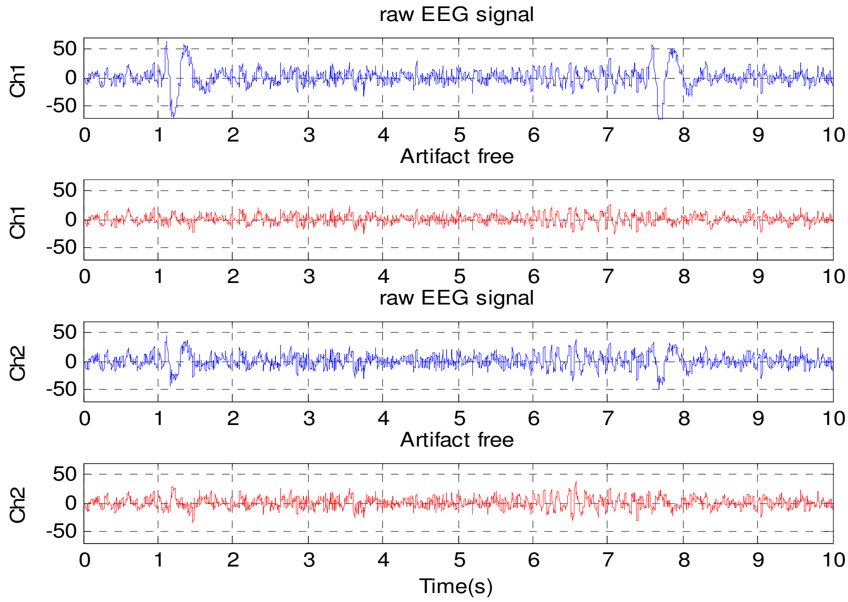


Figure 3.13: Ten seconds of data recorded from C3 (Ch1) and C4 (Ch2) of Subject 3 in the current dataset, where the artifact free signal has similar dynamics all over the interval except the contaminated intervals (1 s - 2 s) and (7 s - 8 s). Using online artifact correction methods each EEG channel can be cleaned from the eye blink online for further processing. By applying the proposed method, the independent component related to artifact is selected automatically by setting a threshold 5.5. ICs are extracted from the sliding 2s windowed EEG signal. Unit on y-axis is μV and unit on x-axis is *s*.

3.5.2 Feature Space

The first feature group consists of normalized band power features which were extracted from the segmented EEG signal. It is assumed that for each segment, which lasts between 2s to 5s, the signal is considered stationary (McFarland et al., 2006); meaning the statistical properties of the signal do not change over time (Haykin, et al., 2007). In Figure 3.14, power band features have been calculated in 2s-windows with 90% overlap after applying a band pass filter (0.1 - 45 Hz) and the EOG artifact correction as described before. Five major groups of frequency ranges were extracted from each channel of data and normalized in each window and used as a feature for the classification. The average power in Theta band is higher than in other frequency bands and it shows around 10% decrease during the imagination interval (7 s - 12 s in Figure 3.14). In Beta and Gamma band a short-term power increase is seen at the start of the imagination interval. In Beta1 the increase happened in the beginning of both imagination and relaxed phases (1.5 - 2 s and 7.5 - 8.5 s).

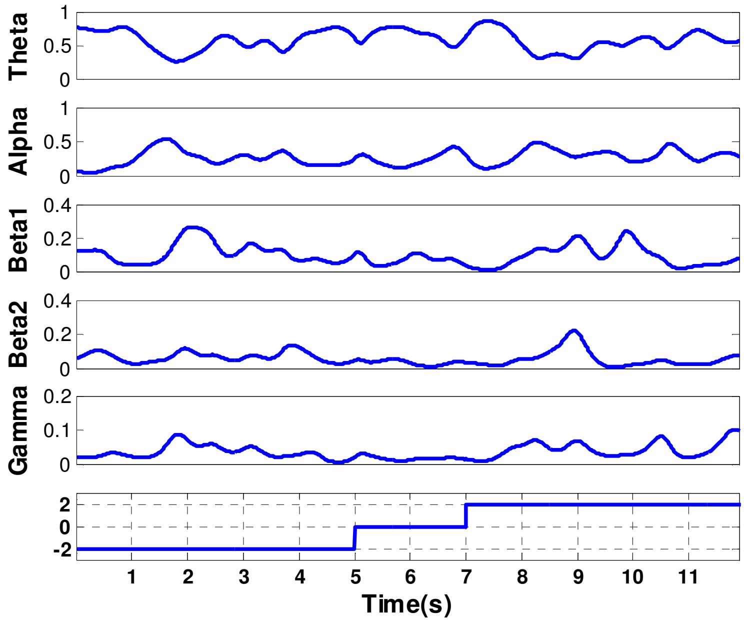


Figure 3.14: Changes of different EEG frequency bands over time of a subject in one trial of the experiment. Unit on x axis is seconds and unit of y-axis is normalized band power. For the first five seconds of recording (1 s - 5 s) the subject is instructed to be in a relaxed state. After a 2s transition, the next five second segment (7 s -12 s) is recorded in which the subject is asked to imagine movement.

3.5.2.1 Feature extraction based on fractal components

In this section I introduce a new class of features based on fractal components for BCI. An application of nonlinear science to investigate the alteration of EEG signals in two brain states, i.e. idle state and imagination of movement. Based on Coarse Graining Spectrum analysis (GCSA) we extract fractal components from the recorded EEG signal in the frequency domain.

The proposed method is based on the calculation of fractal exponents from the power spectral density in signals characterized by a frequency power law. Biological signals may possess periodic components in addition to $1/f^{\beta}$, and which is also a well-known phenomenon in EEG signal (Yamamoto, et al., 1993). CGSA was introduced to separate simple harmonic and fractal components from each other in the frequency domain. It is based on the spectral analysis of windowed signals using a Fast Fourier transformation partially modified according to (Yamamoto, et al., 1993). If the total spectral power of a signal consists of both harmonic and non-harmonic (fractal) components, it is possible to isolate the latter, because the fractal component is scale- invariant when rescaled. It will still retain its power when cross correlated with the original data (Pereda, et al., 1998). The non-linear nature of EEG exhibits random fractal structure with $1/f^{\beta}$ spectrum ($1<\beta<3$) (Kobayashi, et al., 1982). In contrast, rescaling of harmonic components causes a complete loss of spectral power when cross-correlated with the original signal. Here, β can be obtained as a negative slope of the fractal power versus frequency, in a log-log scale (Pereda, et al., 1998). Here the frequency range of 4 - 45 Hz was considered, where the spectrum presents the clearest $1/f - \beta$ dependence within it. As the processing was done on 2 s ongoing EEG data and in order to be equivalent with the first feature set slow frequencies (Delta band) were not considered in this study.

3.5.2.2 Coarse Graining Spectral Analysis (CGSA)

According to the definition introduced by Mandelbrot and van Ness (Mandelbrot, et al., 1968), fractal time series $x(t)$ satisfy the equation (4.5) for any $h > 0$ and t_0 , where $\stackrel{\text{def}}{=}$ implies that the distribution function is equal on both sides and H is the Hurst exponent (Yamamoto, et al., 1993). It explicitly demonstrates the nature of random fractal time series, where changing the time scale doesn't affect the dynamic of the signal.

$$x(ht + t_0) - x(t_0) \stackrel{\text{def}}{=} h^H\{x(t + t_0) - x(t_0)\} \tag{3.5}$$

Without loss of generality, we assume $\mathrm{x(t_0)} = 0$, thus we have $x(ht + t_0) = x_h(t, t_0)$. That means the original time series is related to its renormalized version (Yamamoto, et al., 1993), consequently the discrete version of this relationship can be defined as:

$$X_h(i, i_0) = X(hi, i_0) \stackrel{\text{def}}{=} h^H X_1(i, i_0) \tag{3.6}$$

Where $X(i)$ is the discrete version of $x(t)$. The new time series $X_h(i, i_0)$ is called the "coarse grained" subset and the new sequence is formed by selecting every h sample from the original time series. An auto power spectrum and cross power spectrum from $S_{XX}(n)$ and $S_{XX_h}(n)$ then can be calculated as:

$$S_{XX}(n) = \frac{1}{N_{subset}} \sum_{i_0}^{N_{subset}} \left\| \frac{1}{N_{data}} . \sum_{k=0}^{N_{data}-1} X_1(k, i_0) . e^{-j2\pi kn / N_{data}} \right\| \tag{3.7}$$

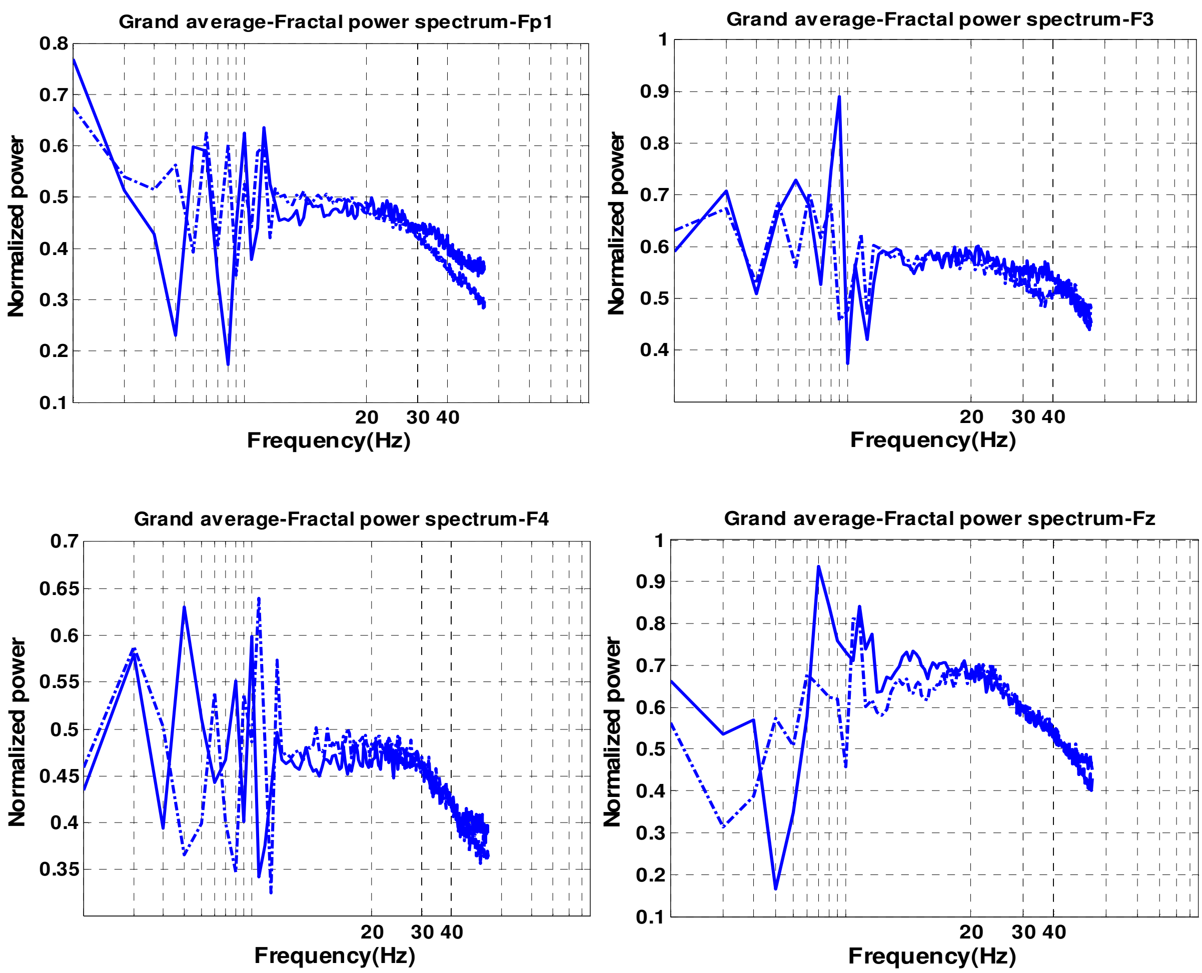
Where $n = 0, 1, \dots, N_{data} - 1$ and N_{subset} is the number of different i_0 chosen from a given time series.

$$S_{XX_h}(n) = \frac{1}{N_{subset}} \sum_{i_0}^{N_{subset}} \left[\begin{array}{c} \frac{1}{N_{data}} . \sum_{k=0}^{N_{data}-1} X_1(k, i_0) . e^{-j2\pi kn / N_{data}} \times \\ \left(\frac{1}{N_{data}} . \sum_{k=0}^{N_{data}-1} X_h(k, i_0) . e^{-j2\pi kn / N_{data}} \right)^* \end{array} \right] \tag{3.8}$$

In a simple harmonic signal the value of $S_{XX_h}(n)$ that is equivalent to Fourier transform of the cross correlation function between two orthogonal sinusoid, tended to be zero when $N\_data \to \infty$. On the other hand, it is proven that the corresponding value in a fractal motion defined in the equation (3.6) never goes to zero. Indeed it could be concluded that $\| S_{XX_h}(n)\| / h^H$ can be considered as a fractal component in the auto power spectrum without contribution of simple harmonic motions (Yamamoto, et al., 1993). For random fractals, the spectral exponent β is linked to H with the relationship of:

$$\beta = 2H + 1 \tag{3.9}$$

We extracted and measured the linear correlation between CD and β in both relaxation and imagination of movement states. Features computed in this fashion represent the fractal part of EEG signal, which is reported in Figure 3.15 for recorded channels. The spectrum shows some course changes in lower frequency bands. Especially there are noticeable differences between spectrum calculated over imagination and relaxation states in C3, Fz, Cz and Pz. The changes are distinguishable in lower frequencies corresponding to theta and alpha bands.



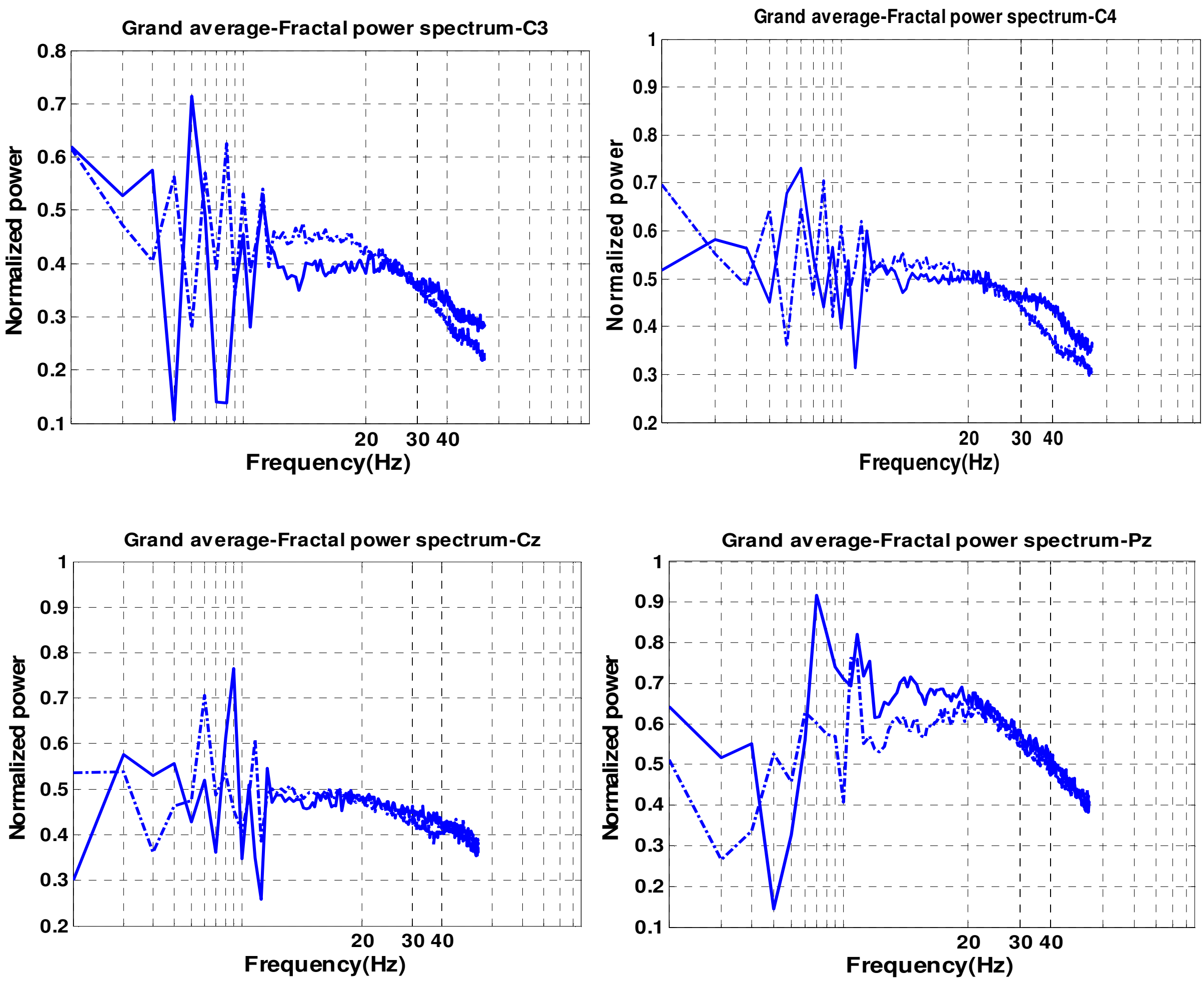


Figure 3.15: Grand average of EEG fractal power spectra obtained from CGSA in relaxed (solid) and imaginary (dashed) states.

An increase of power in Beta1 is observed in C3 during the movement imagination. In general fractal power spectrum shows similar fluctuations in comparison to the band power spectrum (Figure 3.8, above).

In this section by calculating fractal component (β) via CGSA from EEG signal, a new set of features for classification of two different brain states was introduced. In order to assess the quality of features, a machine leaning technique should be applied where the computer learns a decision function based on the training dataset. Soft Margin SVM was used to classify extracted features from the recorded EEG, distinguishing between two classes, motor imagination and relaxation.

3.5.3 Soft Margin Support Vector Machine

The main structure of a support vector machine was introduced earlier in chapter 2. A SVM q-norm soft margin classifier is used to handle the high dimensional classification problems where the data are not linearly separable (Steinwart et al., 2008). The difference between the standard approach and the soft margin method is that the latter allows making a few mistakes and placing some points inside or on the opposite side of the margin in order to remedy for noise, outliers or nonlinearity in the data structure. Mathematically the existence of misclassified samples should be paid by modifying the cost function by subtracting the distance between the real data point and the margin requirement, which can be seen in Equation 4.10. This can be implemented by adding slack variables ξ_i. The formulation of the SVM under the condition for the optimal hyper-plane can be modified by including an extra term:

$$y_i(X_i^T W + b) \geq 1 - \xi_i\,, \quad i = 1, \ldots, m \tag{3.10}$$

For minimum error, $\xi_i \geq 0$ should be minimized as well as $\|W\|$, and the objective function becomes:

$$\min \; W^T W + C \sum_{i=1}^{m} \xi_i^q \tag{3.11}$$

Subject to:

$$y_i(X_i^T W + b) \geq 1 - \xi_i\,, and \;\; \xi_i \geq 0; \quad i = 1, \ldots, m$$

Here C is a free but fixed parameter. It is called the regularization parameter and controls the trade-off between maximizing the margin and minimizing the training error (Hastie et al., 2008). The small value of C tends to emphasize the margin while ignoring the outliers in the training data, whereas large C may overfit the training data (Vapnik, 1999).
Note that the condition $\xi_i \geq 0$ is dropped, as if $n\xi_i < 0$, we can set it to zero and the objective function is further reduced. Alternatively, if we let k =1, the problem can be formulated as:

$$\min \; W^T W + C \sum_{i=1}^{m} \xi_i \tag{3.12}$$

Subject to:

$$y_i(X_i^T W + b) \geq 1 - \xi_i\,, and \;\; \xi_i \geq 0; \quad i = 1, \ldots, m$$

This is called the 1-norm soft margin problem. The algorithm based on the 1-norm setup, when compared to a higher-norm algorithm, is less sensitive to outliers in training data (Zhang, 2010). When the data is noisy, the 1-norm method should be used to ignore outliers. The support vectors are normally selected out of a small part of the training data. But under certain circumstances, e.g. for the nonlinear case or for problems which are non-separable or has a narrow margin, each non-zero data point within the margin which is misclassified should be added to the set and consequently the set of support vectors grows very fast. Large set of points will lead to the SVM slowdown during the test time. The generalization capability is directly related to the size of the margin. As shown in Figure 3.16 the positive and the negative samples are separated by a hyperplane illustrates. The points x, that lie on the hyperplane should satisfy $x^T\beta + \beta_0 = 0$. Here β is the vertical distance from the hyperplane to the origin, which is normalized to the hyperplane $\beta_0/\|\beta\|$ ($\|\beta\|$ is the Euclidean norm of β). When the problem is linearly separable, the algorithm simply searches for the separating hyperplane with the maximum margin. In fact, it is a convex optimization problem which under linear inequality constraints minimizes the quadratic function (Meyer et al., 2003) (Steinwart et al., 2008). Figure 3.16 shows the concept of slack variables, when the problem is not linearly separable. The maximum margin is set to $2M = 2/\|\beta\|$. The points on the wrong side of the margin are labeled ξ_j and are calculated as $\xi_j^* = M\xi_j$. The points on the correct side have $\xi_j^* = 0$. Thus the value $\sum \xi_j^*$ represents the total distance of all points on the wrong side (Steinwart et al., 2008). It measures the overlap in relative distance, which changes with the width of the margin M. By adjusting the parameters we can improve the quality of BCI systems.

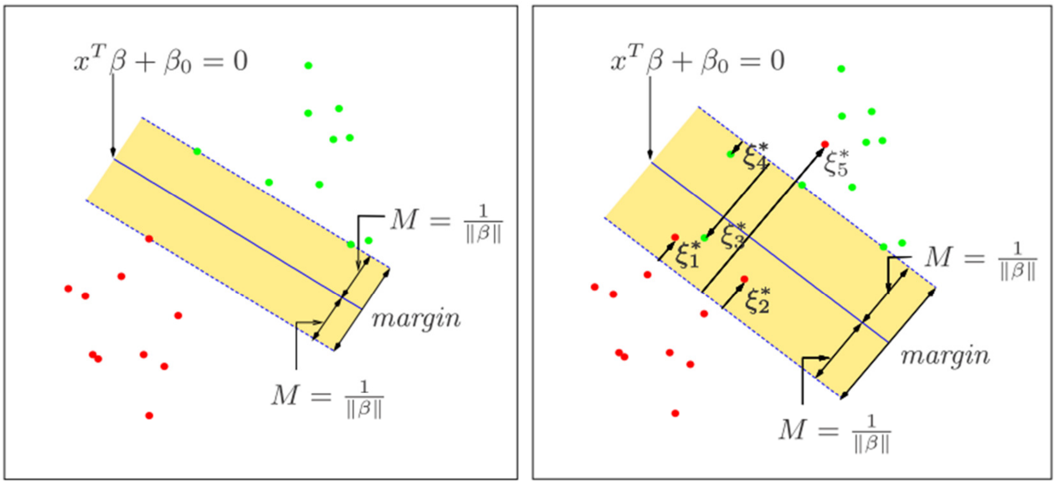


Figure 3.16: Support Vector Classifier: The solid line represents the decision boundary and the hashed lines represent the maximum margins $2M = {}^{2}/_{\|\beta\|}$. (a) In a separable case, all the trained dataset are classified correctly (b) The right graph shows the non-separable case and the concept of slack variables (adopted from (Hastie et al., 2008)).

3.5.4 Decision making: Bagging approach

At this section we apply the concept of ensemble classification, which was introduced in Chapter 2. Bagging (Bagg), shortly for bootstrap aggregating was employed for decision making section. It is an ensemble (jury) of classifiers. Random sampling with replacement is used to generate training sets for each classifier. Decisions are finally made by majority vote (Grandvalet, 2004). As described in Table III, the algorithm repeats in M iterations to train classifiers.

Table III - Bagging algorithm
1. for m= 1to M a) Select S_m bootstrap samples out of the data (with replacement) b) Train the classifier C_m from S_m
2. For each test example a) Try all classifiers C_m b) Predict the class that receives the highest number of votes

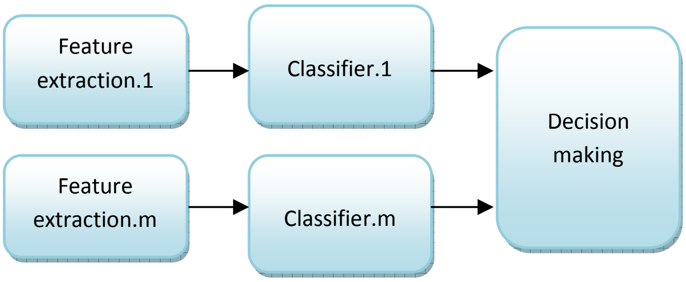


Figure 3.17: Decision making was based on the Bagging algorithm. A subset of feature set is used to train a classifier; the final decision is made with the majority of votes.

3.6. Statistical measures of performance

In this dissertation we used several statistical measures of the performance of a binary decision maker. True negative (TN) is when the test makes a negative result (relaxed state prediction) and the subject actually intended it. Similarly, true positive (TP) is when both test and the subject has positive results (here imagination of the hand movement). False positive is incorrect rejection of a positive result and false negative is the failure to reject a false result (Hastie et al., 2008). False positive (FP) and false negative (FN) are called error type I and II respectively (Hastie et al., 2008).

3.6.1 Classification accuracy

Accuracy is the most common evaluation measure in classification problems. It is defined as the sum of all correctly recognized samples divided by the total number of samples in an experiment. It is a general measure of how well a classifier can classify the samples. We used this measure in order to evaluate our algorithms.

$$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)} \tag{3.13}$$

3.7. Results

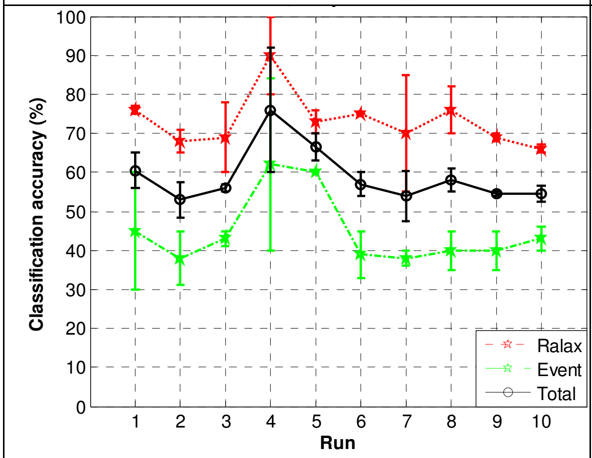
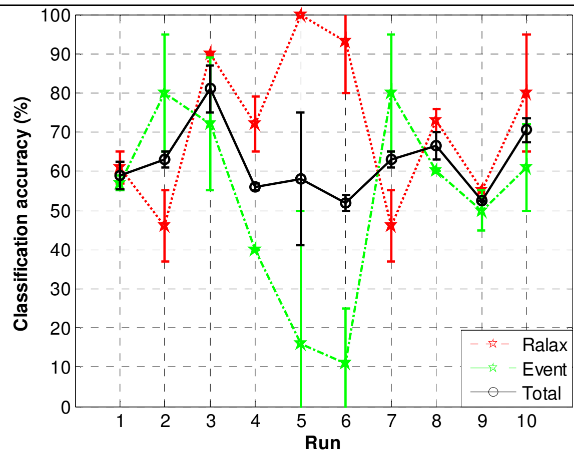
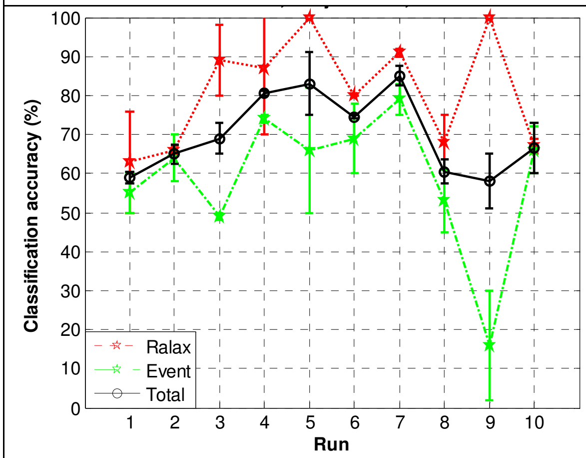
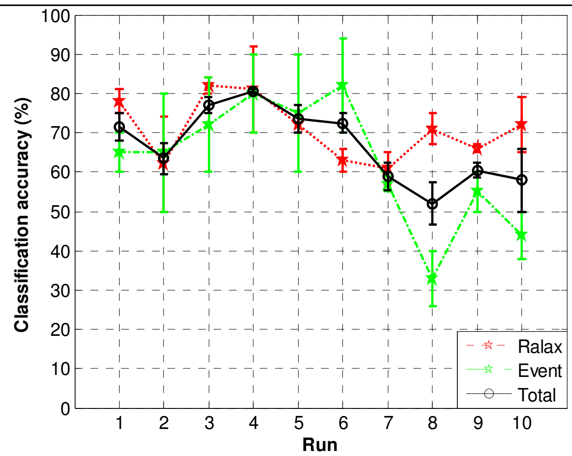
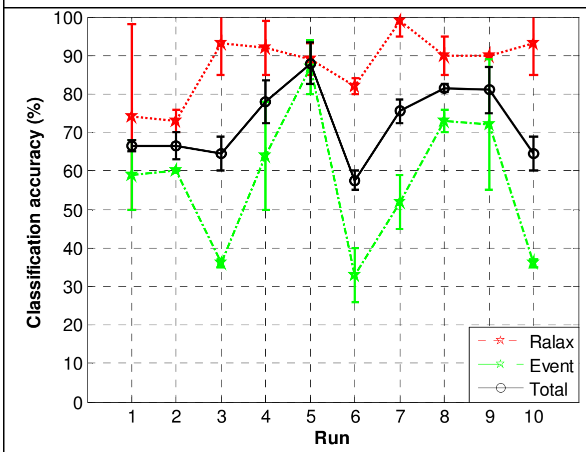
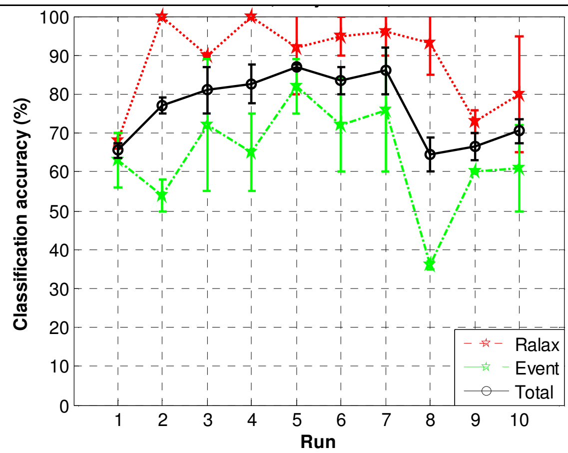
The mean values and standard deviation (SD) of the classification accuracy during different runs of the experiment averaged over all sessions are illustrated in Figure 3.17 for each individual. Results presented in Figure 3.17 were achieved using the simulated online scheme, where the classification was done for each single segment of the ongoing EEG.

Two different combinations of feature extraction and classification were applied on this dataset. The first group was comprised of 35 power band features from the spectral power of EEG signal in theta, alpha, lower beta, upper beta, and gamma frequency bands of 7 channels (all except Fp1) and the soft SVM classifier. The corresponding results are shown in the left column of Figure 3.17. The right column shows the results achieved by employing the second group of features (n=35) included fractal components calculated by the CGSA method and the same classifier. The black line reports the classification accuracy in average, which presents the general performance of each subject. The red and green lines depict the result for relaxation and movement trials, respectively.

For power-band features, in average the classifier for the majority of participants (all except S4) was more toward choosing the relaxation class. The relaxation accuracy (red line) is higher than movement accuracy (green line) for S1, S2, S3, and S5. For S6 only in the first five runs of the experiment relaxation accuracy is higher. Finally for S4, movement accuracy is higher all during the experiment. S4 was not very good in keeping himself relaxed during the experiments and his total performance was measured around random in this simulation. S3, S5 and S6 were able to increase their performance over the time. Despite of the increase in performance, S6 seems to be in average less relaxed in the last five runs of the experiment.

Assuming that the fractal features extract another kind of information from EEG signals, which are not available (or partly available) in power band features, the changing patterns are not completely similar for two groups.

Using fractal features, in general a fewer number of participants show the classification biased towards relaxation phase in comparison with power-band features. Moreover, both features results in similar changing pattern in some subjects but highly dissimilar in other subjects. For fractal features, relaxation accuracy (red line) is higher than movement accuracy (green line) for S3 and S6. For S5, movement accuracy is higher in the beginning of the experiment but decreases after run 3. Results from S1 and S2 are almost equally distributed between two classes over different runs. For S4 relaxation accuracy decreases in the first four runs and then starts to increase till the end of the experiment. In particular using the fractal features for S4 caused a decrease in his relaxation percentage in the middle of the experiment which led to slightly higher classification accuracy in general. The final decision was made based on the output of seven classifiers trained for each channel separately. In each session the classifier parameters were calculated using data recorded in the probe run. Thus, the entire structure including both proposed algorithms are highly nonlinear and the wide range of variations, seen in the results, are due to the combination of the multidimensional feature space with the nonlinear classifiers.

Power band features (a)	Fractal spectrum features (b)	
		**S1**
		**S2**
		**S3**

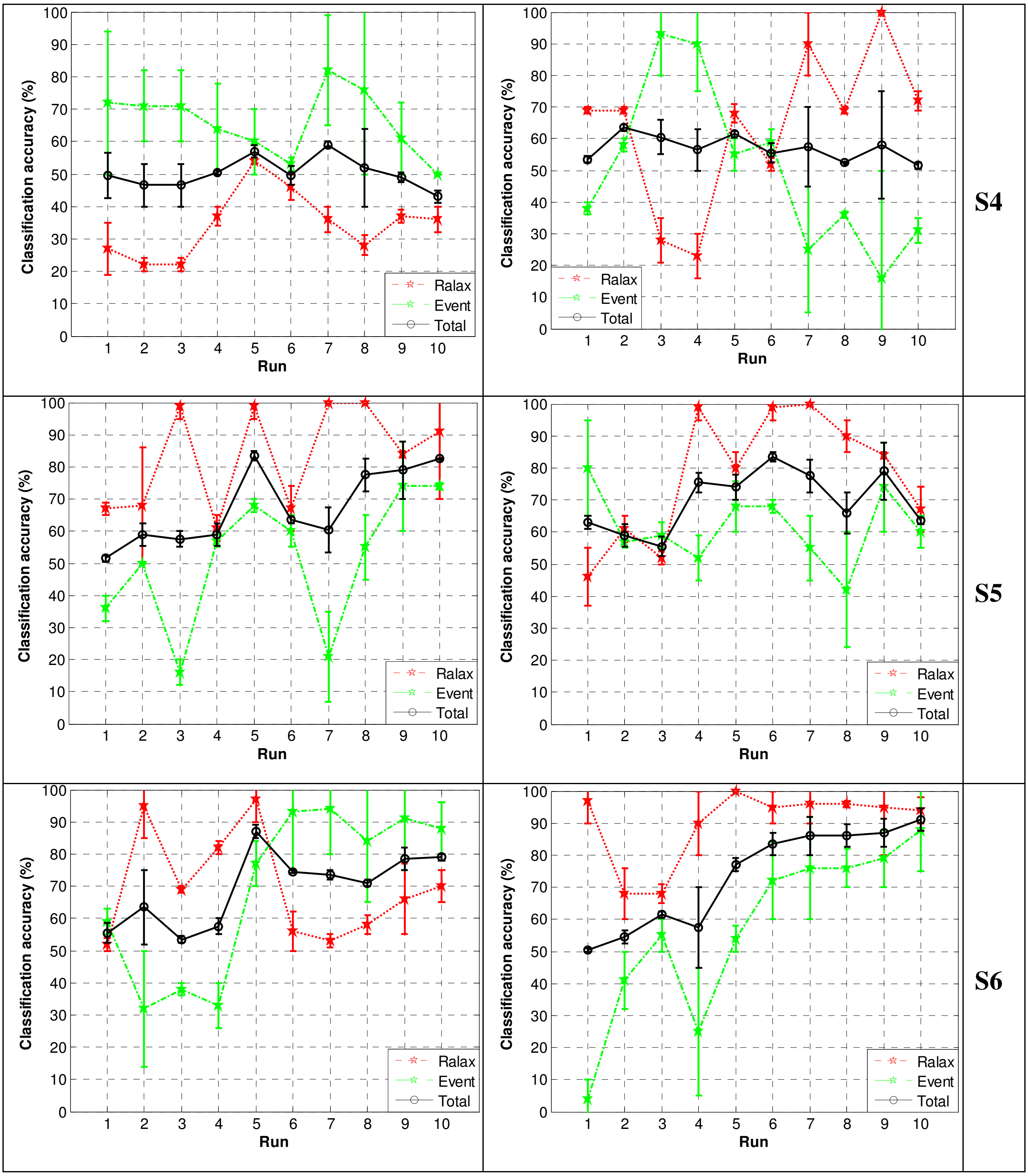


Figure 3.17: The averaged classification accuracy over different sessions for each subject in a simulated online paradigm using (a) power band and (b) fractal features and the 1-norm soft margin SVM as classifier

The above results were achieved in as simulated online paradigm. For calculating the classification accuracy in an offline scheme, as can be seen in Figure 3.18 and Table 3.2, we considered 80% of the data as training data for each subject and repeated the calculation over each training group of data following the cross validation scheme to verify the results. The grid search was done using a 5-fold cross-validation. For this purpose the dataset was randomly split into five subsets. The training and validation was carried out five times for each subset and the remaining four for training. Figure 3.18 depicts the average classification accuracy and its variance for each subject using power band and fractal power spectra as feature set and 1-norm soft margin support vector machine as classifier. Results were given as the average over all sessions for each subject. Here also an improvement is seen in the result achieved from data collected from S2, S3 and S6 after employing the fractal features. The average classification accuracy is low for S4 in both cases (<70%). Only in one case, S5, recruiting the new features did not yield some improvement on average; the variance was lower when using fractal features.

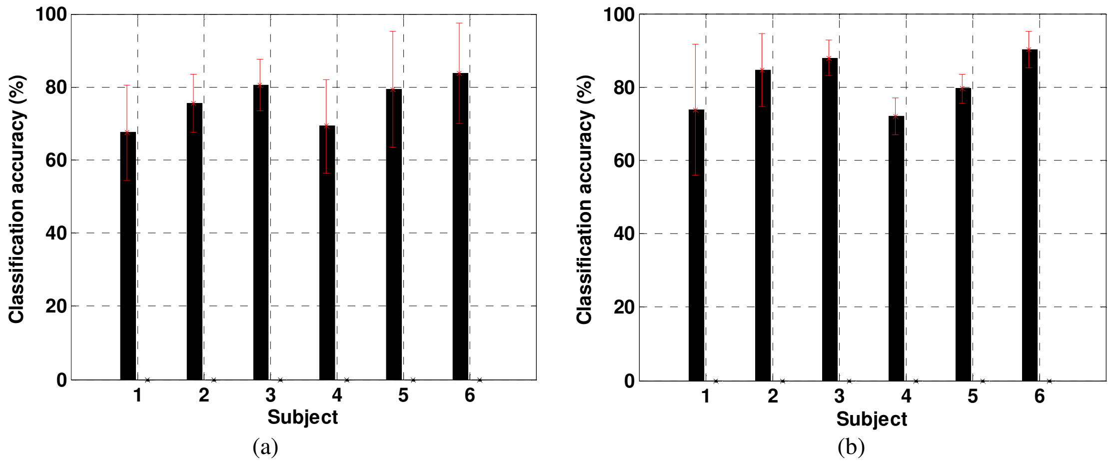


Figure 3.18: Average classification accuracy for each subject using power band (a) and fractal power spectra (b) as feature and 1-norm soft margin support vector machine as classifier

As reported in Table 3.2 the performance of the model is consistent for all subjects. In order to investigate the effect of the feature selection method, in both cases the classifier was kept fixed. Applying these components as features to a soft margin SVM classifier, enables us to evaluate how informative the fractal based attributes are.

Table 3.2: The average classification accuracy using three different feature spaces. In Average, the combination of two feature groups led to slightly higher classification accuracy.

Classification accuracy (%)	Power band Features	Fractal components	Combination of two Feature groups
S1	67.50	73.72	72.14
S2	75.48	84.59	85.90
S3	80.06	87.95	88.73
S4	69.33	71.90	73.01
S5	79.36	79.54	78.45
S6	83.72	90.01	91.32
Average	75.90	81.30	81.52

An average accuracy of 75.9% and 81.3 % in classification over the entire sessions for all six subjects was achieved using the band power and fractal features, respectively. Results show the performance in six subjects, corroborating the idea that the nature of this brain activity can be considered as chaotic.

In order to analyze the results in a statistical scheme we performed a matched pair t-test to investigate the overall statistical differences between two conditions, across all subjects (5 degree of freedom (df)). A statistical test, e.g. t-test, is usually applied to the sets of data in order to determine the degree of significantly difference that exists between them (Zhang, 2010). In this application we define these two sets as the feature space extracted from relaxed state and imagination state, respectively. Commands *[h, p]= ttest(x)* in MATLAB were used to calculate the p-value of the matched pairs test and the value of *h* which determines the acceptance or rejection of the null hypothesis. The result of the test is returned the probability of p, which is indeed the probability of observation of extreme values in the test under the null hypothesis (Parasuraman, 2007). The t-test performs a null hypothesis at the 5% significant level that data in the vector *x* are random samples from a normal distribution with zero mean and unknown variance, against the alternative that the mean is not zero.

The averaged classification accuracy was lower in imagination than in relaxation states using power band features (p =.041) but not in fractal based features (p = .23) with df = 5 and the t-values 2.72 and 1.34, respectively. For the combined features space t-test results in not statistically significant bias toward any of two classes; p = .35 with t-values = 1.02.

Combined accuracy over the two states, as compiled in Table 3.2, was higher for the fractal based features than for the power band features (t = 3.9, p = .011). As Table 3.2 shows, the combination of fractal and power features yielded slightly higher accuracy on average than the fractal features alone. However this difference was far from significant (t = 0.57, p = .58).

Thus, the experimental results show that the extracted fractal components using GCSA are suitable discriminators of EEG signals. They are able to extract significant differences between two predefined brain states for all subjects.

3.8. Discussion

It can be assumed that the complexity or predictability of the EEG signal differs in different brain states including before and during the performance of a mental task (e.g., motor imagery) (Phothisonothai, et al., 2007). The change of fractal attributes has been discussed in several studies like cognitive tasks, sleep, and different types of diseases as well as mental states. For instance in (Georgiev, et al., 2009), a higher dimensionality, i.e. complexity, of imaginary EEG states was recorded as compared to actual perceptual processing. However in most cases, biological signals do not consist solely of fractal dynamics. They include well-defined harmonic oscillations as well (Yamamoto, et al., 1993) (Pereda, et al., 1998) (Adlakha, 2002) discussed that in the natural signals the value of index β is very similar in linearly correlated noise and chaotic systems. Both CD and LLE have been demonstrated to act as poor indices to discern between different mental tasks (Theiler, et al., 1996). In the frequency domain, these oscillations are recognized as relatively sharp peaks in the power spectra although the peaks are usually superimposed on some type of noise spectrum which has been thought to reflect the underlying fractal dynamics (Kobayashi, et al., 1982). The capability of using fractal based features in a BCI system has been proposed in (Adlakha, 2002).

Considering this fact, we employed a novel method called Coarse Graining Spectral Analysis (CGSA) (Yamamoto, et al., 1993) to calculate random fractal components in the frequency domain of human EEG signals set to two different brain states. The method is capable of separating simple harmonic and fractal components from each other in the frequency domain (Yamamoto, et al., 1993). An EEG-based Brain Computer Interface could benefit from this remarkable property. By analyzing the EEG signal, we investigated the applicability of extracted features using CGSA in distinguishing the predefined brain states.

The fractal components extracted from the logarithmic power spectrum have been used as an index of the complexity of the signal, demonstrating the non-linear nature of EEG signal or the non linear dynamic of the brain. An application of chaos and fractal properties, which are one of the essential tools in nonlinear analysis, has been presented in analyzing two different brain states including idle state and imagination of hand movement in the human EEG. We extracted and measure the linear correlation between CD and β in these two states. However due to its lower computational cost, use of β is more appropriate in online applications. In order to investigate the demand of using real EEG data, we extracted the fractal components in the frequency domain and compared the effects of imagination of motor activation tasks of the human EEG compared to relaxed state, where a lucid frequency power dependency is visible and fractal components appeared to be lower in movement imagination compared to the baseline condition. In general it is concluded that fractal analysis could be a powerful method in investigating brain activities during motor movements and is a promising option for the design of BCI systems.

In a multidimensional feature space we compared the classification accuracy over different combinations of feature groups including the power band features, fractal components and their combination. With combining the extracted features with common oscillatory features in a high dimensional space, we tried to mine maximum information related to brain state hidden in EEG signal. The combined features did not lead to the significant higher classification accuracy in comparison to the power band features or the fractal features alone. A possible speculation is that, as the grand averages in the Figure 3.8 and Figure 3.15 suggest, the two feature sets have common characteristics or redundant information which makes them more similar to each other in a higher dimensional space than expected. This is a serious issue in each machine learning problem. Simply adding new features does not lead to the better classification results. We applied the concept of features selection in the structure of the next BCI applications in order to avoid this inconsistency. By applying a feature selection techniques the best m features under the criterion of minimum mutual information are selected, which in principle results in higher classification accuracy in lieu of smaller feature space.

Looking at the current results a possible reason why the classifier accuracy was lower during the imagination phase than in the relaxation phase could be the overall concentration of subjects and their tendency towards relaxation state during the experiment. The reason could be also sought in the weaker classifier performance during the imagination phase. SVM classifier is an optimum margin classifier. In our soft SVM classifier the concept of slack variables was used, where a limited number of misclassification during the training is allowed. It may lead to a slight imbalance behavior of the classifier in multi-dimensional space. As discussed in Chapter 2, in theory there is an unavoidable trade-off between bias and variance of the classifier. High bias and low variance classifiers are usually the stable ones, but not necessary leads to the highest classification accuracy over test data. Here a nonlinear classifier has been trained and tested over feature space. It is interesting that the results achieved using the combined feature spaces were less biased toward any of two classes. Questions of robustness of the method and data requirements could be yet discussed.

Chapter 4

Controlling a hand prosthesis using brain signals – A simulation study[1]

4.1. Introduction

With the advancement of technology neuroprosthetics can become an applicable solution for paralyzed people. The design of neural-machine interfaces for prosthesis, however, is limited in academia due to the shortage of affordable robotic hand availability and the overall accuracy of the system. Commercial hands are able to mimic the human hand with a good resolution but they are bulky, large and expensive (Oung, et al., 2012). "At the University of Lübeck, institute for signal processing (ISIP) researchers developed a new low-cost anthropomorphic robotic hand. The project is envisioned to be a new cost-effective robotic hand that can initially serve as a research tool for neuroprosthetic engineers. With the ability to be fully as dexterous as the human hand, it let us effectively develop new and innovative interfaces that will allow users to control a robotic hand intuitively"[2]. We applied such a concept to prosthetic control, which could handle many tasks autonomously, only requiring high-level commands from the user. This chapter, thereby, presents an application based on motor imaginary BCI systems with artificial hand control for simulated hold, grasp and open a hand using a previously collected dataset. The focus of this project was to establish the robust hardware connection between the BCI and the already developed anthropomorphic robotic hand and at the same time to investigate a new and reliable machine learning approach for this setup. In order to reveal non-Gaussian characteristics of the EEG signal we applied higher order statistics momentums along with power band features to create

[1] Part of this work was published in : **M. Kh. Hazrati**, and A. Erfanian, " An On-line BCI Without Training for Controlling the Sequence of Hand Grasping, Opening, and Holding Using Adaptive Probabilistic Neural Network ," *Journal of Medical Engineering and Physics*, April 2010.

[2] Stephen Oung, Design and Development of a tendon-based anthropomorphic robotic hand", Master thesis, 2012

feature space for this application. Considering the nonlinear and chaotic behavior of brain signals, we hypothesize that these features are able to distinguish between two states of the brain which are of our interest in this application. To my knowledge using the proposed parameters were not investigated in any online BCI application so far.

4.2. Project specifications

4.2.1. Robotic hand configurations

Figure 4.1 shows Lübeck's anthropomorphically designed robotic hand (LAnDRoH). The main motivation to design LAnDRoH was to provide of a cost-effective, fully dexterous, and lightweight robotic hand (Oung et al., 2012). The entire hand was bolted onto a solid base to prevent shifting of the hand during flexion. The initial work, shown in Figure 4.1, was focused on the electromechanical design and control. Thus the preliminary test bed was designed to perform initial measurements of the control algorithms. The robotic hand implemented an *N+1* flexible link actuation system that allowed independent control of each of the 20 degrees of freedom. The system was controlled using an ATmega2560 microcontroller (SmartProjects Italy), which allowed control from either a haptic glove or a computer. An adult human hand only weighs 530g but is capable of exerting a flexion speed of 40 rad/s and a grip grasp with an exertion force of 540N (Oung, et al., 2012). Although the maximum speed is remarkably high, studies have shown that for performing the majority of activities of daily living (ADLs) the human hand only requires speeds of 3-4 rad/s (Bundhoo et al., 2005). Higher speeds are only required for specialized tasks such as catching a thrown object. The average speeds of the robotic fingers for extension and flexion are comparable to current robotic hands and prostheses (Bundhoo et al., 2005).

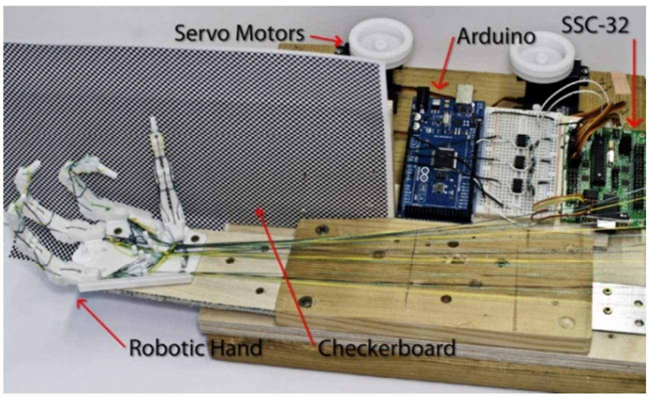


Figure 4.1: (a) The main motivation to setup the Lübeck's anthropomorphically designed robotic hand (LAnDRoH) was to provide of a cost-effective, fully dexterous, and lightweight robotic hand. The entire hand was bolted onto a solid base to prevent shifting of the hand during flexion. The average time to fully flex a finger was recorded at 0.64 s (˜ 2.8 rad/s).

The average time to fully flex a finger was recorded at 0.64 s. This translates to an average flexion speed of 2.8 rad/s. Although this is only a fraction of the speed capable of the human hand, it is comparable to current robotic hands and prostheses. The construction details can be found in Oung's master thesis.

4.2.2. Lübeck BCI design

The Lübeck brain computer interface (LBCI) was designed based on a MATLAB graphical interface connected to the robotic hand and with the option of connecting to a g.USBamp (Guger Technologies, Graz, Austria). The connection with robot was established using a direct SSC-32 connection and a serial object from MATLAB and Real-Time Windows Target for Windows XP. An initialization command was sent to the robot hand to set its status to fully open before the start of the experiment. Written m-file function *Fingermove* tests the connection and set the joint angle by sending the predefined speed and direction values to servomotors (GWS, Taiwan) with maximum speed 360°/1.28 s. The position is controlled by a pulse width modulation (PWM). The initial home position is determined with a PWM of 1500 μs and a full rotation is given by an increase or decrease of 1000 μs (Oung et al., 2012).

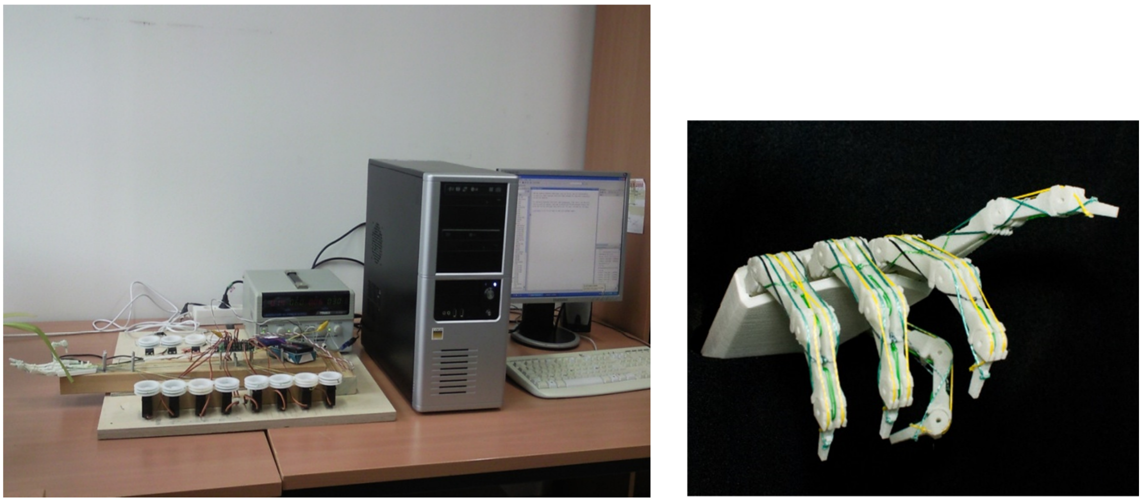


Figure 4.2: (a) Experimental setup for hand prosthesis control. BCI paradigm consisted of a GUI interface to demonstrate the virtual hand and a link to the robotic hand. (b) 3D rendering of Robotic Hand, each finger requires a specific control command.

The interface consists of both offline and online modes. In the current configuration a separate command will be sent to each finger leading to one step "Grasp", "Hold" or "Release" actions. I modified the BCI set-up described in the previous chapter (Hazrati et al., 2011) and linked it to the LAnDRoH (Oung et al., 2012). In the following we investigate how a BCI can use the same framework as in the "virtual hand movement" and benefits from the hardware connection established between the artificial hand custom while both interactive virtual reality environment and LAnDRoH are controlled by advanced machine learning algorithms. The system is able to read the data in a real-time scheme from a data set previously recorded. It means that in our current evaluation system a data pool was used instead of an actual recording system. The result of the classification was fed to the robot system in order to control the artificial hand and simultaneously a feedback was displayed on the screen. In the following the primary experimental setup and the implementation of the entire system are explained.

4.2.3. Data Collection and Primary experimental paradigm

Data collection experiments were conducted in the BCI Laboratory, Neural Technology Center (NTC), Departments of Electronic Engineering at Iran University of science and technology (IUST) in Tehran. Ten healthy volunteer subjects (Five male, Five female, mean age: 27.6) participated in the experiments of this study. All subjects except S5 were right handed. The subjects were naïve without any previous experience for BCI experiments. Monopolar EEG signals were recorded by means of a g.USBamp at the sampling rate of 256 (samples/s) from positions F3, F4, Fz, C3, C4, Cz, and Pz by Ag/AgCl scalp electrodes mounted according to the 10-20 system and then were low-pass filtered with a cutoff of 45 Hz using Butterworth filter order 10. During the experiments impedances were kept below 5 kΩ. Each epoch of 3840 points comprised 26 Hanning-windowed, 512-point data windows with 75% overlap. An electrode was placed above the left eyebrow line to record the eye blinks. All channels were recorded with the reference electrode located on the right earlobe and a ground electrode at the left earlobe. A resting period of about 5 minutes was enforced between each run. The experiment for each individual consisted of different sessions and each session was conducted on a different day. Each session consisted of at least 10 runs and each run consisted of 10 trials with online feedback.

To classify two states of brain activity, relaxation and imagination of hand movement (close and open), a scenario containing those states for data collection was designed. In this BCI paradigm subjects tried to control their attention and follow the instruction for the imagination of predefined hand movement and relaxation modes. At the start of trial, a blank screen was shown to the subject for 1 s. Then the subject observed an opened hand on the screen which indicates the onset of the relaxation phase. In this period the subject should not perform any specific mental task but to try keeping the hand open for 5 s. Following the relaxation phase, as displayed in Figure 4.3, the hand re-opened again (in case it was partially closed before) a ball began to fall and by reaching the virtual palm the second phase of the trial started. This transient phase was fixed and consisted of four images displaying every 0.5 second (total of 2 s). This was considered as a cue to convey the message of starting the imagination. At 7 s after onset of the trial, an active feedback phase was started. The entire feedback phase lasted 10 s. In the first 5 s, the user was asked to imagine of hand grasp action by grasping the ball (i.e., closing phase). Independent of the subject's success in this part a closed hand holding a green ball appeared on the screen at 5 s and cued the subject to imagine the opening hand. During the next 5 s interval which represented the last 5 s of the entire trial the subject was asked to open the hand by corresponding imagination of hand opening. Starting from the point when the red ball appeared on the screen till the end of imagination trial was treated as a single class during the recording experiments; meaning that a single binary classifier was responsible to distinguish between the relaxed and movement classes.

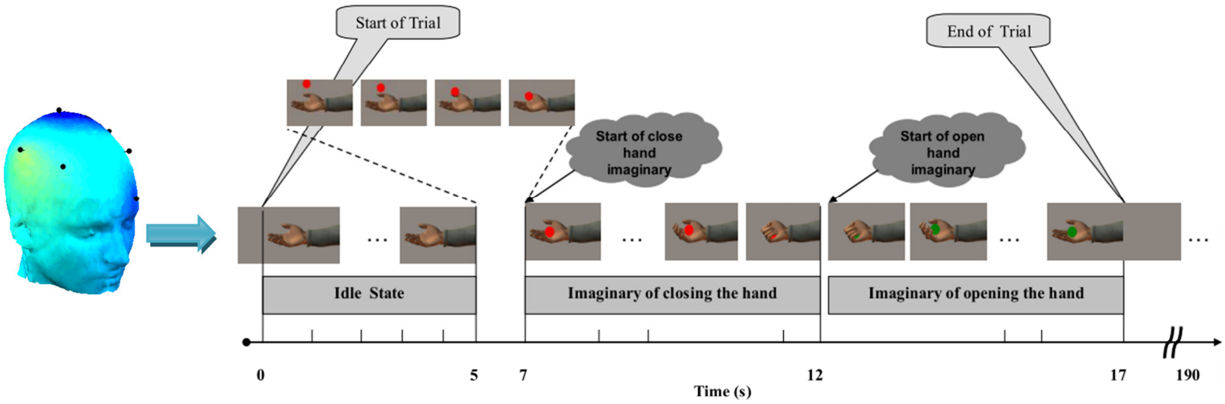


Figure 4.3: Experimental set up for virtual hand grasp and open control. Ten subjects participated in online sessions. Each trial lasted 17s and consisted of 5s idle state, 2s preparation for imagination followed by 5s of imagination of hand grasp and 5s of imagination of opening the hand.

Data recorded during each run of experiment was saved automatically at the end of the experiment in separate MATLAB files in format (*.mat) containing variables:

OnlineData =

- time: [43560x1 double]
- signals: [1x8 struct]
- blockName: [1x58 char]

Using data collected in these series of experiments, we extended our previous BCI paradigm to an EEG-based BCI for online control of hand grasping and holding and opening in a virtual reality environment as well as simultaneously by the robot hand. In this study we defined a binary problem for classifying between the relaxation and the imagination of movement (both closing and opening as one class). Therefore results are presented for 5 s of the relaxation and 10 s of the imagination of movement intervals.

4.3. Evaluation of the recorded EEG data

4.3.1. Averaged potentials

Similar to the previous chapter signal averaging was used to enhance the signal-to-noise ratio. Grand-mean EEG waveforms (i.e., mean over trials and over participants) for this dataset are presented in Figure 4.4 for the relaxation and imagination phases. Here we consider the entire trial which last 17 s, starting from the onset of relaxation trial at -7 s. The time point -2 s is when the red ball appears on the screen and 2 s later at the zero point the imagination phase for closing the hand starts. In Figure 4.4 units on x-axis are ms and unit on y-axis are μV. The first 5 s of recording (relaxation phase) was considered as the baseline. Looking at the data after baseline removal reveals the fact that the imagination phase has lower average in comparison to the relaxed phase.

The first evoked potential happened at -6800 ms with negative peak followed by a positive peak at -6700 ms maximum at Pz. This large positive peak is most probably a P300 component according to its timing and topography. It decreases from parietal to frontal lobes.

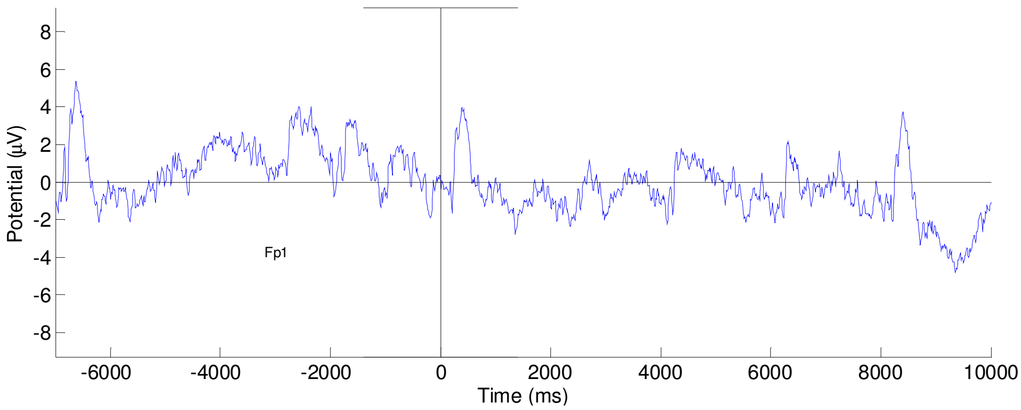
In the middle of the relaxed phase at -4500 ms regular fluctuations, possibly due to evoked activity, can be seen over F3 and F4. This fluctuation exists rather weaker in the other recoding locations.

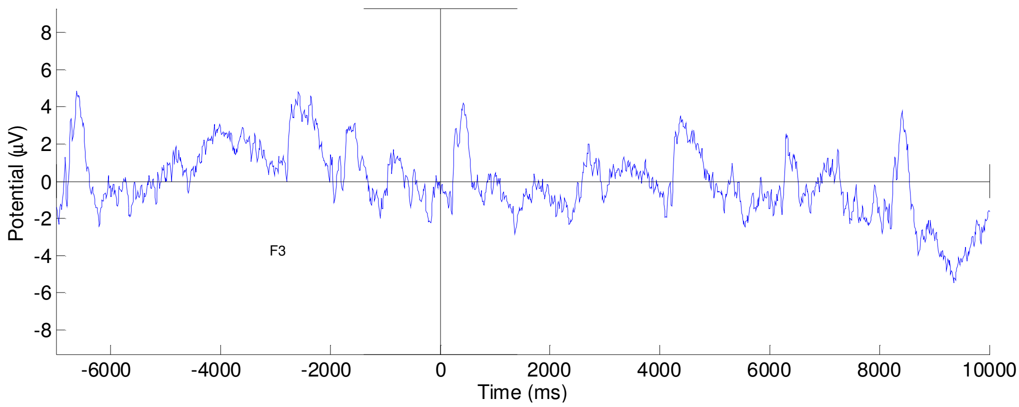
Another positive peak can be seen at -2400 ms close to the end of the relaxation phase (which is at -2000 ms) maximum at F4. The reason for it is not clear and can be the results of an artifact, eye movement, with stronger pattern over the frontal lobe. This is followed by a positive peak -1700 ms maximum at Pz, which is the evoked potential of the event at -2000 ms and probably a P300 (happened 300 ms after the red ball appears). The maximum amplitude decreases from parietal to frontal sites.

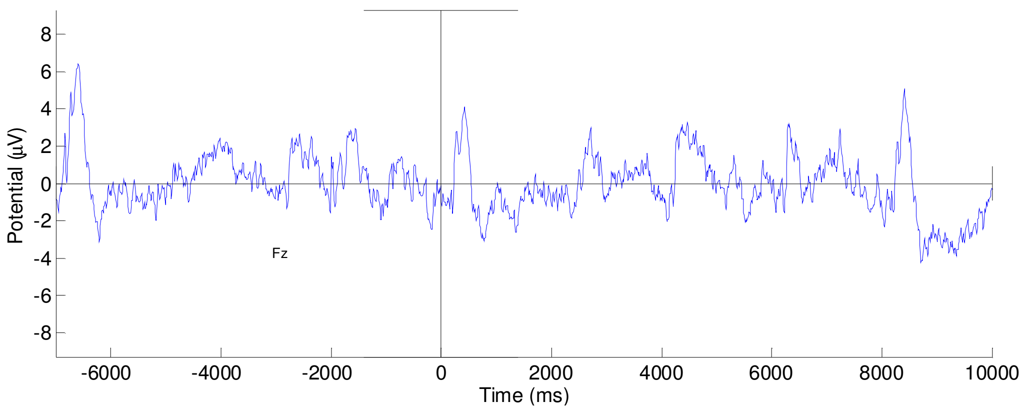
After the start of the imagination phase at 200 ms, the negative peak of the evoked potential, maximum at Pz and C4, is followed, especially at Pz and Cz, by another positive peak after 500 ms, most probably a P300 component according to its timing and topography. It can be seen over all recoding sites and has its maximum at Pz.

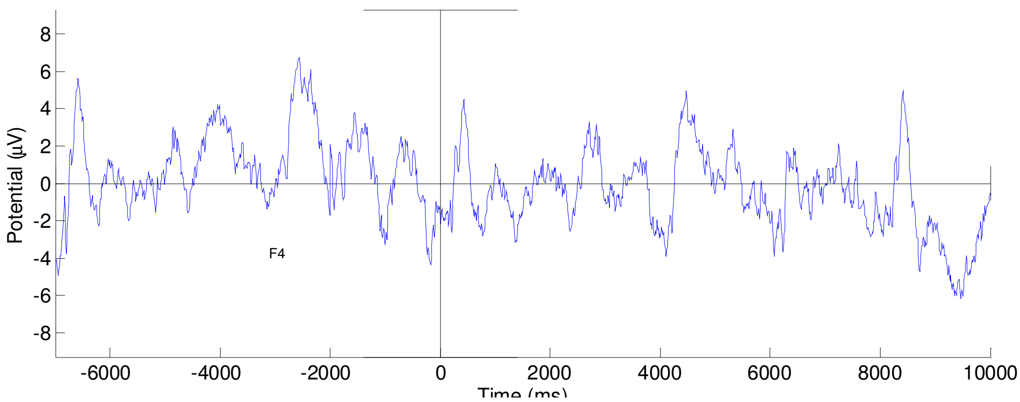
At 4300 ms another evoked potential exists which maximizes at F4 and reaches its peak somewhat later at Fz and F3, at 4520 ms.

Two late deflections, consisting of peaks at about 6300 ms and 8400 ms, maximum at Pz and C4, can be observed during the imagination of opening hand phase. These are probably positive peaks of two evoked potentials at 6 s and 8 s from the biofeedback.


Potential (µV)
Time (ms)
Fp1


Potential (µV)
Time (ms)
F3


Potential (µV)
Time (ms)
Fz


Potential (µV)
F4

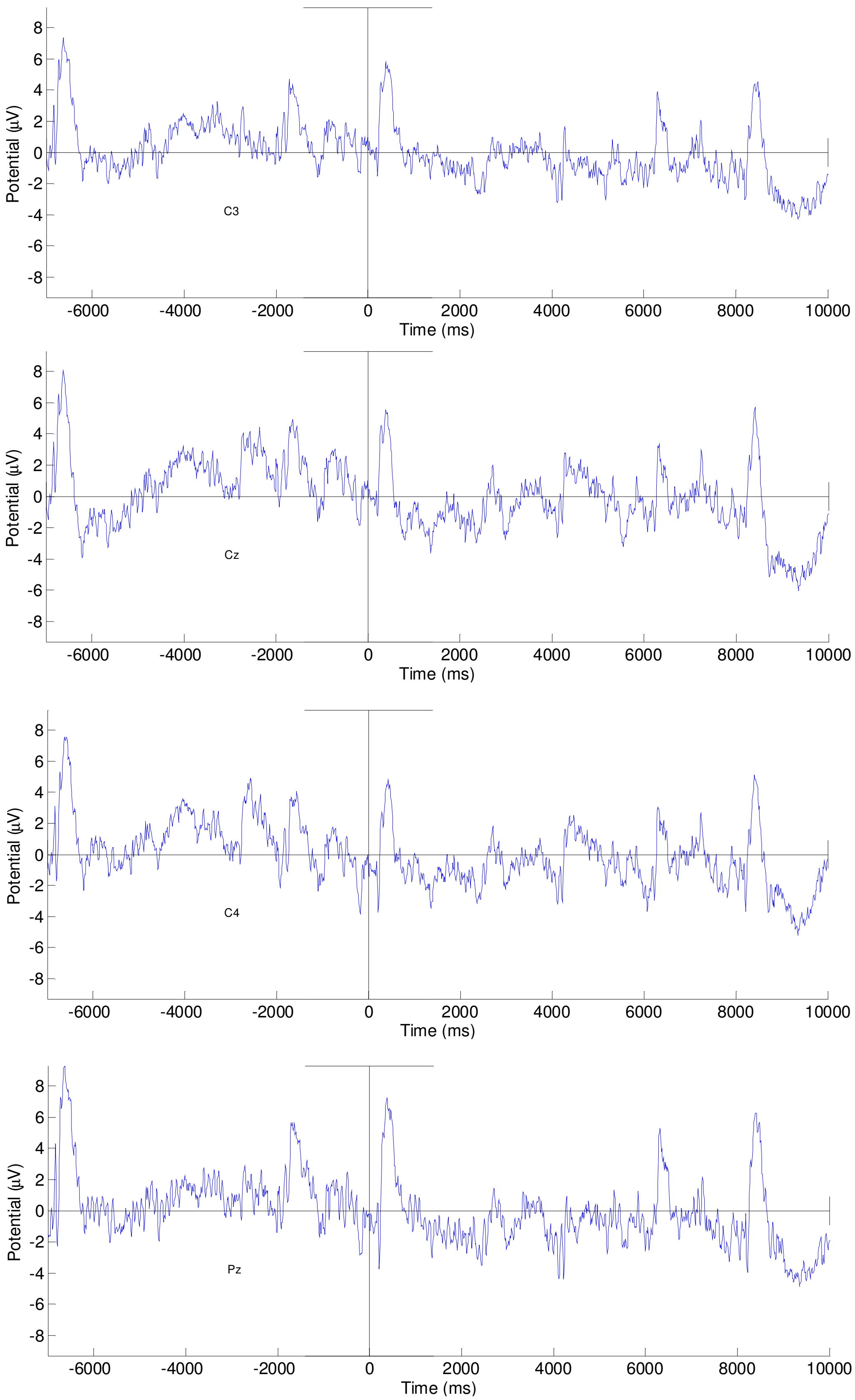


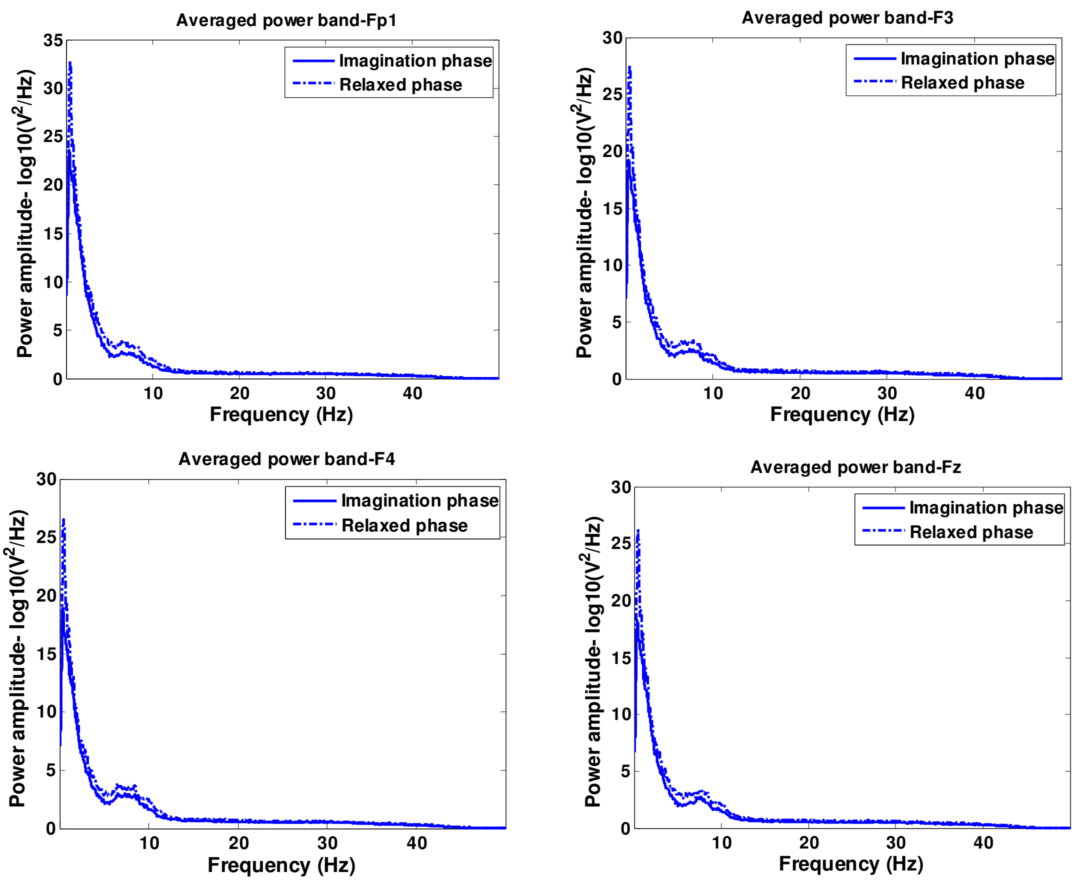
Figure 4.4: Grand mean waveforms for all 10 participants at all electrode locations (positions Fp1, F3, F4, Fz, C3, C4, Cz and Pz). y-axis range from -9 µV (bottom) to +9 µV (top).). At -7000 ms, the relaxation phase started when the picture of an empty hand appeared on the screen. After 5 s, at -2000 ms, the transient phase started when the red ball appeared on the screen. Imagination of the closing hand commenced at 0 ms and after 5 s the imagination of the opening hand started (at +5000 ms).

4.3.2. Power band spectrum

The average power band spectrum was depicted for the imagination and relaxation signals acquired from each individual channel in Figure 4.5. The amplitude is calculated in log10 (V^2/Hz) unit. The calculation is similar to the one described in 3.5.

In the current dataset values of the EEG power in delta (1-3 Hz) and alpha bands (8-12 Hz) of relaxation phase were found to be noticeably higher than those in imagination phase over all recording sites. The alpha peak has tendency to the lower border at Fp1 and F3 and the peak is seen in the lower frequency (around 8 Hz) in the data corresponding to relaxation in comparison to the imagination. The maximum exists at 10 Hz at Pz, as expected.

For theta band (4-8 Hz) a wide range ERD is seen during the imagination from recording sites overlying the primary hand-motor cortex (M1) (at C3 and C4) as well as at parietal site (Pz). The theta ERD topography changes over channels. At C3, C4 and Cz it differs in a wider range of frequencies between the two conditions. There is a slight decrease in the power of beta band in imagination phase over C3, C4 and Pz. No distinguishable changes can be seen in the gamma band (30-45 Hz) in average. However, it could be seen sparsely in some individuals for some session of experiments.



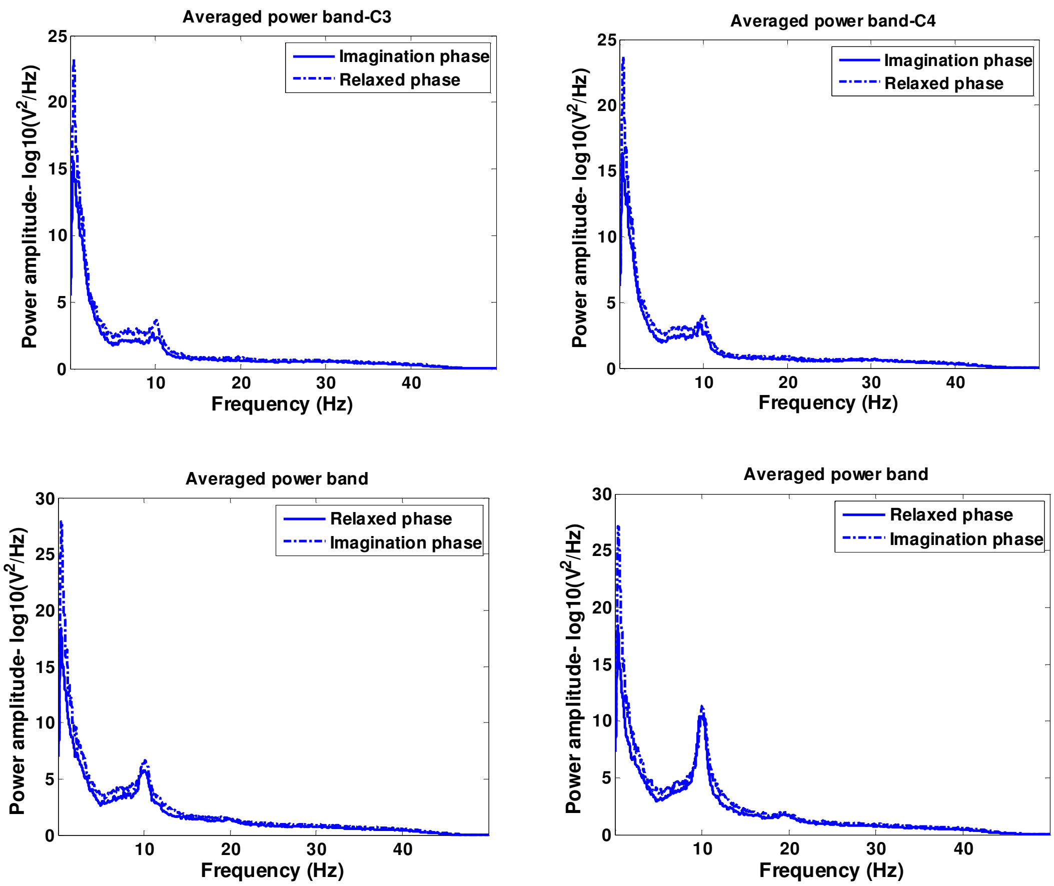


Figure 4.5: Average power band over the entire dataset for two different brain states demonstrated over Fp1, F3, F4, Fz, C3, C4, Cz and Pz - Unit on x-axis is Hz, unit on y-axis is V^2/Hz.

4.3.3. Time-Frequency analysis employing ERSP procedure

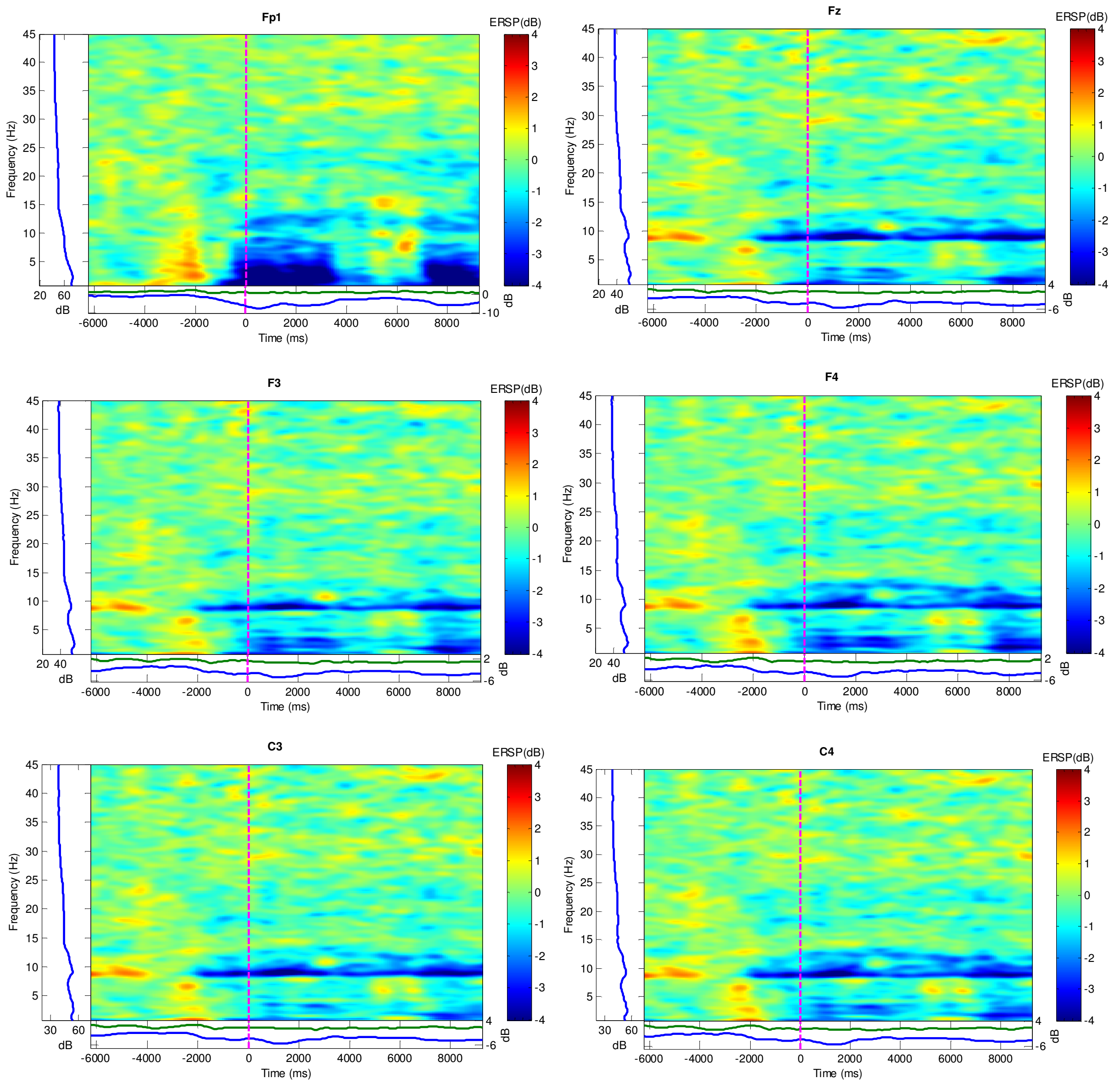
Figure 4.6 depicts ERSP analysis for all recorded electrodes averaged over all trials and subjects for this BCI experiment. Unit on x-axis is seconds. It starts from the onset of the relaxation phase at -7 s till the end of the imagination phase at 10 s. Unit on y-axis is Hz and varies between 0 to 45 Hz. The first 7 s of data is considered as baseline for the rest of the trial. The first and last one second of data was truncated because of processing distortions.

The first time-frequency fluctuation in Figure 4.6 is an ERS in alpha and upper Beta bands that happens between -6 s to -4 s almost continuously in Pz and over central but sparsely over frontal sites. It cannot be observed in Fp1. Sporadic Beta ERS starts later at -5 s. Delta and theta ERS can be seen between -3 s and -1.5 s, when the falling red ball appeared on the screen.

The most distinguishable fluctuation in Figure 4.6 is an alpha ERD that starts from -1 s and lasts till the end of the entire trial, all over the hand movement imagination phase.

Sparse ERD in beta band exists simultaneously over all recording sites. At +3 s an increase of power around 12 Hz can be observed maximum at Pz. There is also a theta ERS between 5 s and 7 s all over channels with maximum at Fp1. It can be the result of switching between opening and closing hand segments. A gamma ERS can also be observed in C3, Cz, F3 and Fz between 6 s and 8 s. It occurred after the start of imagination of opening hand at +5 s.

The global ERS in alpha band, as seen in Figure 4.6 in the previous section, is the result of absolute value of the alpha band. However, by applying ERSP method we can observe ERSs and ERPs for each frequency proportional to the baselines. Blue and green curves below ERSP diagrams are averaged values of ERSP over frequency with and without considering the baseline corrections, respectively.



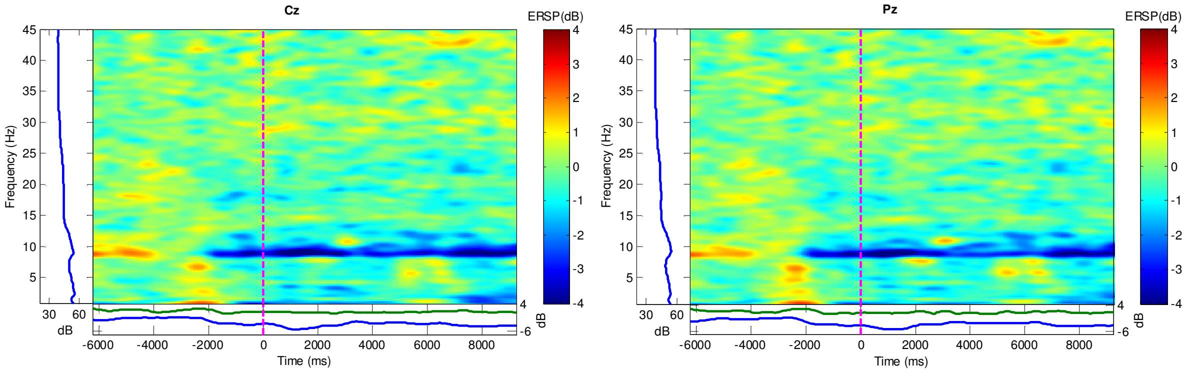


Figure 4.6: Time-Frequency analysis based on ERSP during relaxation and imagination phases. Zero point indicates the start of hand movement imagination by the subject. Unit on x-axis is s, unit on y-axis is Hz. The color bar represents the absolute values of ERSP in dB. Left and below diagrams are averaged values of ERSP, over time and frequency, respectively. Green line corresponds to the values before subtracting the baseline. Plotted using EEGLab toolbox in MATLAB.

4.4. Materials and Methods of BCI

The main structure of the current BCI consists of five major elements: Preprocessing (artifact reduction and filtering), feature extraction, feature selection, classification and decision making. The recorded EEG signal was treated as a simulated online scenario. The ongoing EEG in each trial of the experiment is processed in two-second windows, overlapping with 75% overlap. In the first step an artifact correction and filtering technique was applied on this window. Here we propose a novel information-based adaptive filter for the artifact reduction. A similarity index is introduced to evaluate the method and to choose the optimum parameters for the online setup. Then a feature vector was calculated using normalized power band features along with higher order statistic features. A feature selection algorithm was then used in order to select the more informative features for each channel. In this simulation study we compared two classification approaches. In the first method all selected features were fed to a single SVM classifier and the binary decision was made based on the output of this classifier. In the second approach again one single SVM classifier was used, but separately for each channel, and the final decision was made based on the voting technique. The classification accuracy was calculated regarding the number of correct binary outputs of the ensemble classifier. In the following I will explain each element of BCI thoroughly. At the end the result are presented and discussed.

4.4.1. Online Ocular Artifact Removal by Information-Based Adaptive Filtering

In order to implement an artifact removal system online, we used an adaptive filter approach in the time domain. We applied this method as an alternative to an ICA-based approach, expecting that its computation would be considerably faster, thereby qualifying as a suitable alternative for online schemes.

As depicted in Figure 4.7 an adaptive system requires two inputs, one being the reference input and the other being the desired signal. In ocular artifact removal from EEG, the desired signal input is the EEG recorded from the required site. The signal recorded from the Fp1 electrode is used as the reference input. In the current setup, it is the closest electrode to the eye and with the left-ear reference it mainly captures the ocular artifacts related to vertical eye movements. The reference input is low-pass filtered by passing through a moving average filter and then is applied to the adaptive filter.

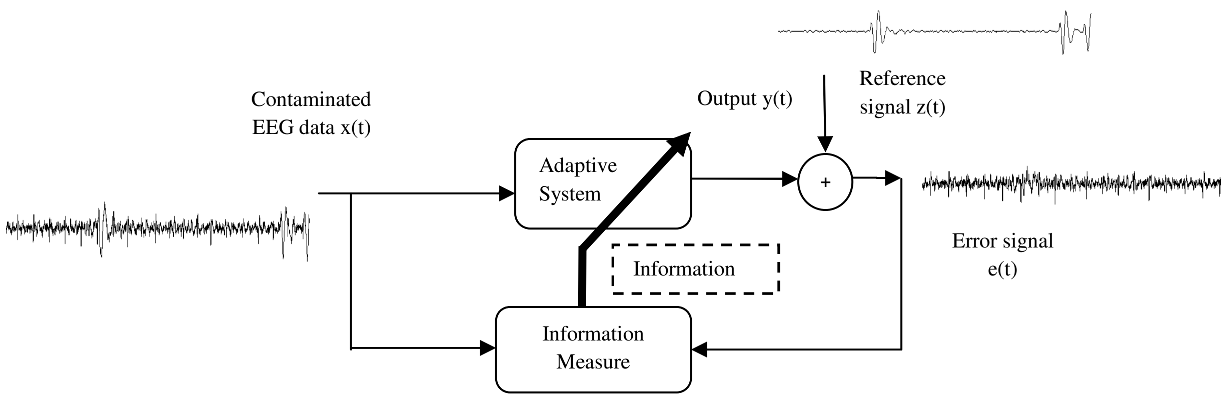


Figure 4.7: A general block diagram of an adaptive noise canceller (Principe, 2010)

Minimizing the error entropy in fact minimizes the distance between the probability density functions (PDFs) of the input and output of the adaptive system (Principe, 2010). In the proposed approach we apply Renyi's Entropy (Renyi, 1960) instead of Shannon's definition of entropy since it needs less computational effort to be estimated directly from samples. For a random variable X whose PDF is given by $f_X(x)$ Renyi's entropy of order α is given by:

$$H_\alpha(X) = \frac{1}{1-\alpha} log \int_{-\infty}^{\infty} f_X^\alpha(x)dx = \frac{1}{1-\alpha} lo\, g(E_X[f_X^{\alpha-1}(x)]) \tag{4.1}$$

For $\alpha = 2$ the equation reduces to:

$$H_2(X) = -lo\, g(E_X[f_X(x)]) \tag{4.2}$$

The argument of the logarithm is called the information potential (IP). It is easier to estimate and work with the information potential rather than the entropy itself. A recursive equation for estimating the IP directly from the samples of a non-stationary signal can be calculated (Wolpaw, et al., 1998) as:

$$V_{k+1} = (1-\lambda)V_k + \frac{\lambda}{L}\sum_{i=k-L+1}^{k} \kappa_\sigma(x_{k+1} - x_i) \tag{4.3}$$

Here λ is the forgetfulness factor and L is the length of the window of the past samples taken into consideration. $\kappa_\sigma(x)$ is an even symmetric kernel function whose kernel size is σ. We have then implemented a linear adaptive filter, which means the intermediate output is given by:

$$f(\vec{x}, \vec{w}) = \vec{w}^T \vec{x} \tag{4.4}$$

$\vec{x}$ is the contaminated input signal, and $\vec{w}$ is the weight vector. The error signal is the difference between the desired signal and the intermediate output of the filter:

$$e(n) = z(n) - \vec{w}^T \vec{x} \tag{4.5}$$

The error entropy is computed using the estimator described in the previous section. The filter weight update recursion is given by:

$$\vec{w}_{k+1} = \vec{w}_k - \mu \frac{\partial V_k}{\partial \vec{w}_k} \tag{4.6}$$

Where (Jian-Wu, et al., 2003),

$$\frac{\partial V_k}{\partial \vec{w}_k} = (1-\lambda)\frac{\partial V_{k-1}}{\partial \vec{w}_k} + \frac{\lambda}{L}\sum_{i=1}^{k-1} \kappa'_\sigma(e_k - e_i)\left[\frac{\partial e_k}{\partial \vec{w}_k} - \frac{\partial e_i}{\partial \vec{w}_k}\right] \tag{4.7}$$

The quantity $\frac{\partial V_{k-1}}{\partial \vec{w}_k}$ can be approximated by $\frac{\partial V_{k-1}}{\partial \vec{w}_{k-1}}$ and in the case of the linear adaptive filter the quantity $\frac{\partial e_k}{\partial \vec{w}_k} - \frac{\partial e_i}{\partial \vec{w}_k}$ can be approximated by $\vec{x}_k - \vec{x}_i$.

The parameters involved are filter order, forgetfulness factor (λ), window length (L), kernel (κ), kernel size (σ), and learning rate (μ). All these parameters were optimized before the proposed algorithm has been applied in our online BCI experiment. The optimum values of these parameters using the current dataset can be seen in Table 4.1.

4.4.1.1. Implementation

The critical factor in the implementation of the adaptive filter is the choice of the kernel and hence, the kernel parameters. In order to investigate the effect of each element, the filter was implemented with two different kernels, Gaussian $G_\sigma(x)$and Gabor $k_{\sigma,f}(x)$. Both were implemented with and without adaptation of their kernel parameters.

$$G_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} \tag{4.8}$$

$$k_{\sigma,f}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} \cos 2\pi f x \tag{4.9}$$

The adaptive filter was implemented with and without the adaptation of the kernel parameters. Adaptation of kernel parameters was carried out using a regression method.

$$\sigma(n+1) = \sigma(n) + \eta \frac{\partial J_{KL}(\sigma(n))}{\partial \sigma(n)} \tag{4.10}$$

The cost function used was the Kullback-Leibler (KL)-divergence between the estimated PDF of the error and the true PDF of the error signal. η is the learning rate.

$$J_{KL}(\sigma) = \frac{1}{N}\sum_{i=1}^{N} log\left(\frac{1}{N}\sum_{j=1}^{N} k_\sigma\left(e_i - e_j\right)\right) \tag{4.11}$$

The results varied depending upon the training sample upon which the filter was applied. However some general trends could be seen. It was observed that the Gabor kernel performed better than the Gaussian kernel on average. This could probably be due to the fact that the shape of the Gabor kernel is similar to that of the artifact which was mostly produced by blinking. The effects of the non-kernel parameters on performance of the filter were also studied. Since EEG is a non-stationary signal, it is not advisable to set the window length to a large value and include too much prior data to estimate the PDF of the error signal. This is also why without σ adaptation the forgetfulness factor should be close to 1. As expected, it was observed that the performance increased with higher values of the forgetfulness factor (λ). Increasing window length (L) not only increased the computation time by a large amount, but also reduced the performance of the filter. Large computation time is not desirable for an online adaptive filter so L was set to 16 data points for all cases. Similarly, adaptation of kernel parameters also slows down the filter. Even though there is an increase in performance, it is not so much for being worth compromising the response time of the filter.

Table 4.1: The parameter values used in different methods of the adaptive filtering for artifact suppression. For the adaptive cases the noted value is the initial value of that parameter. L is set to 16 data points for all cases.

Method	**Parameters**						
	λ	***σ***	**Kernel frequency**	***μ***	**Iteration**	**MA order**	**Filter order**
Gabor kernel filter with sigma update	0.7	16	0.1	0.05	20000	40	10
Gabor kernel filter without sigma update	0.99	16	0.5	0.5	20000	40	10
Gaussian kernel filter with sigma update	0.4	70	-	0.05	20000	40	10
Gaussian kernel filter without sigma update	0.99	70	-	0.5	20000	40	10

Figure 4.8 demonstrates the output of the algorithm in each step, where the input is the EEG recording from F3 and the reference signal is Fp1. The artifact-free data is depicted in red. Changes in the value of intermediate output start at 7 s and continue till the fluctuation in the main signal exists. As can be seen, the method does not distort the data when no artifact exists. The same procedure was applied simultaneously to the other recording sites.

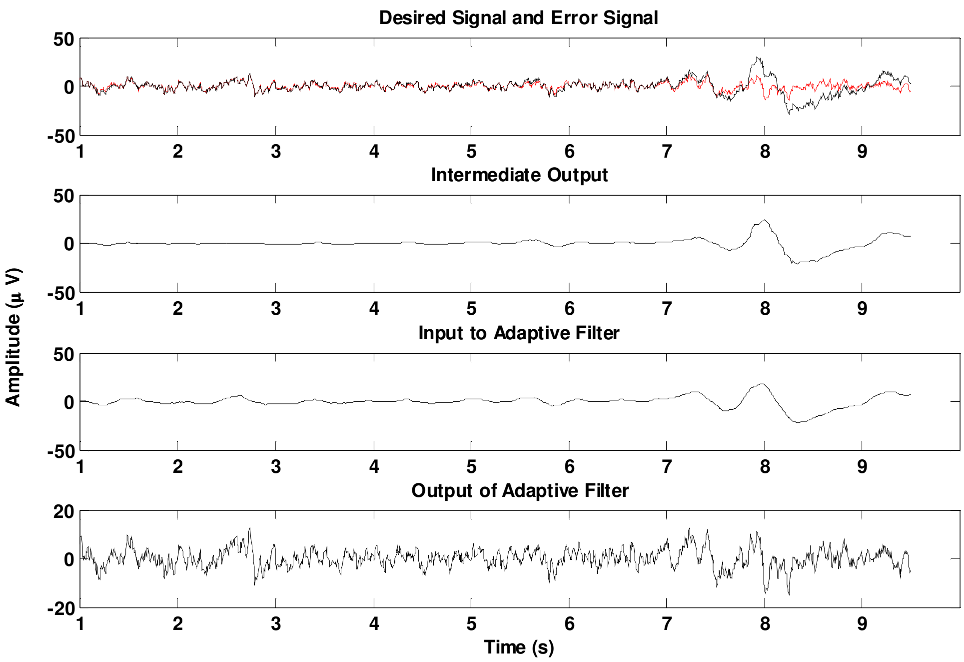


Figure 4.8: Gabor filter with the sigma update over the time applied in 10 s ongoing EEG data. The output of the adaptive filter is the artifact-free EEG. The upper panel shows the measured signal at F3 in black, and the corrected signal in red (output of the adaptive filter). The reference signal (input to the adaptive filter) is the signal recorded from Fp1 smoothed by a moving-average filter.

4.4.1.2. Validation

One of the main concerns in artifact correction is the validation step. The contaminated part of the signal which is corrected by the computational method needs to be compared to an original source which is latent. In order to overcome this issue we introduce a criterion based on the similarity of the information between the clean and artifact segments before and after artifact correction. In order to evaluate the performance of the information theoretic filter as compared to the least-squares filter, we use correntropy (Principe, 2010) as the similarity criterion. The correntropy of two signals X and Y is given by:

$$V(X,Y) = \frac{1}{N}\sum_{i=1}^{N} k(x_i - y_i) \tag{4.12}$$

Here $k(x)$ is a kernel function (usually Gaussian). The correntropy is highest when the two signals are closest to each other (Weifeng et al., 2006). The performances of the different implementations of this filter were compared to those of a LMS adaptive filter. The general similarity measure used was the correntropy of the output of the filter $e(n)$ and the desired signal $z(n)$.

$$similarity = V(e_{total}, z_{total})\% \tag{4.13}$$

Similarity values and the total computational time can be seen in Table 4.2 for each method using the signal depicted in Figure 4.8. However, in order to evaluate the performance of the method in details we defined an overall performance measure as the proportion of the sum of correntropy calculated for the signal divided by the sum of correntropy calculated for artifact segments. The reason is that because performance does not intuitively depend upon the similarity of output and the desired signal in areas where artifacts are present, the signal and artifact regions of the EEG recording had to be distinguished.

$$correntropy_{signal} = V\left(e_{non-artifact}, z_{non-artifact}\right) \tag{4.14}$$

$$correntropy_{artifact} = V\left(e_{artifact}, z_{artifact}\right) \tag{4.15}$$

$$performance = \frac{correntropy_{signal}}{correntropy_{artifact}} \tag{4.16}$$

Table 4.3 summarizes these calculations for the current dataset. Artifacts were considered those segments of the signal where the output of the adaptive filter differs more than 10% from the input signal.

Table 4.2: Comparison between our proposed method and two commonly used approaches for artifact reduction

Evaluation method	**Comparison**	
	Similarity %	**Computational timing**
Entropy-based Gabor kernel with $\boldsymbol{\sigma}$ adaptation	78.96	34 ms
Entropy-based Gabor kernel	74.24	20 ms
Entropy-based Gaussian kernel with $\boldsymbol{\sigma}$ adaptation	75.65	28 ms
Entropy-based Gaussian kernel	72.14	17 ms
ICA	77.80	120 ms
LMS	69.33	15 ms

Table 4.3: Evaluation of the proposed method based on the averaged comparison between our proposed method and LMS based on the performance factor for the current dataset.

	Evaluation method		
	Correntropy signal	**Correntropy artifact**	**Performance**
Gabor kernel filter with $\boldsymbol{\sigma}$ adaptation	0.8996	0.1256	7.162
Gabor kernel filter without $\boldsymbol{\sigma}$ adaptation	0.7508	0.1208	6.215
Gaussian kernel filter with $\boldsymbol{\sigma}$ adaptation	0.6984	0.1053	6.635
Gaussian kernel filter without $\boldsymbol{\sigma}$ adaptation	0.8747	0.1263	6.925
LMS	0.5585	0.092	6.071

The information filter performed up to 1.2 times better than the LMS adaptive filter with respect to the signal *performance* measure. This means the information filter preserves the original EEG signal in the regions where there is no artifact present. The LMS filter however distorts the signal where there is no artifact. Even though the performance on the *artifact correntropy* parameter is good for the LMS adaptive filter, which means that the filter removes artifacts satisfactorily, the distortion of the original signal information overrides the effective artifact removal. The increment in performance between the Gabor and Gaussian information-based adaptive filters was not reliably greater over all training examples. In any case, Gaussian or Gabor, with or without adaptation of kernel size, the information filter employing Renyi's entropy as the cost function performed on average 20% better than the LMS adaptive filter on the training samples with respect to the *correntropy* parameter (Table 4.3). In our online setup, the Gabor filter with sigma adaptation was used.

4.4.2. Feature extraction based on higher order statistics (HOS)

The term higher-order statistics (HOS) describes the functions of third or higher order (Petropulu, 2000). The higher-order spectrum as an extension of the Fourier spectrum uses higher moments for spectral estimates. By eliminating the effects of Gaussian random processes, HOS is able to reveal non-Gaussian and nonlinear characteristics in complex patterns such as EEG time series (Mendel, 1991). In what follows we employ higher-order spectrum for the extracting of the hidden attributes of EEG signals. Based on this analysis a multidimensional feature space for the current BCI application is proposed.

4.4.2.1. Higher order momentums and cumulants

Moments are quantitative measures to describe the distribution of data points. In the one-dimensional case, the *ith* central moment of the signal $x(n)$ is defined as: $m_i\{x(n)\} = E\{(x(n) - \mu)^i\}$, where $E\{.\}$ denotes the ensemble expectation operator and μ is the mean of the random process. For reasons of mathematical convenience, specific nonlinear combinations of these moments, called cumulants, are often used instead of moments (Mendel, 1991). The first-order and the second-order cumulants are the mean and the auto covariance sequence of the process. Under the assumption of a stationary real-valued process with zero mean, the first-, second-, third- and fourth-order cumulants can be expressed by following equations (Petropulu, 2000) , where n_i denotes the timing lag in each equation.

$$
\begin{aligned}
&C_{1x} = E\{x(n)\} \\
&C_{2x}(n_1) = E\{x(n)x(n+n_1)\} = M_{2x}(n_1) \\
&C_{3x}(n_1, n_2) = E\{x(n)x(n+n_1)x(n+n_2)\}
\end{aligned}
\tag{4.17}
$$

$$
\begin{aligned}
C_{4x}&(n_1, n_2, n_3) \\
&= E\{x(n)x(n+n_1)x(n+n_2)x(n+n_3)\} - C_{2x}(n_1)C_{2x}(n_2 - n_3) \\
&- C_{2x}(n_2)C_{2x}(n_1 - n_3) - C_{2x}(n_3)C_{2x}(n_1 - n_2)
\end{aligned}
$$

For calculating the third-order moment two independent lags n_1 and n_2 should be set. The formulation of higher order moments is done by adding lag terms in a similar way. The zero-lag cumulants have special names: $C_{2x}(0)$ is the variance and the normalized quantities of $C_{3x}(0,0)$ and $C_{4x}(0,0,0)$ are usually referred to as skewness and kurtosis (Zhang, 2010).

$$skewness = \frac{m_3[x(n)]}{\sigma^3} \quad , \; kurt = \frac{m_4[x(n)]}{m_2^2[x(n)]} - 3 \tag{4.18}$$

Skewness is a measure of symmetry of the distribution of the data set. A symmetric distribution looks the same in the both sides of the mean point. Skewness is negative if the distribution is more to the right of the mean point (Mendel, 1991). As defined earlier in chapter 3, kurtosis is a measure of Gaussianity. A Gaussian distribution has zero kurtosis (but not vice versa), positive kurtosis implies a sparse distribution, "super-Gaussian" (leptokurtotic). We used these two parameters as features for the current BCI setup.

4.4.2.2. Bispectrum and Bicoherence

Any Gaussian signal can be thoroughly defined by its first and second order statistics i.e., mean and variance. So the higher order spectrum of Gaussian signals either contains redundant information or has zero moments. For instance, a Gaussian signal has zero for the 3rd order or higher moments (Mendel, 1991). Phase information is not present in the second-order measures (such as the autocorrelation functions or the power spectrum). For this reason, those phase coupling that are associated with nonlinearities cannot be correctly identified by second-order statistics (Maclaughlin et al., 1995-add) (Petropulu, 2000). The autocorrelation function is the second-order cumulant, as defined above (Eq. 4.17), and its Fourier transform is the power spectrum of the signal. Similarly, bispectrum is the Fourier transform of the third-order cumulant (Petropulu, 2000). In bispectrum and bicoherence $B(\omega_1, \omega_2)$ the prefix *bi-* is used to indicate the relation between two frequencies of a single signal, ω_1 and ω_2 which differs from the coherence relation between two time series.

$$B(\omega_1, \omega_2) = \sum_{n_1=-\infty}^{+\infty} \sum_{n_2=-\infty}^{+\infty} x(n)x(n+n_1)x(n+n_2) \quad e^{-j(n_1\omega_1+n_2\omega_2)}$$

$$\cong \frac{1}{N}\sum_{i=1}^{N} X_i(\omega_1)X_i(\omega_2)X_i(\omega_1+\omega_2) \tag{4.19}$$

Here the discrete Fourier transform of the *ith* section of the EEG signal is shown by $X_i(\omega)$ (Sanei, 2007). At coordinate point (ω_1, ω_2), the magnitude of the bispectrum calculates the degree of phase coherence between the three frequency components ω_1, ω_2 and $\omega_1 + \omega_2$. So the magnitude of the joint Fourier spectrum also affects the magnitude of bispectrum $|B(\omega_1, \omega_2)|^2$. The normalized bispectrum, also called bicoherence, is defined upon the following relationship:

$$b^2(\omega_1, \omega_2) = \frac{|B(\omega_1, \omega_2)|^2}{P(\omega_1)P(\omega_2)P(\omega_1, \omega_2)} \tag{4.20}$$

Where $P(\omega) \cong \frac{1}{N} \sum_{i=1}^{N} X_i(\omega)\ X_i(\omega)^*$ is the power spectrum of the signal $x(n)$. In principle, HOS can be a powerful tool in distinguishing the biological signal and the Gaussian background noise from each other. The reason is that the majority of biological signals have non-Gaussian distributions, but measurement noises have usually Gaussian distributions. For this purpose, the second-order measures cannot be as effective as HOS, because they cannot capture the differences and therefore they are more affected by the background noise. In the present BCI application, six features based on HOS characteristics of the EEG signal were calculated using modified functions of HOSA toolbox in MATLAB. Figure 4.10 illustrates the estimation of bispectrum averaged over the dataset. Phase coupling during the imagination interval happens in lower frequencies, but during the relaxation it shows a wider distribution and reaches its maximum in the alpha and beta bands. The red color shows the maximum value of the bispectrum which can be seen in the lower frequency during the imagination phase.

Relaxation — Imagination of movement

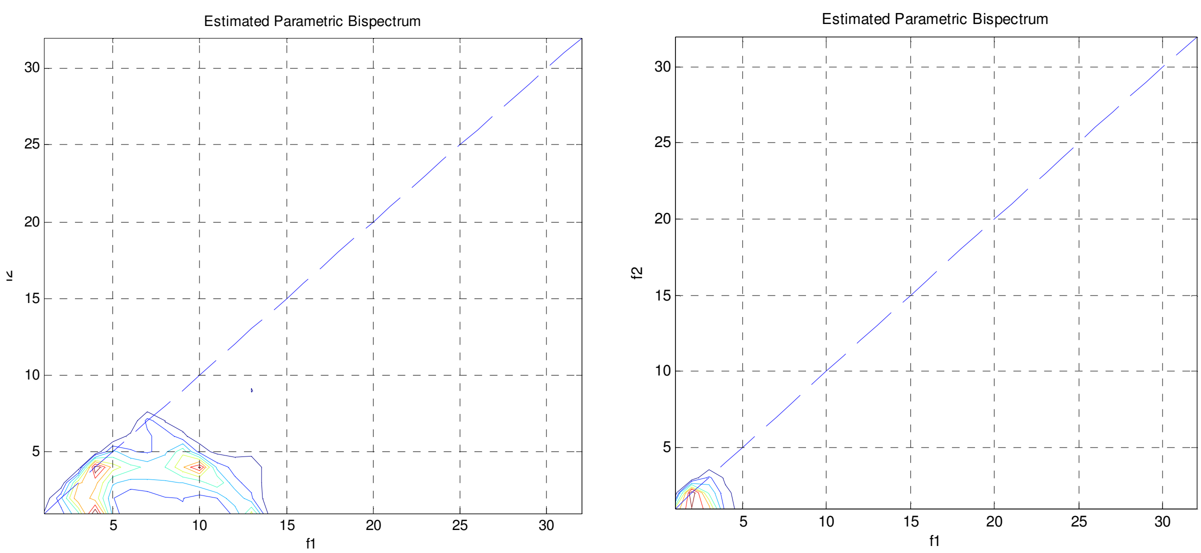


Figure 4.10: Estimated parametric bispectrum demonstrates the higher order dependency in different brain states (Imagination of movement state vs. relaxation state) in the current dataset. Units on x- and y-axis are frequency (Hz). Colors show the magnitude of bispectrum in V^3/Hz^2 , with blue = 2×10^{-3}, and red = 200. White areas between the isolines show interpolated values.

Figure 4.11 shows the extended bicoherence estimation via direct FFT method for the relaxation state (left column) and the imagination state (right column) averaged over all subjects and trials.

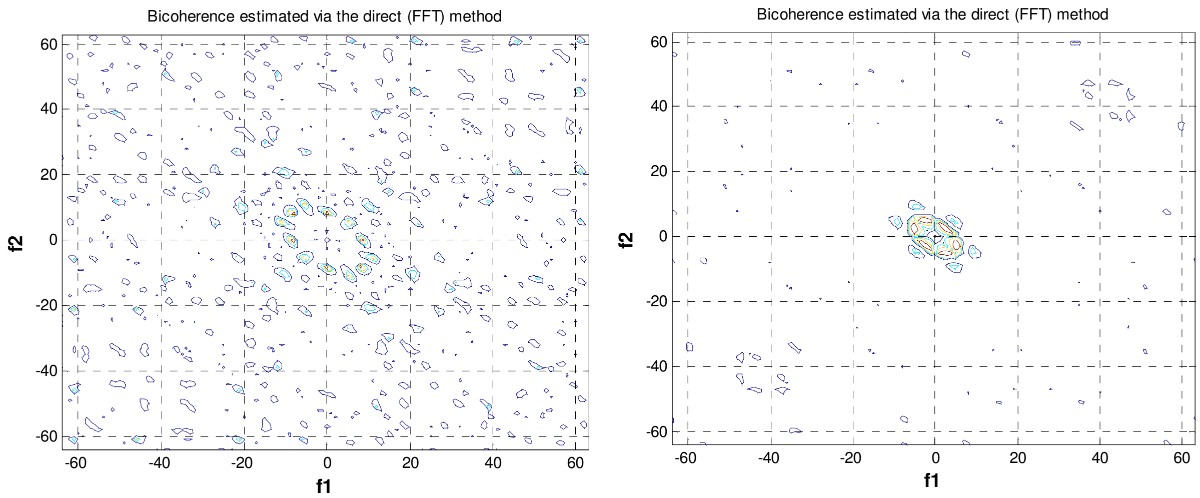


Figure 4.11: A contour demonstration of the bicoherence estimated for the relaxation state (left panel) and the imagination state (right panel) averaged over all subjects and trials. Units on x- and y-axis are frequency (Hz). Negative frequencies are interpreted as negative phase shifts. Colors show the normalized magnitude of bicoherence, with blue = 0, and red = 1. White areas between the isolines show interpolated values.

The results of this analysis show that the distribution of EEG signal bispectra on the bi-frequency planes is smaller during imagination than relaxation and the distribution during relaxation tends more to be more chaotic as may be seen from the coupling between different frequencies in Figure 4.10. It means, there is more frequency coupling during the relaxation state in comparison to the movement state. Plots are symmetric regarding the frequency axes.

In order to construct a feature set based on the higher order spectrum, six characteristics were extracted from each EEG trial: Maximum and minimum values of bicoherence, average power and the distance from the center (maximum frequency coherence) along with kurtosis and skewness values (cf. Figure 4.12 for an illustration). These six bispectrum estimates were extracted from each of the 7 EEG channels (= 42 features) along with 4 broad-band parameters from the power spectrum (theta, alpha, beta, gamma bands, × 7 = 28). Figure 4.12 shows an example of the segmented signal during the imagination and relaxation states recorded for S1 while working with the online BCI system (left panels in first and third rows). Its corresponding normalized power spectrum is depicted below the signal (left panels I second and forth row). The right panel shoes third-order cumulant distributions with different lags (first and third rows), and the calculated bicoherence (second and forth row). Color bars show the magnitude of cumulants (V^3) and bicoherence (V^3/Hz^2). Lags for calculating the third-order cumulant theoretically may vary from -∞ to +∞, but the range is restricted to -100 ms to +100 ms in Figure 4.12, which directly affects the resolution of the corresponding bicoherence (its Fourier transform) in range of –50 Hz to 50 Hz.

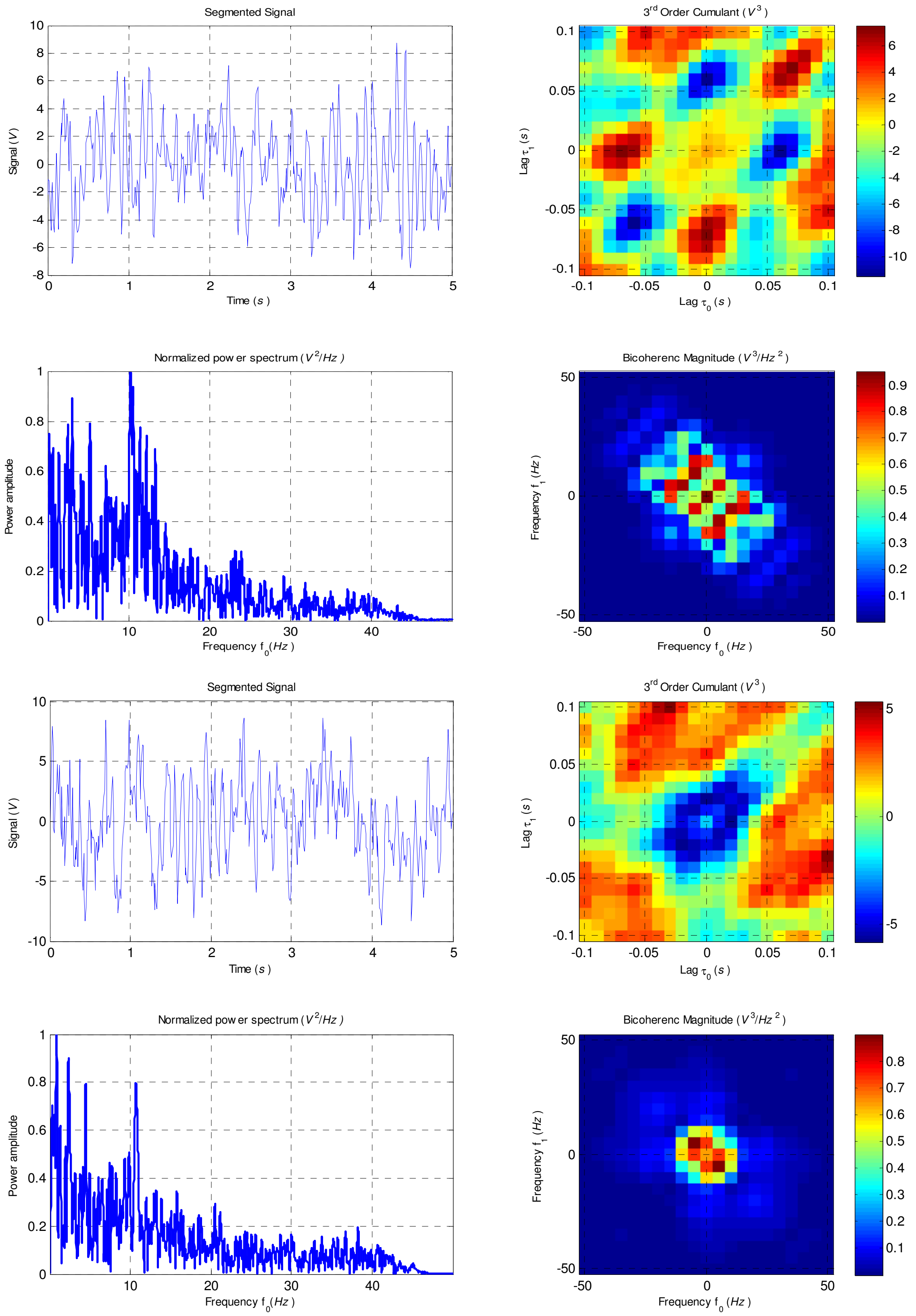


Figure 4.11: A sample of the segmented signal, normalized power spectrum, third-order cumulant distribution and bicoherence magnitude for the imagination state (upper four subplots) and the relaxation state (lower four subplots). Units were inserted for each subplot separately.

In our analysis selecting greater lag values up to 1000 ms for each segmented EEG is possible but it has direct effect on the increasing the speed of computation, which is not desired in our BCI setup. These displayed cumulant distributions show smoother connections in the relaxation state than in the imagination state. The values of the six extracted parameters are in this example for the imagination and relaxation states: Maximum bicoherence value 0.852 vs.0.772, minimum bicoherence value 0.09 vs. 0.02 (considering non-zero values in the positive frequencies), average power 0.264 vs. 0.093, max frequency coherence 18 Hz vs. 7 Hz (visible as red dots in the positive frequencies in the bicoherence plots distance of the dark red squares from the zero point). Skewness was calculated as normalized value in the center of the third-order cumulant distribution and amounted to 1.74 vs. -1.52. Kurtosis values for the two signals amounted to 3.12 vs. 2.71 respectively. Empirically the kurtosis value of a distribution is linearly proportional to its squared of skewness (Schopflocher & Sullivan, 2005), therefore these two values are not independent.

The constructed 70-dimensional feature space was fed to a feature selection module (4.4.3) and the selected features finally were used in our BCI to send a control command to a robotic hand. Averaged HOS extracted features for the imagination state calculated for the ten subjects can be seen in Table 4.2.

Table 4.2: Averaged HOS extracted features from EEG signals calculated for the ten subjects for the imagination state.

Subject	**Max and min values of bicoherence**	**Max frequency coherence**	**Average Power**	**Kurtosis**	**Skewness**
S1	0.737, 0.335	18	0. 625	4.32	1.98
S2	0.726, 0.407	27	0.633	3.15	2.25
S3	0.899, 0.501	15	0.747	4.50	2.07
S4	0.938, 0.678	22	0.820	3.91	2.16
S5	0.990, 0.631	21	0.755	3.42	2.12
S6	0.834, 0.551	16	0.711	4.57	1.92
S7	0.805 ,0.375	12	0.606	2.73	2.06
S8	0.912 ,0.466	15	0.615	3.06	1.72
S9	0.972 ,0.560	11	0.846	4.13	1.73
S10	0.885 ,0.365	14	0.645	4.62	1.84

The two most effective features out of 10 features were selected by the mRMR selection algorithm for each channel during the training mode. The output command is formed from the

7-channel EEG signal every 0.5 s using a 2 s data segment with 75% overlap on the preceding and following segments.

4.4.3. Feature selection: Maximum relevance minimum redundancy

Selecting a subset of superior features using the maximal statistical dependency criterion is a well-known feature selection method in machine learning (Koprinska, 2009). The Max-Dependency criterion is based on mutual information but is mathematically difficult to calculate in a real-world application. An alternative approach for selecting features based on the maximal relevance minimum redundancy criterion is the mRMR algorithm which has been evaluated by extensive experimental comparison of the proposed algorithm and other methods using different classifiers and different data sets (Peng et al., 2005). The results confirm that mRMR leads to promising improvement on feature selection and consequently increasing the classification accuracy (Peng et al., 2005). In the mRMR approach, M features are selected based on the maximum mutual information calculated between each two features and also between each feature and the class labels. The criterion is to find, at the same time, the minimal redundancy amongst features and the maximal relevance to the class labels. Peng et al. reported a detailed description of the mRMR algorithm (Peng et al., 2005). Based on their formulation, the algorithm first finds Max-Relevance or $\mathrm{D(S, c)}$ satisfying Eq. 4.21. The equation is calculated based on an approximation all mutual information $I(x_i; c)$ between individual feature x_i and class c and then the mean is calculated over the feature space:

$$\max D(S,c)\,, D = \frac{1}{|S|}.\sum_{x_i \in S} I(x_i; c) \tag{4.21}$$

It is likely that features selected according to Max-Relevance could have rich redundancy, i.e., the dependency among these features could be large. Therefore, removing or adding one of them will not alter the respective classification accuracy. Peng et al. proposed the second step in their algorithm and added the minimal redundancy (Min-Redundancy) condition. Eq. 4.22 selects mutually exclusive features when the sum of mutual information between each two is minimized (Peng et al., 2005):

$$\min R(S)\,, R = \frac{1}{|S|^2}.\sum_{x_i, x_j \in S} I\left(x_i; x_j\right) \tag{4.22}$$

We used the mRMR method for the current BCI paradigms in order to decrease the feature space to the two most relevant features (total of 14 features). Selection of the most appropriate features based on mRMR increased the classification accuracy in our current and following BCI applications. Table 4.3 shows the selected features for each classifier trained on the data of each recording channel. The two most informative features based on mRMR selection contain both HOS and power band features.

Table 4.3: The two most informative features selected for each EEG channel during the training phase of S1.

	EEG channels						
	F3	**F4**	**Fz**	**C3**	**C4**	**Cz**	**Pz**
Feature1	Max frequency coherence	Alpha power	Theta power	Theta power	Kurtosis	Kurtosis	Max frequency coherence
Feature2	Average Power	skewness	Max bicoherence	Max frequency coherence	Alpha power	Average Power	Max bicoherence

4.4.4. Classification

For this application after selecting the features we employed a kernel-based support vector machine (SVM) classifier in order to deal with non-linearity of the new feature space (N=2 × 7). SVMs are used with kernel functions as transformation function of the input data to the higher dimension feature space (Müller et al., 2003). By exploiting the idea of a kernel, the linear maximum margin classifier, described in Chapter 2, can be extended to nonlinear classifiers. Furthermore in the case that the transported features in the new space are not balanced, they are practically separable by adjusting the parameters of a maximum margin classifier. This transformation of the input data to the feature space is done without explicitly specifying the transformer (Vapnik, 1999). Figure 4.12 depicts the steps of the algorithm. It consists of three essential parts: the dual form of linear algorithms, nonlinear mapping and the kernel function (Liang, et al., 2011). There are some commonly used kernels:

- Linear kernel: $K(x_i, x_j) = a_0 {x_i}^t x_j + b$
- Radial basis function (RBF): $K(x_i, x_j) = exp(-\gamma . \|x_i - x_j\|^2)$
- Sigmoid kernel: $K(x_i, x_j) = \tanh(a_0 {x_i}^t x_j + b)$

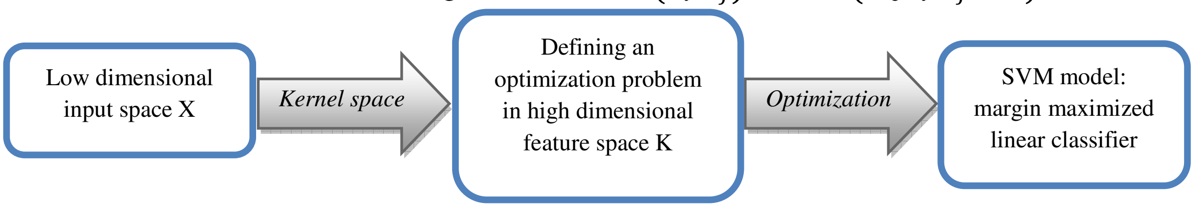


Figure 4.12: The kernel-based SVM model comprises three essential steps. The low dimensional input space X is projected to a high dimensional feature space K by using a kernel trick. After that the optimization problem is solved in order to find the maximum margins (Liang, et al., 2011).

We employed LIBSVM as a standard MATLAB package for computing the SVM (Chang et al., 2001). A grid search was used to determine the optimized values for parameters of the RBF kernel, i.e. the kernel parameter γ and the regularization parameter C. The parameter C indicates a tradeoff between the training error and the model complexity. If the C value is small, it means that the model is simple and also there is a small deviation between the test error and the training error. But the training error might be large per se. On the other hand, a large C value guarantees that the training error is small. It should be noted that larger C values leads to higher complexity of the model. It is also more probable that the increase in test error exceeds significantly the increase in the training error. Hard-margin SVM is the extreme case when the value of C goes to infinity (Schouenborg et al., 2010). Various pairs (C, γ) values were tried and the one with the best cross-validation accuracy was picked for each channel. As proposed in the literature, we tried exponentially growing sequences of C and γ which is a fast and practical approach to identify the best parameters. We swept the range between 2^{-3} and 2^{10} to find the optimum C and γ in our dataset. Optimized values of SVM classifiers calculated for each EEG channel were used in an ensemble scheme to produce the final command to the hand prosthesis.

Table 4.4: Optimized value of SVM classifier calculated for each EEG channel.

	Channels EEG						
	F3	**F4**	**Fz**	**C3**	**C4**	**Cz**	**Pz**
C	2^3	2^5	2^5	2^6	2^8	2^4	2^9
γ	2^{-4}	2^{-3}	2^{-3}	2^{-2}	2^{-3}	2^{-2}	2^{-4}

4.4.5. Ensemble classification for decision making

A way to reduce the influence of signal variability in the classification problem is to use an ensemble of classifiers approach. Indeed, averaging of the classifier outputs is known to be an efficient technique to soften the classifier variability. The simplest way to combine different classifiers in a process of decision making is voting. It is fast and efficient to implement and usually has a satisfactory result in the practical problems (Kuncheva, 2004). In this method a predefined set of classifiers is trained using a subset of the dataset. In our application for hand prosthesis control, each classifier is trained using the features extracted from an EEG channel. The final decision is made based on the highest number of classifiers which predicted the specific class. As an alternative we compared the results in an offline study for fusion and concatenation strategies for decision making. Figure 4.13 depicts the block diagram for each method.

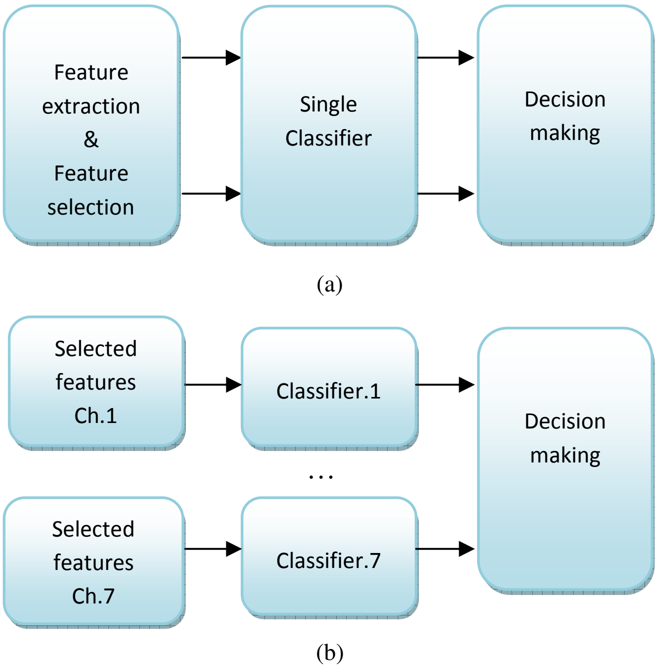


Figure 4.13: Two different strategies for ensemble classification employed on this BCI application, (a) Fusion at classification level, (b) Concatenation: A single feature vector was fed to different classifiers and the voting technique was applied for the decision making.

4.5. Assessment : Statistical measures of performance

In order to evaluate the generalization capability of our system, we have calculated specificity, sensitivity, and Matthew's correlation coefficients (MCC) besides the popular accuracy concept in this dataset (Altman, et al., 1994) (Krishna et al., 2012). These concepts are based on some primary definitions: True negative (TN) is when the test yields a negative result (here: predicts the relaxation state) and relaxation was indeed required. Similarly, true positive (TP) is when the test results in positive (+1 value as the output of the classifier) and the hand movement imagination was indeed required. False negative is incorrect rejection of a positive result and false positive is the failure to reject a false result (Hastie et al., 2008). False positive (FP) and false negative (FN) are called error type I and II respectively (Hastie et al., 2008). Table 4.5 summarizes these definitions based on the relation between the condition and the result outcome.

4.5.1. Sensitivity and Specificity

Sensitivity is a measure of evaluation which measures the percentage of TP divided by the sum of TP and FN.

$$Sensitivity = \frac{TP}{TP + FN} \tag{4.23}$$

Specificity is a measure of evaluation which measures the percentage of TN divided by the sum of TN and FP.

$$Specifity = \frac{TN}{TN + FP} \tag{4.24}$$

Table 4.5: A summary of evaluation parameters based on the test outcome

		Condition		
		Condition positive	Condition negative	Predictive value
Test outcome	Test outcome positive	True Positive	False Positive (Type I error)	Positive
	Test outcome negative	False Negative (Type II error)	True Negative	Negative
		Sensitivity	**Specificity**	

The terms sensitivity and specificity are employed to characterize a rule in medical classification problems. They are defined as the probability of predicting disease given a true state of disease and the probability of predicting non-disease given a true state of non-disease respectively (Hastie et al., 2008). In the BCI application we can assume movement and a relaxation brain state as the presence and absence of disease in the medical problem.

4.5.2. Matthew’s correlation coefficient

Matthew’s correlation coefficient is a measure of the quality of binary classification and ranges from $-1 \leq \text{MCC} \leq 1$. The best possible prediction is achieved when the MCC value is equal to 1 and the anti-correlation or the worst possible prediction happens when MCC = -1. Random prediction results would be expected by MCC = 0. We used this measure in order to evaluate our current BCI system.

$$MCC = \frac{(TP * TN - FP * FN)}{\sqrt{(TP + FP)(TN + FN)(TP + FN)(TN + FP)}} \tag{4.25}$$

4.5.3. Applying ANOVA (ANalysis Of VAriance) test

By using several statistical measures the performance of the binary classifiers and the feature selection method were evaluated by parametrical statistical tests: t-test and ANOVA. There are two types of these tests, depending on whether the measurements to be compared were taken in different subjects or were taken in the same subjects (as in the present application). In the latter case, the measures are correlated, and therefore what is actually tested is the difference between the measures. In the case of 2-level variables, t-test and one-way ANOVA provide identical p-values (Hastie et al., 2008). The advantage of ANOVA is that the effects of several independent variables on one dependent one may be simultaneously measured. We used SPSS statistical software (IBM SPSS statistics, version 20) for this analysis.

4.6. Results

We tested the feasibility of the entire system for real-time execution. Figure 4.14 shows the averaged classification accuracy for all subjects employing both single SVM (all features + one classifier) and ensemble SVM (classifier per channel + voting) classification techniques.

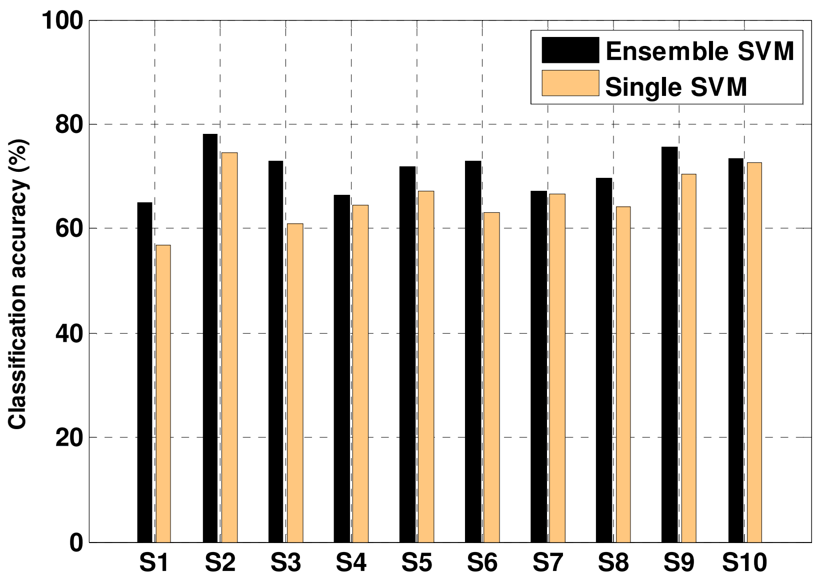
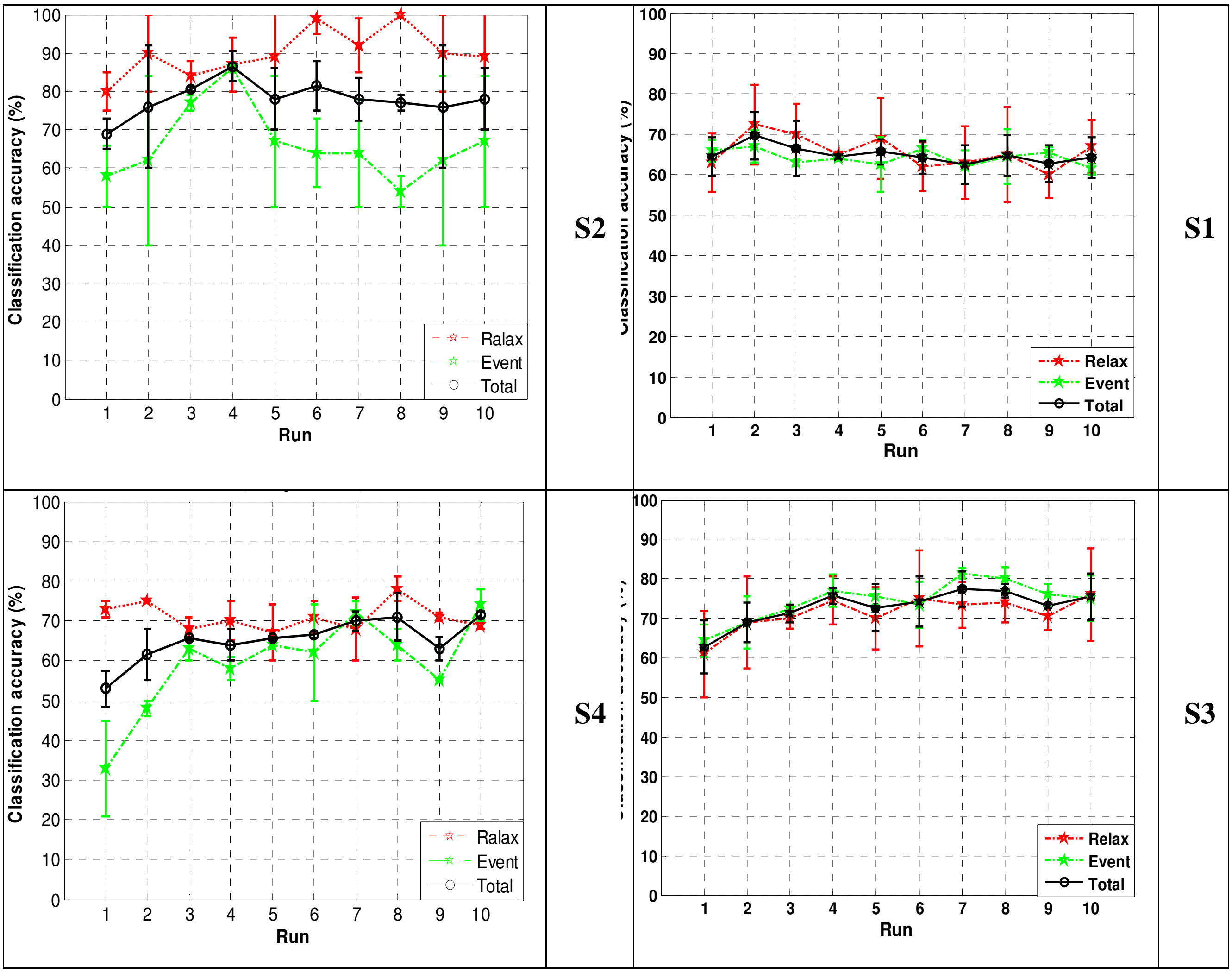


Figure 4.14: The averaged classification accuracy for all subjects employing both single SVM (all features + one classifier) and ensemble SVM (classifier per channel + voting) classification techniques

Presented results in this section were achieved by running the system in a simulated mode, in which EEG signals were read in a simulated online scheme from our previously recorded database. The time scale on which commands are sent to the robot should be congruent with the BCI paradigm and hardware limitations, which is every 0.5 s. Figure 4.15 shows the averaged classification accuracy for all subjects over different experimental sessions using the ensemble classifier technique. S2 showed the highest classification accuracy among the ten subjects but at the same time shows more variation in comparison to the results calculated for the other subjects. The classification accuracy during the relaxation phase is higher for this subject. For S5 the accuracy for relaxation is lower than for imagination. In the latter case two phases of progress can be seen from first to third run and from fourth to eighth run. Most subjects (all except S1 and S9) showed an increase in the classification result from the first run of the experiment till the end of the experiment. For S2 the averaged classification accuracy increased in the first three runs and then it was almost robust till the end of the experiment. S3, S6 and S10 have a continuous improvement pattern with low variance between classes. Comparing the accuracy values in the first and last runs of the experiment show a significant progress for all subjects (df=9, t=5.22; $p < .001$)



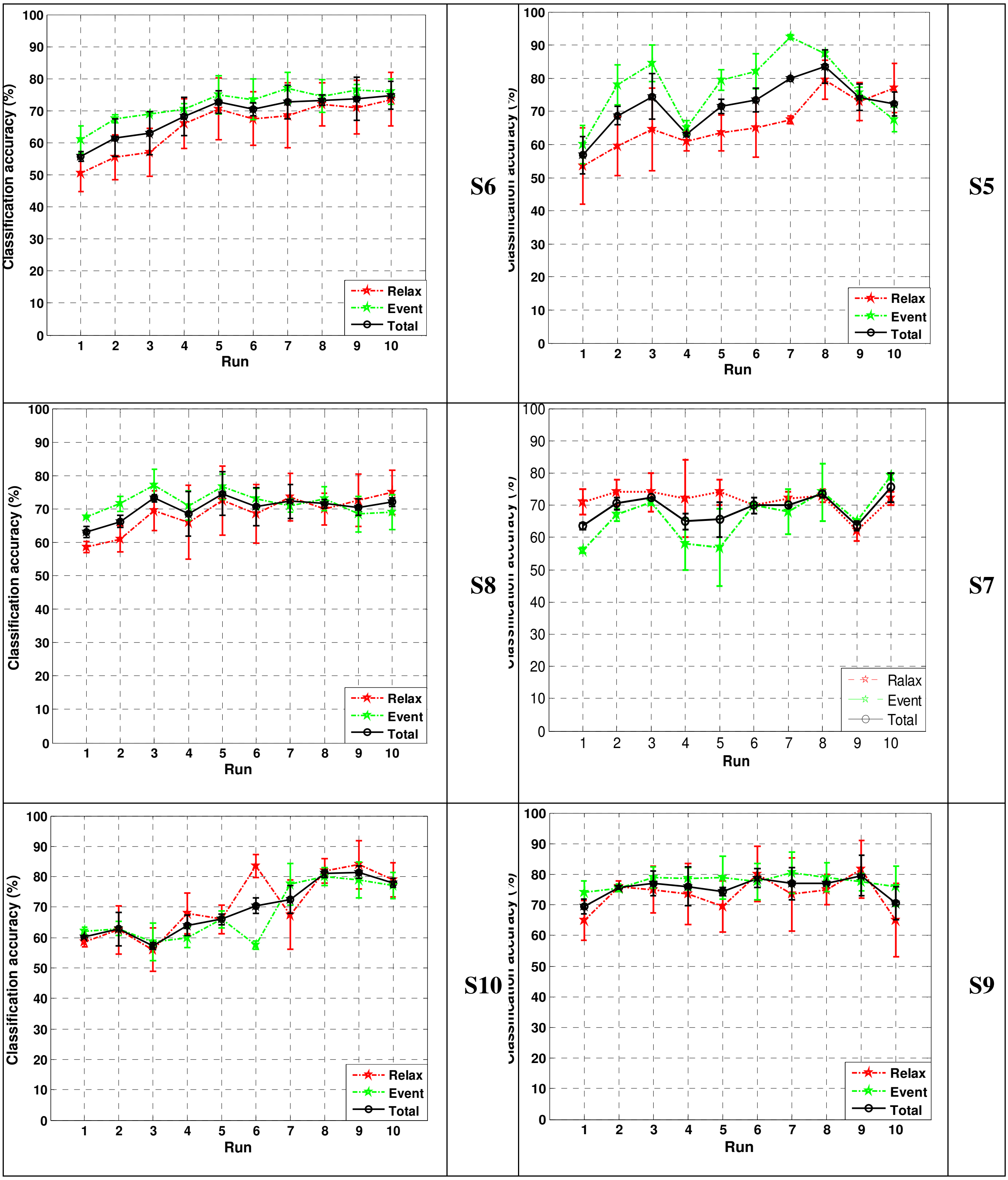


Figure 4.15: The classification accuracy for all subjects averaged over different experimental sessions. "Event" in the panel legends refers to the imagination state.

In this offline study we applied a single SVM classifier fed by the entire feature vector in order to investigate the effect of ensemble classification. It differs from the online processing policy, when the BCI system only has to compute a few relevant features for the control. We compared both classification techniques using MCC and ITR factors. In this experiment commands are issued every 0.5 seconds. Therefore maximum ITR can be theoretically 120 BITS/min with 100% accuracy. For the two mental tasks (relaxation and imagination of the movement) according to Figure 4.2 the information transfer rate in this experiment was found between 24 and 72 bits/min in average with a maximum of 80 bits/min, not far from maximum reported ITR of 90 bits/min (Gao X. et al., 2003). Table 4.6 shows the averaged classification accuracy, MCC, and bit rate for each participant with each paradigm. Simulated online accuracy was significantly higher for the ensemble SVM, 71.24%, than for the simple SVM, 66.56%, $t(9) = 4.78$, $p = 0.001$.

False negative (FN) commands were minimized by using an ensemble SVM classifier based on the bagging algorithm, which is designed based on the selected features from those channels with more than 70% accuracy during the training mode. Finally the majority of classifiers vote for the command to be sent to the robot. This provides a more robust decision making scenario. For example, for subject S2 using the single SVM vs. ensemble SVM specificity, sensitivity and MCC are calculated as:

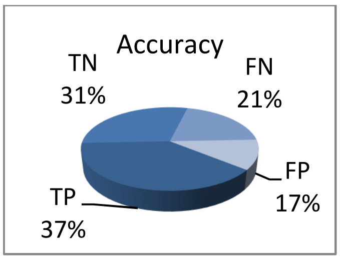
$$Specificity = \frac{TN}{FP + TN} = 100 \times \frac{32}{4 + 32} = 88.89\ \%\ vs.\ \frac{31}{17 + 31} = 64.58\%$$

$$Sensitivity = \frac{TP}{TP + FN} = 100 \times \frac{46}{46 + 18} = 71.87\ \%\ vs.\ \frac{37}{37 + 21} = 63.79\ \%$$

$$MCC = \frac{(TP * TN - FP * FN)}{\sqrt{(TP + FP)(TN + FN)(TP + FN)(TN + FP)}}$$

$$= \frac{(46 * 32 - 4 * 18)}{\sqrt{(46 + 4)(32 + 18)(46 + 18)(32 + 4)}} = 0.5833$$

$$vs.\ \frac{(37 * 31 - 17 * 21)}{\sqrt{(37 + 17)(31 + 21)(37 + 21)(31 + 17)}} = 0.2825$$



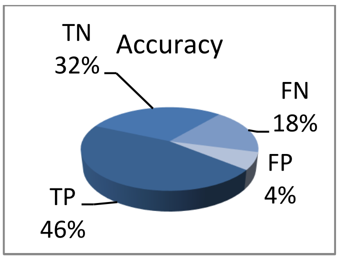


Figure 4.16: The voting algorithm in time and space ("ensemble SVM") improves on average the TP and TN values and leads to higher ITR. The left diagram shows the average values of the accuracy measurement across all trials in detail for S2 for the single SVM method. The right diagram shows the similar calculations when the ensemble technique was employed.

Table 4.6: Evaluation of the classification techniques applied on the current dataset.

	Single SVM			Ensemble SVM		
Subject	Accuracy %	**MCC**	**ITR** BITS/min	Accuracy %	**MCC**	**ITR** BITS/min
S1	56.7	0.12	51	64.9	0.47	73
S2	74.6	0.28	54	78.1	0.58	80
S3	65.8	0.18	49	72.9	0.41	68
S4	64.5	0.22	42	66.3	0.48	53
S5	67.1	0.36	35	71.8	0.45	44
S6	63.2	0.41	38	72.9	0.47	48
S7	66.6	0.28	25	67.1	0.33	24
S8	64.2	0.33	41	69.6	0.48	57
S9	70.4	0.40	55	75.5	0.51	72
S10	72.5	0.32	39	73.3	0.49	54

The voting algorithm on average improved the TP and TN values and, consequently, the MCC factor and led to higher ITRs. Figure 4.17 demonstrates the MCC values from Table 4.6. Ensemble classification has higher MCC values in all subjects with smaller inter-subject variation (mean=0.467, var=.0042) in comparison to single SVM (mean=0.29, var=.0089). The result shows the higher robustness for the proposed method.

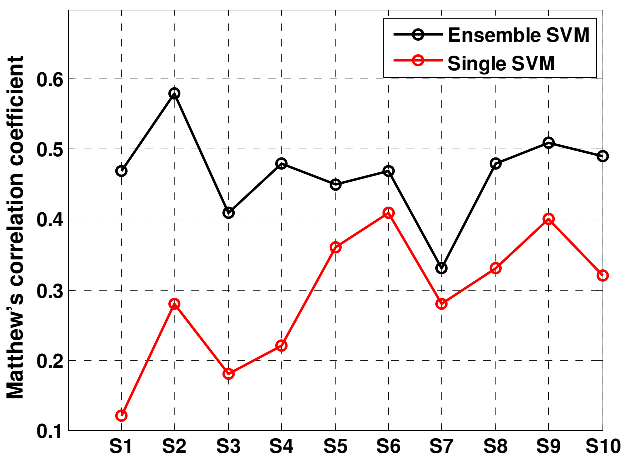


Figure 4.17: A graphical view of averaged MCC values for the current dataset calculated for ten subjects using two different classification techniques.

As described above, each feature vector was composed of the 4 power spectral and 6 HOS values, for each recorded channel. Here, using the feature selection technique we have pruned the original feature set by eliminating the possible redundant features. The word "redundant" means that the final classification performance will not be remarkably influenced by adding or removing them to/from the feature set. Therefore, we extracted and classified several feature groups in a reasonable range in order to find the optimum feature set for employing in the simulated online experiment. Different feature vectors were extracted based on mRMR algorithm using 1, 2, 5 or 10 features from each of the 7 channels (= 7, 14, 35, or 70 features). Table 4.7 shows different evaluation parameters for this study. Both the classification accuracy and MCC factor are maximized when the 14-feature subset was applied. The classification accuracy drops both in training and test by adding more features. Thus the extra features either contain redundant information or are not appropriate candidates for this application.

Table 4.7: Averaged classification accuracy for different feature selection techniques applied on the current dataset.

	Ensemble SVM				
Total number of Feature subset	**Sensitivity %**	**Specificity %**	**MCC**	**Test Accuracy %**	**Training Accuracy %**
7 features	68.18	82.29	0.51	74.80	77.59
14 features	69.23	85.71	0.52	75.25	78.15
35 features	66.67	82.35	0.46	72.60	77.10
All features	64.18	78.79	0.40	69.45	75.66

4.7. Discussion

In this chapter we introduced a practical BCI system for hand prosthesis control. The system was evaluated using a combined features space (power band and HOS features). The effect of a feature selection technique was investigated, which has two important advantages for designing every BCI. The new feature space has lower dimension and thus reduces the computational cost and at the same time improves the classification accuracy by ignoring the irrelevant features. Here the final feature subset was selected based on mRMR algorithm in an offline study. On the one hand the mRMR technique searches for the most independent features by minimizing the mutual information between the features and on the other hand it finds the most informative feature which leads to the higher classification accuracy. HOS features were mostly selected as the first or the second important features in each channel. It can be argued that these features along with common power band features are appropriate candidates for designing BCI systems.

For the classification purposes an ensemble classifier was proposed. By applying this method the combination of robot and the BCI system achieved a higher information transfer rate (ITR) and MCC. For a few number of subjects (S2, S4 and S5) the classifier tends to be biased towards a class. In general, the classifier has low variance in different runs, which shows the robustness of the entire system. The results suggest that the proposed system may be feasible for real life applications.

Chapter 5

BCI for gaming:
A four class BCI for Robot control[1]

5.1. BCI for gaming

The previous chapters explored BCIs which can be applied to assistive applications. Here we delve into another popular application of the BCI. In the past few years, emerging simple and cheap EEG recorders opened up the market and made BCI appealing for a new and large group of people who initially started to use BCI for fun and entertainment purposes (Brutsch et al., 2011). The main purpose of the new generation of BCIs is to expand the limited BCI application to more applicable scenarios in everyday life. The purpose of using BCI in the lives of healthy people is to render it a useful tool in the situations that the normal and conventional interfacing means such as a mouse, keyboard or joystick cannot be an option, i.e. space, military or marine projects (Bradley, et al., 2009) or to improve attention, concentration or motivation (Schuler et al., 2011). As an example, virtual keyboards allow users to compose phrases and sentences just by thinking, mouse control might facilitate the interaction with the computer programs and enable the subject to access the internet (Schalk et al., 2007) (Daly I., 2012). To date a few BCI prototypes are available in the domain of gaming and entertainment (Krepkiy, 2008) (Debener S. et al., 2012) (Kaplan et al., 2013). These systems use brain signals in addition to traditional physical and mental abilities to adapt and to control a game environment.

5.2. Multi-Class BCIs

One of the first steps in designing a BCI based on imagination is to select the number and the type of mental tasks that subjects should practice. It was found that classification

[1] Part of this work was published in: **M. Kh. Hazrati**, V. Subramanian and U. G. Hofmann," A Four-Class Brain Computer Interface for Robot Control", *ISSNIP Biosignals and Biorobotic Conference, Vitoria, Brazil, January 2011*

accuracy is affected by the type of mental task and is subject-dependent (Pfurtscheller et al., 1997). Physiological signals originating from thoughts corresponding to natural intent are preferable. Different mental tasks have been used in existing BCIs including relaxation, imagination of left hand, right hand, and foot movements, calculation, and spatial rotation (Naeem, et al., 2006). One of the first experiments done in this field was a simple application for moving a cursor on a monitor in two- (Vidal, 1973) and later in three-dimensional space (Taylor, 2002). The simple one-dimensional control can be interpreted as steering a robot vertically or horizontally. One of the first instances of robot control using a BCI was presented in 2004 by Millán et al. This three-class asynchronous BCI allowed users to give commands to a robot in real-time (Millan et al., 2004). Another three-class asynchronous BCI was used by Geng et al. to control a software-simulated robot (Geng et al., 2007). In 2008 the same group reported a four-class BCI based on two binary linear discriminant analysis (LDA) classifiers (Geng, et al., 2008). Bell et al. presented a noninvasive BCI to control a humanoid robot (Bell, et al., 2008). Similarly, a new concept the classification of the left and the right wrists in a four-class imaginary-based BCI was presented by (Vuckovic et al., 2006). This noninvasive BCI combines two binary classifiers. First, it classifies movements of left and right wrists, and then it classifies movements of wrist flexion and extension of the selected hand. Noninvasive methods however will probably have limited use of fine controlling; this could be compensated by combining with the robot intelligence (Taylor, 2002) (Lange et al., 2011). In the next section we investigate theoretically how the number of classes affects a BCI system in terms of information transfer rate (ITR) or signal to noise ratio (SNR) parameters and whether it is reasonable in practice to design multi class BCIs. Then our proposed BCI system is explained.

5.2.1 Multi class BCI and Information transfer rate (ITR)

It is estimated that the brain can handle over four hundred billion bits of information per second while using its peripheral nervous system which is directly affected by fatigue and mood (Dispenza, 2008). Raw EEG data obviously comprises much less information because it is the outcome of synchronized activity of diverse and random sources inside the brain. Based on some previous research reports, the amount of information decreases down to approx. 5 to 25 bits/min when one tries to extract mental tasks form EEG signals (Wolpaw et al., 2000). Some sparse reports in the literature, under restricted conditions and paradigms, report BCIs with high ITR (also called throughput) ranging from 30 bits/min (Blankertz et al., 2007) to slightly above 60 bits/min (Friman et al., 2007) or even 90 bits/min in a steady state visual evoked potential paradigm (Gao X. et al., 2003).

The binary decision structure is one of the most popular structures in designing BCIs, when only two tasks at each level are classified (Kronegg, et al., 2005) (Moore-Jackson et al., 2005)(Geng, et al., 2008). Moving from a binary decision to higher class systems is a logical approach to speed up the system by providing more choices for the user (Obermaier et al., 2001). Moreover the number of mental tasks used in similar research studies is limited to 5 (Schlögl et al., 2002). Obermaier et al. reported that by increasing the number of mental tasks, the classification accuracy decreases (Obermaier et al., 2001). Mason et al. categorized BCI in two major groups of synchronized and self-paced systems. In the first group, in which the majority of available systems are included, the number of tasks that the subject is asked to perform is limited (*N*) and is time lagged by the computer. Here we are interested in the information rate that can be transferred by synchronized BCI.

In Pierce's book (1980) "Introduction to information theory", the number of bits transmitted *B* was computed based on the number of possible targets *N* and the probability of hitting the target *P* (Arndt, 2001).

$$B = log_2 N + P log_2 P + (1-P) log_2 \left(\frac{1-P}{N-1}\right) \tag{5.1}$$

Here bit rate (bits/min), can be computed by dividing *B* by the trial duration in min. Based on this formulation, a common assessment of performance accepted by the BCI community is presented (Wolpaw et al., 2000). It is based on the assumption of unbiased BCI, in that all classes have the same probability (equiprobable classes) (McFarland, et al., 2003). Wolpaw et al. modified the formula in Eq. (5.1) by multiplying *V*, the application speed in trials/second. This assessment is not appropriate for self-paced BCIs (Moore-Jackson et al., 2005), but covers a wide range of BCI systems including imaginary-based BCI setups, where *V* can be tuned easily and represents the number of thoughts that can be recognized per second.

$$B = V \left[log_2 N + P log_2 P + (1-P) log_2 \left(\frac{1-P}{N-1}\right)\right] \tag{5.2}$$

According to this formula BCI performance may increase by increasing *N* which indirectly decreases the probability of hitting each target. Kronegg et al. have shown that Wolpaw's ITR is very close to Shannons's ITR when the number of targets, e. g. mental tasks, is smaller than five (Kronegg, et al., 2005). The results from previous studies indicate that there is a tradeoff between the two parameters *P* and *N*. Dornhege et al. also studied the effect of the error rate and the number of commands in *ITR*. Based on their finding, *ITR* depends on the number of targets as well as on the error rate; however the effect of the first parameter (*N*) is not noticeable compared to the accuracy measure (Dornhege et al., 2003). Schlögl et al. derived an equation to calculate the information rate in a BCI system which worked with two-class movement imaginary data. They applied the entropy difference as a measure of the separability of two classes of EEG patterns using a linear classifier LDA (Schlögl et al., 2002). The authors stated that one cause of unsuccessful BCI is an insufficient amount of information from feedback, i.e., from the output of the classifier.

McFarland et al. discussed the selection of effective parameters in a cursor control paradigm (McFarland et al., 1997). His group carried out an experiment over 6 subjects with different target numbers and variable trial durations. The results show that a greater number of targets could increase system performance, since more targets provide more information. At the same time, a greater number of targets could decrease system performance by decreasing accuracy (McFarland, et al., 2003). Figure 5.1 demonstrates the theoretical relation between the information transfer rate *B* and different accuracies with more than chance levels (*100/N*) for various numbers of classes based on Eq. 5.2. Here *V* was set to *1* trials/second. Later in this chapter, we will calculate the ITR in our BCI application in a similar way.

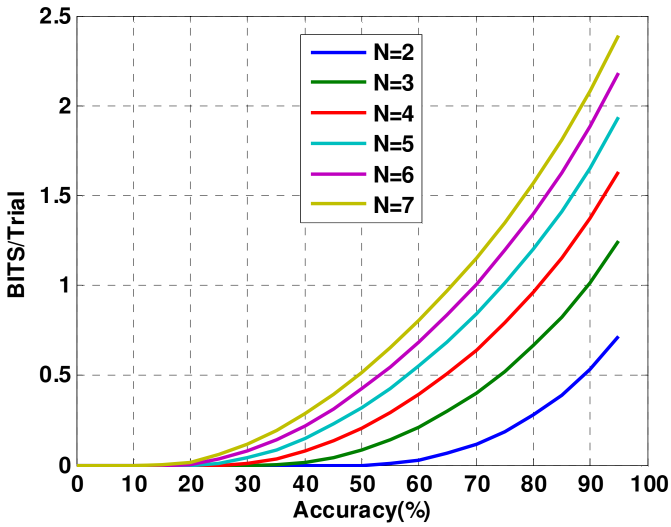


Figure 5.1: Information transfer rate versus accuracy in a BCI system for (N =2 to 6) mental tasks (imagination of movement) according to Eq. (5.2)

It is clear that every additional EEG pattern to be classified decreases the chance level (*100/N*). Increasing the number of mental tasks or classes, however, also decreases the reliability of classification (Obermaier et al., 2001). This might be due to increasing areas of overlap features in the feature space which increases the probability of misclassification (Obermaier et al., 2001).

In their cursor control experiment, Obermaier et al. achieved information transfer rates varying from 0.42 to 0.81 bits per trial and concluded that the upper limit of different mental tasks for a BCI system is three (Obermaier et al., 2001). They measured the highest classification accuracy for two classes of control (*N=2*). Kronegg et al. presented a model *ITR* in an EEG-based synchronized BCI (Kronegg, et al., 2005). Based on their model and an experimental validation, they concluded that the ITR improvement in BCI is not significant from *N=2* to $N = N_{opt}$ and the optimal number of mental tasks N_{opt} depends on the BCI design and user experience. Therefore the modified formula is proposed as:

$$B_{max} = \alpha_s.V\left[log_2N_{opt} + Plog_2P + (1-P)log_2\left(\frac{1-P}{N_{opt}-1}\right)\right] \tag{5.3}$$

Where α_s is the subject specific performance which varies between 0 and 1 and B_{max} is the maximum bit rate achievable for the subject. $\alpha_s = 0$ can describe the users who are categorized under "BCI illiteracy" (Vidaurre et al., 2010).

5.2.2 Multi class BCI and Signal-to-noise ratio (SNR)

Considering N-class BCI applications, one can claim that in each trial the output of the classifier in a BCI system contains two kinds of information. The first element is entirely dependent on the process of mental task ("signal") and the second part is independent or irrelevant to that process ("noise") (McFarland, et al., 2003). In a BCI system, the main goal of signal analysis is to maximize the SNR in the EEG or more accurately to find the best set of feature vectors and classifiers to discover the user's thought (Kronegg, et al., 2005). Based on this assumption, the synchronized BCI can be modeled as depicted in Fig. 2.13.

Features are extracted from EEG data in each trial. If we assume that *p* features are extracted from the time segment *t* related to class C_1, then $F_t^{C1} = [a_{1,t}^{C1}, a_{2,t}^{C1}, \dots, a_{p,t}^{C1}]$ is a vector of features for this segment. The set of contaminated features of interest in the receiver x_t is the summation of original features plus the unwanted signal, $x_t = v_t + s_t$, which reflects the side effect of background activity of the brain in the EEG signal in addition to the effect of external noise. We can assume that the channel noise $V \sim N(0, \sigma_v^2)$ is a white Gaussian noise (WGN). Therefore the probability of a correct decision according to the features extracted from $\mathrm{x_t}$ is defined as Gaussian distribution $p(x_t|s)$ (Eq. 5.4). Although in practice not every feature set is distributed Gaussian, this assumption is widely accepted in the BCI community (Schlögl et al., 2002) (Obermaier et al., 2001).

$$p(x_t|s) = \frac{1}{\sqrt{2\pi}.\sigma_v} \sigma^{\frac{-(x_t - x_s)^2}{\sigma_v^2}} \tag{5.4}$$

So over the decision space R_i, the probability of a mental task to be recognized is:

$$p(C1|F_t^{C1}) = \int_{R_i}^{\cdot} p(x_t|s)\, dx \tag{5.5}$$

Consequently the mean of all possible feature-classifier sets constitutes the SNR formulation (Schlögl et al., 2002):

$$SNR = -10\, log_{10} \left(\frac{\sigma_s^2}{\sigma_v^2} \right) \tag{5.6}$$

SNR is an important parameter in Shannon's calculation of information. In Wolpaw's ITR, the value of SNR significantly affects the acceptable number of mental tasks in a BCI system. As depicted in Fig. 5.2, *N* may rapidly grow with a higher information rate, when SNR is increasing. Schlögl et al. calculated *SNR* at each time segment *t* for a two-class BCI (left versus right hand imagination) (Schlögl et al., 2002) as follows:

$$SNR_t = \frac{2 * var(D_t^{C1+C2})}{var(D_t^{C1}) + var(D_t^{C2})} - 1 \tag{5.7}$$

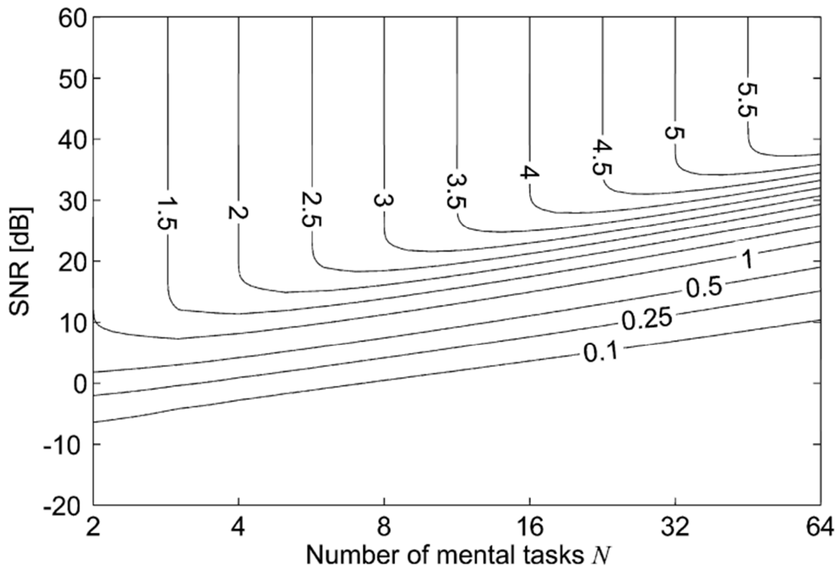


Figure 5.2: Contour plot of Wolpaw's ITR in bits/trial depending on SNR and number of mental tasks (Kronegg, et al., 2005)

Here I generalize the above concept in order to derive a formulation for multi-feature space. Assuming D_t^{C1} is the discrimination function for data related to the class *C1* and it is directly the result of the classifier. In the ideal case, it contains all the segments assigned to this class but in practice due to noise and misclassification it may have some segments which are not directly related to the class. In the general case we have:

$$D_t^{C1,C2,\dots,CN} = f([a_{1,t}^{C1}, a_{2,t}^{C1}, \dots, a_{p,t}^{C1}, a_{1,t}^{C2}, a_{2,t}^{C2}, \dots, a_{p,t}^{C2}, \dots, a_{1,t}^{CN}, a_{2,t}^{CN}, \dots, a_{p,t}^{CN}], W_T) \qquad (5.8)$$

Output D is a time-varying function of features and classifier parameters W_T and indicates information about the classifier's decision. Hence, mutual information can be calculated between the BCI output and D. The mutual information I is determined by calculating the difference between the entropy for the total variance and the within-class variance of D: $I = H(x) - H(x, C)$. The entropy of a Gaussian process is (Arndt, 2001): $H_x(X) = -2 * ln(2\pi e \sigma_x)$. For each time segment of data this information can be calculated using the following equation (Schlögl et al., 2002):

$$I_t = 0.5 * ln(1 + SNR_t) \qquad (5.9)$$

In the case of the two-class BCI discrimination, 100 % accuracy (0 % error) would provide one bit of information when there is no noise input to the system (Schlögl et al., 2002). Consequently using the channel model (see Figure 2.14), we can expect that an increase in the number of mental tasks *N* results in an increase of the *ITR*, subject to the sufficiently high *SNR* (Kronegg, et al., 2005). It implies that only BCIs with high *SNR*, i.e., good accuracy, will noticeably benefit from an increase of the number of mental tasks (Kronegg, et al., 2005).

5.3. Project description

Our first games-based BCI is a four-class control application for steering a miniature educational robot. The main goal of this project is to develop an interactive environment that allows the user to freely decide which action she/he wishes the robot to perform. The robot executes commands online and navigates accordingly and is able to send the information (gyro and images) to a computer. Based on the past success of differentiation between movement imagination classes and the convenience and practicality (Millan et al., 2004) (Naeem, et al., 2006), we have chosen to implement a four-class BCI in which subjects steer a robot to a destination by performing limb imaginations. Each mental state corresponds to a different robot command: left hand movement to turning left, right hand movement to turning right, foot movement to moving forward, and an idle state to stopping. In the following elements of the proposed system are described and finally the functionality of the entire system is evaluated.

5.3.1 EEG recording Hardware and Software

To implement robot control by BCI, appropriate and optimized computer software is required. In our case, we used a gUSBamp (g.tec, Guger Technologies, Graz, Austria) MATLAB® application programming interface under Microsoft® Windows XP for online data acquisition. MATLAB® R2009b (The Mathworks, Inc.) was used for filtering and ocular artifact suppression, feature extraction, feature selection and classification. BlueSoleil Bluetooth® drivers were used to communicate with a miniature e-puck robot (Mondada et al., 2009), and a self-developed graphical user interface (GUI) provided an interactive environment. EEG signals were recorded with an EEG-amplifier at a sampling rate of 128 Hz. Data collection experiments were conducted in the institute for signal processing (ISIP) at University of Lübeck, Germany.

5.3.2 Experimental Paradigm

Experimental runs were carried out with three able-bodied volunteer subjects (two male, one female). In preparation for each run, subjects were seated 0.7 meters from a computer monitor with arms in front and palms flat and facing down on a desk in front of them. A 65-position EEG cap (g.EEGcap, Guger Technologies, Graz, Austria) was fixed on the subject's head, and eight gold electrodes (g.EEGelectrode Au, Guger Technologies, Graz, Austria) were placed in the following locations according to the International 10-20 system: C3, C4, Cz, C5, C1, C6, C2, and Fp1. The electrode for recording eye-blink artifacts were placed at the Fp1 position (on the forehead) above the left eyebrow line. All recording channels were referenced to the right earlobe.

Our experimental paradigm consisted of one offline classifier training run followed by two online classifier evaluations / robot control runs (Figure 5.3). The offline run consisted of 50 feedback trials. Each trial lasted ten seconds, the first five devoted to the imagination of a particular limb movement and the last five to relaxation. During each trial an image of a black circle was continuously displayed for use as a visual fixation point to prevent eye movement artifacts. Prior to the start of each trial, a four-second-countdown cue was given to the subject signaling that the imagination stage was about to begin. One of three images – a left arrow, a right arrow, or a drawing of a foot – was shown for five seconds, and the subject was asked to perform a corresponding imagination – left hand movement, right hand movement, or right foot movement. After five seconds, the picture disappeared, and a screen with only the focal point was shown, indicating that the subject should relax. The information derived from offline experimentation was used to initialize our online setup. Shortening the training time is one of the concerns in designing a BCI. In our design subjects were asked after every 2 minutes of the training phase if they were willing to continue training. This question was embedded in this way to prevent collecting data while the subject is tired or for any other reason an interruption was needed.

Two types of online tests were performed to determine how accurate the training was and how well the user could control the robot. In the first setup, the subject was given a visual cue to begin imagining one of the four actions for which he or she was trained. He or she was given five seconds to imagine the mental state and was instructed to maintain the same thought for the length of the trial. The subject was then given five seconds to relax during which processing of signals was performed and the execution of the robot action took place. The commands in this experiment were set in a logical order to move the robot from point A to point B in an 8-shape grid surface shown in Figure 5.4. For each participant 50 online trials in this type of online test were recorded. In the last session of the experiment for each subject, we used a part of the online data (2 s window from the middle of every trial) to continuously update and improve the classifiers' parameters, which are C and γ for the SVM classifier. The adaptation was continued during all 50 trials of the first online test.

In the second setup, the subject was asked to navigate from point A to point B along the grid on his own. Similar to the first setup, the user was shown a black focal point and was given five seconds to imagine the mental state corresponding to the action he or she wished the robot to perform. After these five seconds, the robot would execute a command while the user was given five seconds to relax. The robot continued to execute the command for five seconds. The data recorded in these series of experiments are not analyzed in this work.

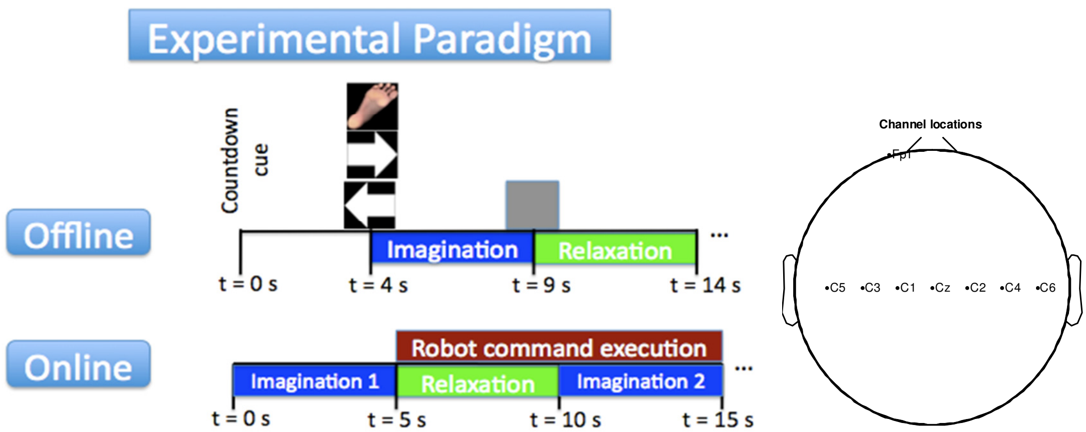


Figure 5.3: Experimental paradigm of offline and online robot control runs. Electrode locations C5, C3, C1, Cz, C2, C4, C6 and Fp1.

Biofeedback was provided during all online experiments, which helped the subjects to learn how to modify their brainwave activity to improve attention. Here the direct output of the decision making module was sent to the robot.

5.3.3 E-Puck Robot control

The e-puck mini mobile robot (about 7 cm diameter) was originally developed at École Polytechnique Fédérale de Lausanne (EPFL) for teaching purposes. The e-puck was designed under open source hardware and software policy. It means that the user has access to every electronic part and will be able to modify or extend the device.

According to the official e-puck community website (www.e-puck.org), the e-puck robot contains several sensors and actuators and its configuration can be listed as following: A Bluetooth interface, a remote control IR receiver, a VGA camera, 9 LEDs, 8 infra-red sensors, two stepper motor wheels, 3 microphones, a speaker, a 3D accelerometer, and a 16 position switch. A commercial development environment called Webots^{TM} has been released for the purpose of simulating, programming and modeling the mobile robots and supports the e-puck robot. We have used a free version of this software for fast prototyping and simulation of the robot. During the online experiments, the ePic2 framework for MATLAB provided interaction with the e-puck.

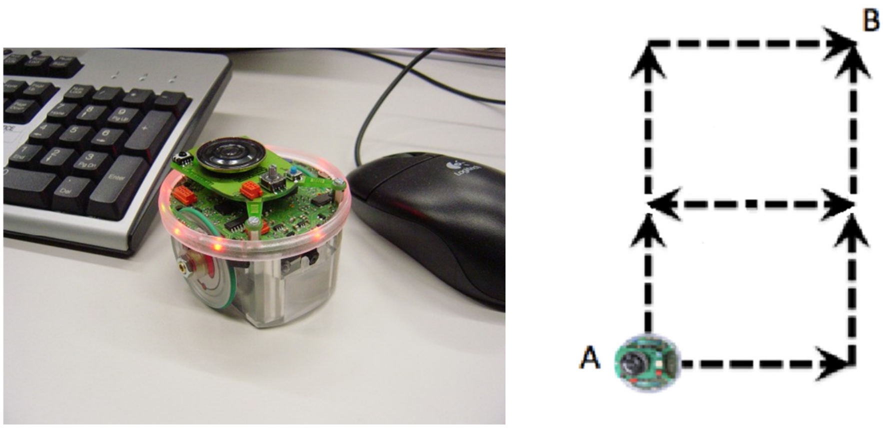


Figure 5.4: (a) A close view of the E-puck robot used in the BCI setup, (b) the subject was asked to navigate from Point A to Point B along an 8-shape grid

In the current project E-puck was set to receive continuous commands through Bluetooth from the computer. Four different codes corresponding to the three different movements and the stop command were set. A wink command was used in the beginning of the experiment to check the connection between the computer and E-Puck. Here is the MATLAB code for setting the speed of the robot in each case.

```
%Assuming one case for each of 3 classifiers, give a command to the robot.
    switch command
        case 1 %left
            ePic = set(ePic,'speed', [0 50]);
            ePic = update(ePic);
        case 2 %right
            ePic = set(ePic,'speed', [50 0]);
            ePic = update(ePic);
        case 3 %foot
            ePic = set(ePic,'speed', [50 50]);
            ePic = update(ePic);
        otherwise %idle
            ePic = set(ePic,'speed', [0 0]);
            ePic = update(ePic);
    end
```

Both advanced robotic and machine learning techniques are required for humans to continuously control a mobile robot. For example in our setup, the robot does not execute the forward command if a barrier is detected. It was set to record and send a photo to the computer in this situation.

5.4. Evaluation of the recorded data

5.4.1 Averaged potentials

Figure 5.5 demonstrates grand average ERPs calculated for the current dataset during the offline experiments using EEGLab toolbox. Each color represents a different class of trials. Here we consider the entire offline trial which lasted 14 s, starting with a countdown cue at zero point, counting back from 3 to 0. The onset of the imagination trial is at 4 s. At this time point an arrow or the foot picture appeared and remained on the screen for 5 s. At time point 9 s the relaxation phase started when the previous image was substituted by a gray square. In Figure 5.5, units on x-axis are ms and units on y-axis are µV. The first fluctuation is a positive peak at around +300 ms, which can be seen over all recording sites possibly due to the appearance of the first number on the screen. It reached its maximum at C5 with 3.1 µV. The second evoked potential happened at +4550 ms for both hand movement imaginations; it reached its maximum at Cz with 7.2 µV and 6.5 µV for trials with the right and left arrows as cue, respectively. In all recording sites the evoked potentials related to the left hand imagination show similar timing and have larger amplitude in comparison to the corresponding right hand or foot trials. For foot imagination trials, the corresponding evoked potential happened earlier at 4380 ms with smaller amplitude of 6.1 µV, again maximum at Cz followed by a negative peak at around 5 s.

The third evoked potential in Figure 5.5 is a positive peak at 9250 ms for the foot imagination trials and at 9500 ms and 9700 ms for the right and left hand imagination trials.

This time the maximum is larger for the right hand imagination trials over the entire recording locations. These two large positive peaks are most probably a P300 component according to their topography.

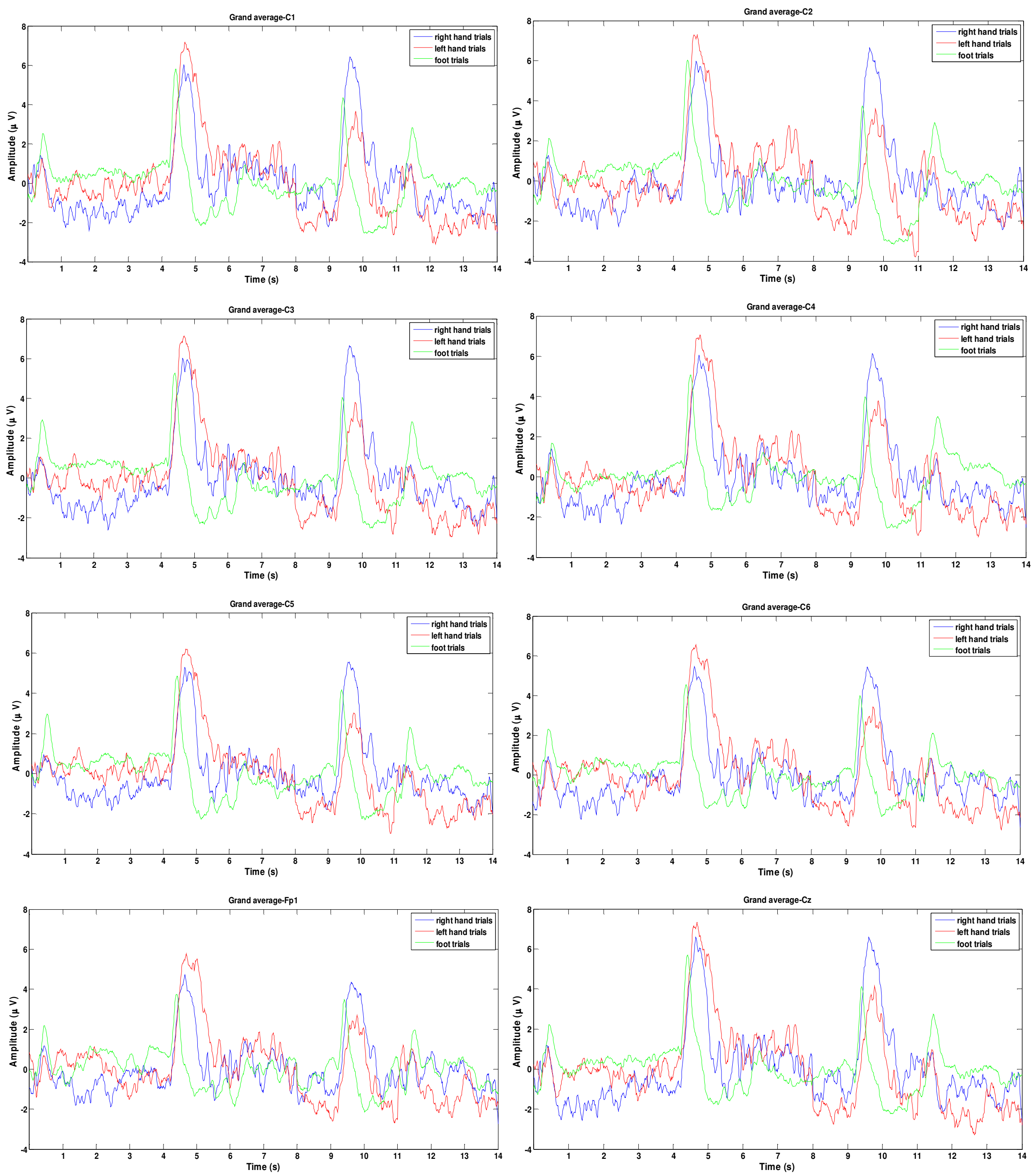
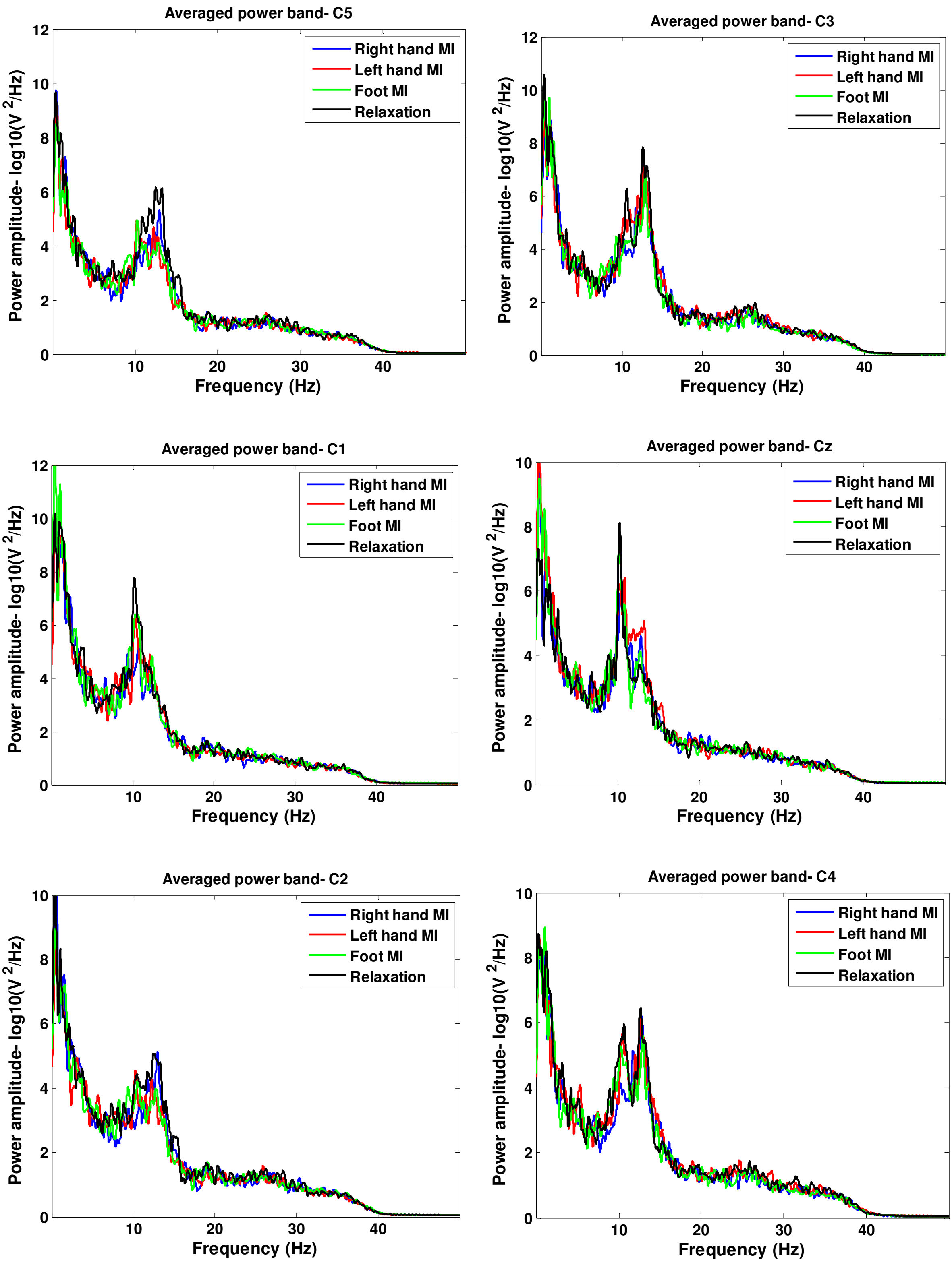


Fig. 5.5: Grand average ERPs averaged over all trials for imagination limb movements calculated for C5, C3, C1, Cz, C2, C4, C6 and Fp1. A countdown started at the beginning of each trial (displaying the numbers 3-2-1-0). At 4 s a cue appeared on the screen showing randomly one of the images corresponding to the imagination of movement. The entire dataset includes imagination of right hand, left hand and foot movements. At 9 s a gray square appeared on the screen as a sign for the relaxation interval which lasted 5 s (9 s – 14 s). Amplitude units are μV, x-axes range from 0 ms (left) to 14000 ms (right).

At Fp1 all positive peaks are smaller in comparison to the central recording sites. In the middle of the relaxation phase for the foot imagination trials, there is another positive peak happening around 11.5 s after the start of the trial, which could be due to an artifact (or even due to the actual robot movement). This fluctuation exists over all recoding locations. Since the recording sites were not diversely distributed in this experiment the ERP figures seem to be more similar over all recording locations, i.e. C5, C3, C1, Cz, C2, C4, C6 and Fp1 .

5.4.2 Power band spectrum

The average power band spectrum is depicted for all four conditions in Figure 5.6. Signals were acquired from each individual channel, including eight recording sites over the motor cortex C5, C3, C1, Cz, C2, C4, C6 and Fp1. The amplitude is calculated in log10 (V^2/Hz) unit. Data from relaxation states are presented in black and data from movement imagination states are presented in color (blue, red and green). In general there are more fluctuations in Figure 5.6 in comparison with similar analysis on the previous datasets. The reason is the lower number of trials available in this dataset (N=3 subjects $\times$ 3 sessions $\times$ 50 offline trials). The calculation is similar to the one described in 3.5. The alpha peak can be seen over all recorded sites. In the current dataset, values of the EEG power in delta (1-3 Hz) and alpha bands (8-12 Hz) of the different imagination phases are comparable to each other over all recording sites. Similar to the previous datasets, in the current dataset values of the EEG power in delta (1-3 Hz) and alpha bands (8-12 Hz) of relaxation phase were found to be slightly higher than those in imagination phases over all recording sites. The alpha peak is not strong at Fp1. The maximum exists at 10 Hz at Cz for the relaxation phase. Over C3, C4 and C4 another peak can be seen in the higher frequency (around 13 Hz) in the data corresponding to relaxation in comparison to the imagination trials. Over C3 and Cz imagination of foot has higher magnitude in delta band. There is a slight increase in the power of beta band (around 25 Hz) in the imagination phases over C3, C4. No distinguishable changes can be seen in the gamma band (30-40 Hz) in average.


Averaged power band- C5
Averaged power band- C3
Averaged power band- C1
Averaged power band- Cz
Averaged power band- C2
Averaged power band- C4
Right hand MI
Left hand MI
Foot MI
Relaxation
Power amplitude- log10(V 2/Hz)
Frequency (Hz)
0
2
4
6
8
10
12
10
20
30
40

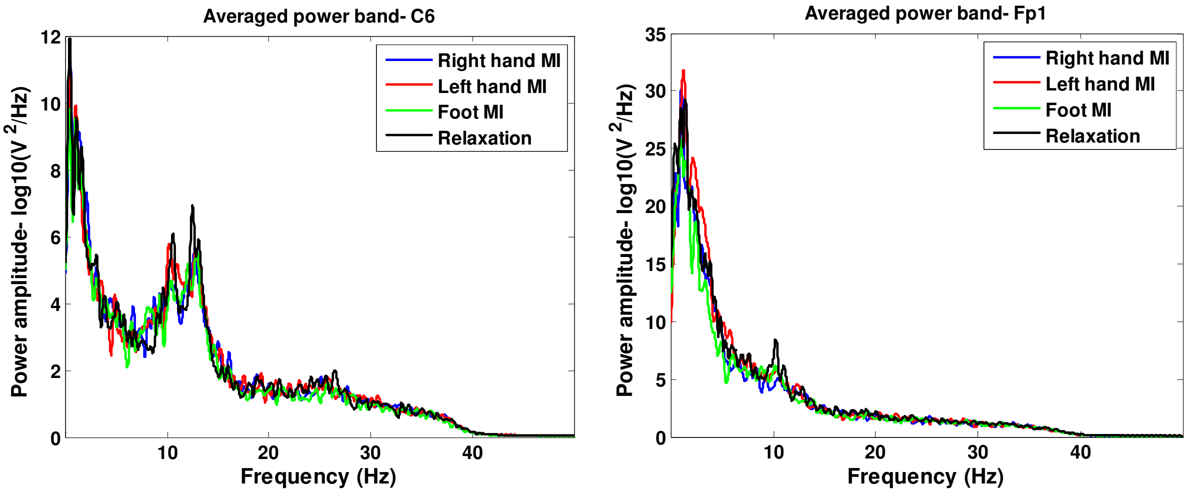
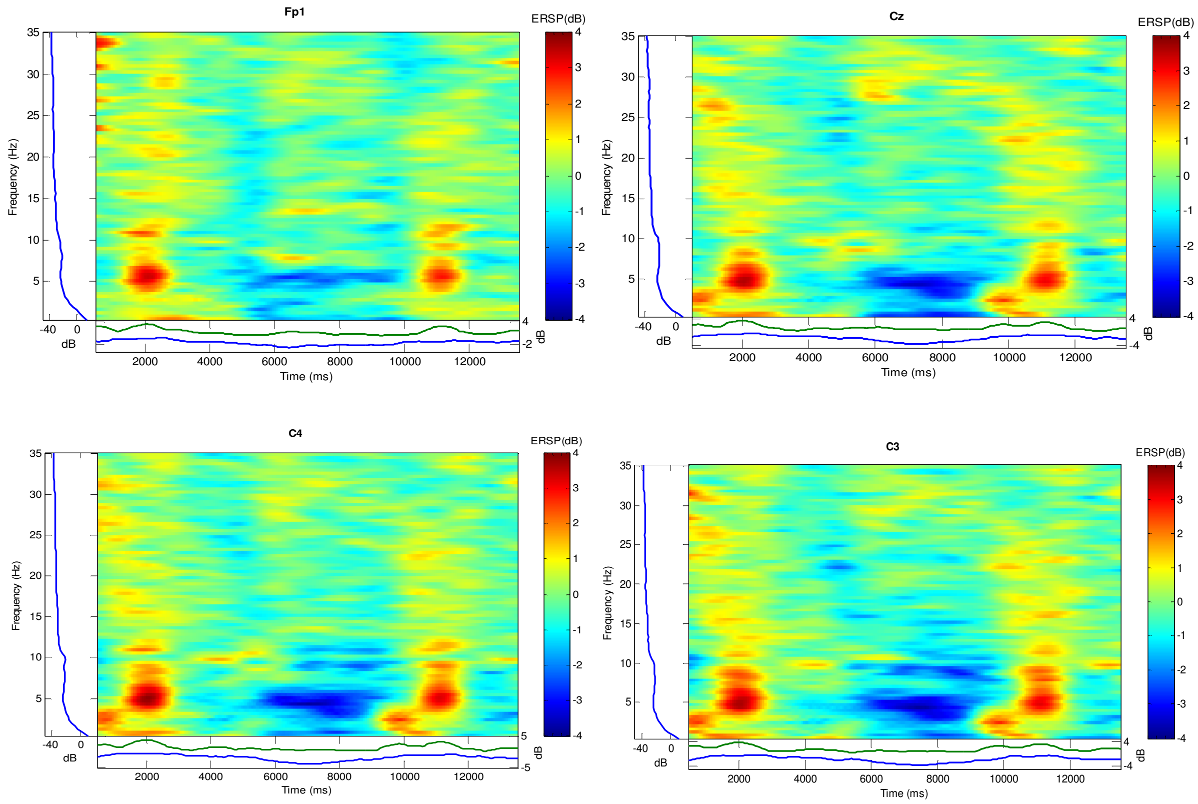


Figure 5.6: Average power band over the entire dataset for the three different imagination tasks and the relaxation state, displayed for C5, C3, C1, Cz, C2, C4, C6 and Fp1. Unit on x-axis is Hz, unit on y-axis is V^2/Hz.

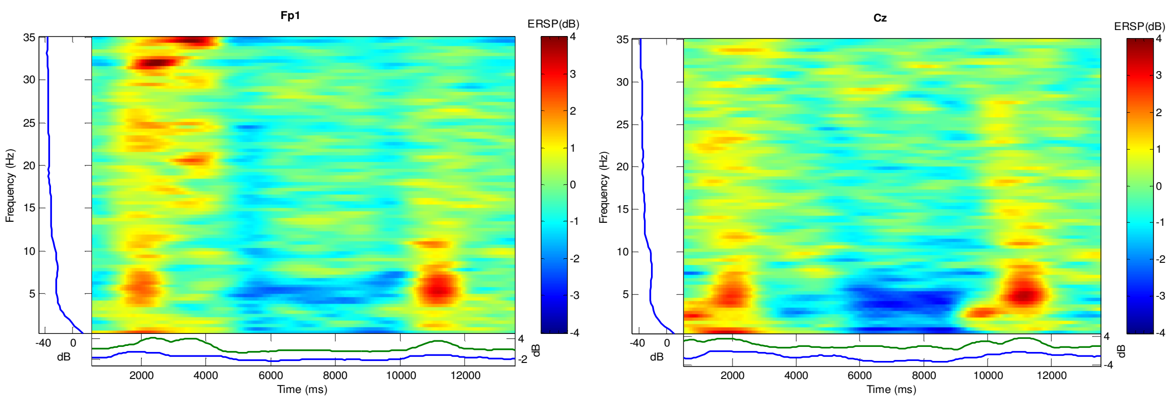
5.4.3 ERSP: Time-Frequency analysis

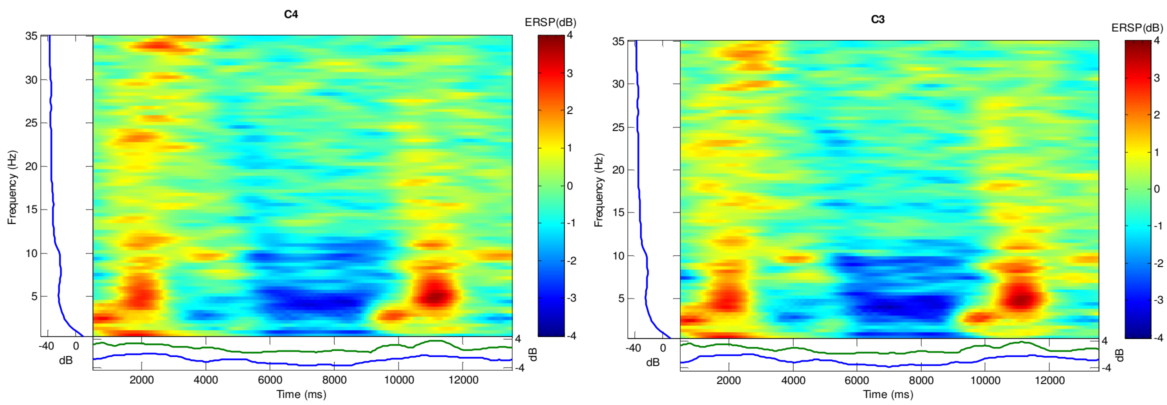
Figure 5.7 depicts ERSP analysis for some selected electrodes (C3, C4, Cz and Fp1) averaged over all trials and subjects of the current BCI experiment. Three groups of ERSP plots are depicted in Figure 5.7; each is calculated using the trials for the imagination of different limb movement. In the first group of subplots the ERSP analyses for imagination of right hand movement can be seen. An increase in power of theta and alpha band exists over all four plotted channels from the start of the trial till the onset of the imagination of movement at 4 s. Some sparse increases can also be seen in higher frequency bands at this period. At time point 4 s, when the cue was displayed on the screen till around 10 s, a noticeable decrease in theta bands over all channels and also in alpha band in C3 and C4 can be observed (overlying left and right hand-motor cortices). At 9 s the cue disappeared and a black dot remained on the screen for another 5 s. A pattern similar to the first 4 s of the trial can be observed at this interval in all ERSP plots. The second four subplots in Figure 5.7 show the similar analyses for the imagination of left hand movement. Here the ERD in C4 (overlying right hand-motor cortex) is clearer than the corresponding one in the first group, in accordance with the contralateral organization of movement control. Gamma ERS before the start of the movement imagination exists in all channels. It has its maximum at Fp1 and is probably due to an artifact. ERSP plots related to the imagination of foot movement can be seen in the last four subplots. The ERD and ERS fluctuations are not as clean and the first and the second group. In C4, an alpha ERS exists at 4 s followed by ERD in alpha and beta band between 5 s and 10 s. An ERS in the theta band can be seen at the beginning of the trial at 1 s and also at time point 10 s over all four depicted sites. An alpha ERD between 4 s and 10 s exists at C4. A shorter Beta ERD starts at 4 s and lasts till 8 s. By the start of the relaxation phase in this group a theta ERS (9 s – 11 s) and another alpha ERS (in the last second) can be seen over are presented channels. There is no clear difference between left and right scalp sites in this case, among others simply for the reason that participants were free in imagining movements of either their left or their right foot.

Imagination of right hand movement



Imagination of left hand movement





Imagination of foot movement

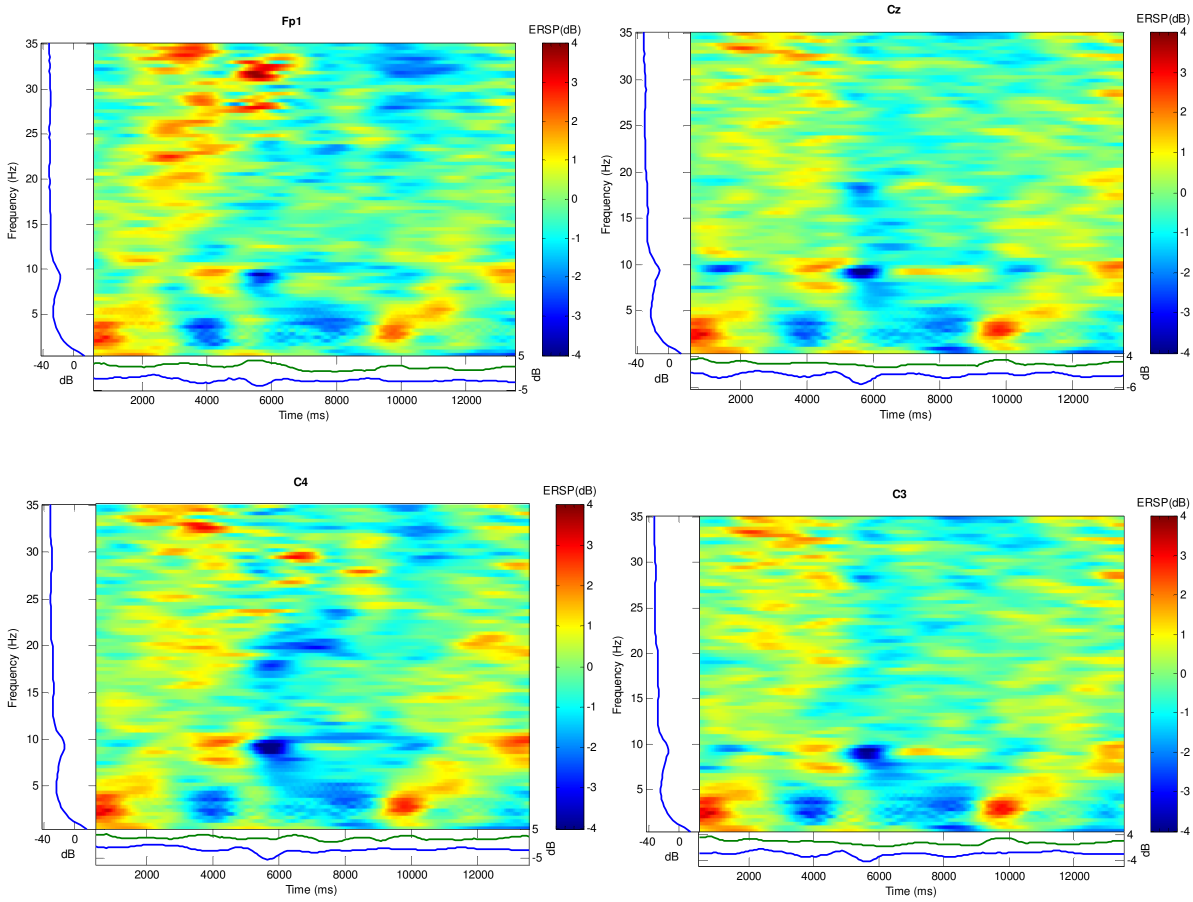


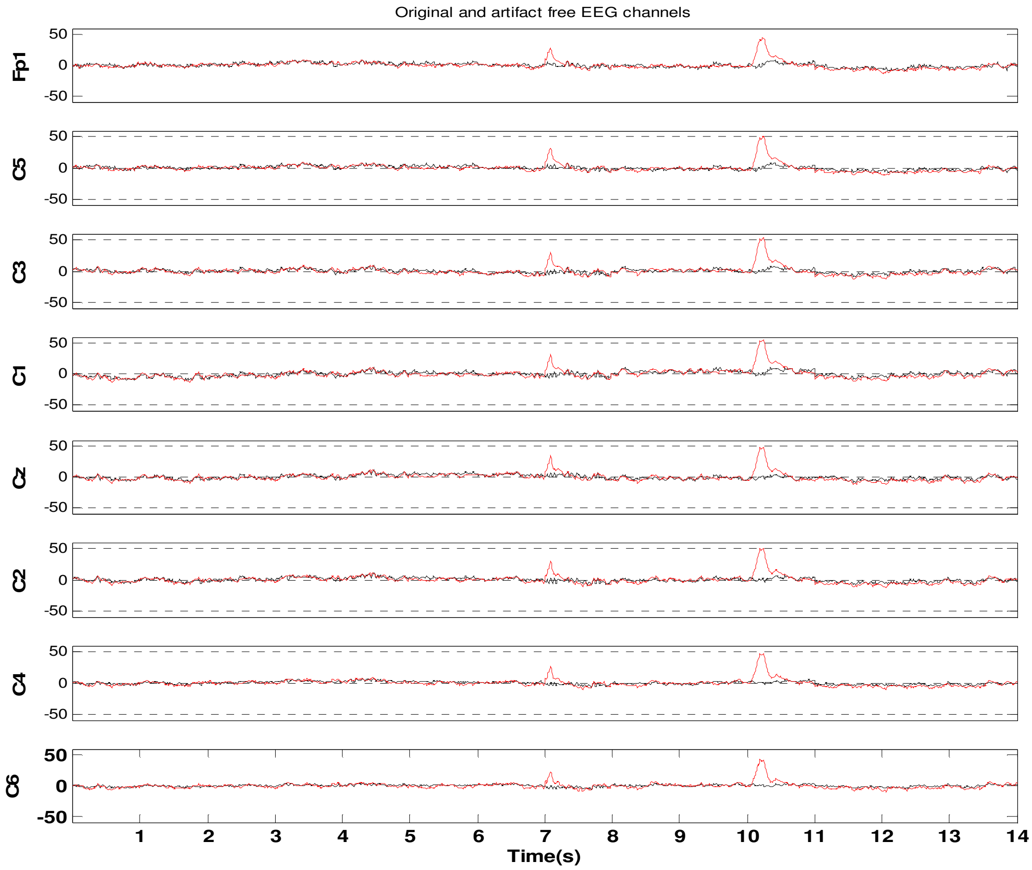
Fig. 5.7: Grand average- ERSP averaged over all trials for imagination of the limb movements calculated for C3, C4, Cz, and Fp1. A countdown started at the beginning of the trial. At 4 s a cue appeared on the screen showing randomly one of the images corresponding to the imagination of movement. Each four subplots correspond to the imagination of right hand, left hand and foot movements, respectively. At 9 s a gray square appeared on the screen as a sign for the relaxation interval which lasted 5 s (9 s – 14 s). Amplitude units are µV, x-axes range from 0 ms (left) to 14000 ms (right). Unit on x-axis is *ms*, unit on y-axis is *Hz*. The color bar represents the absolute values of ERSP in dB.

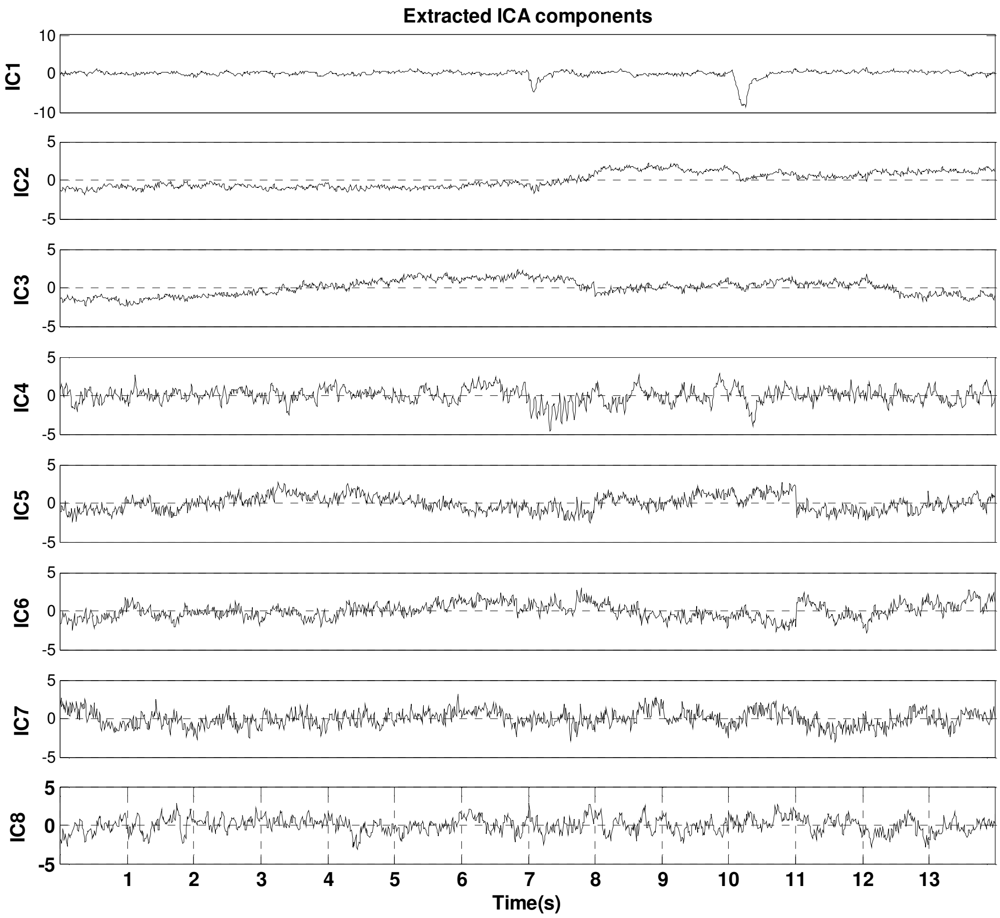
5.5. Materials and methods

For this application we combined the methods of simple feature extraction and ensemble classification. Preprocessing consists of filtering and automatic artifact removal. In contrast to the previous simulated studies feature selection was done solely in space-frequency domain. In this study, we demonstrated that by use of modern machine learning techniques even untrained subjects could achieve the satisfying BCI performance.

5.5.1 Preprocessing the data

Preprocessing consisted of filtering and artifact removal. The signals were filtered using a 10th-order zero-phase Chebyshev low-pass filter with a cutoff frequency of 40 Hz. In each trial the mean values over time of the imagination and relaxation signals were calculated and subtracted from each respective signal. Similar to the first application described in chapter 3, eye blink artifacts were removed using the *fastICA* algorithm in an online scheme. In this experiment we included the eyeblink artifact information detected by the electrode at Fp1. The reason was the topography of electrodes that were covering only the motor area, so it was assumed that the method without direct EOG recording cannot possibly extract enough independent sources from the brain activity. After independent components were retrieved, the component with the largest kurtosis value was removed as long as it exceeded a preset threshold value defined for each individual as described in chapter 3. The threshold values were set at 5.7, 6.5 and 6 for the three subjects estimated from the offline experiment data. Once the blink artifacts were removed, the remaining components were recombined, representing the artifact-free signals for each trial. Figure 5.8 shows a sample of EEG recorded during a 14 s trial contaminated with artifacts at 7 s and 10 s. Cleaned data using fastICA algorithm is plotted in black. Figure 5.8 (b) shows the spatial distribution of independent components for this trial. The first IC with kurtosis 7.3 was selected as artifact component.


Original and artifact free EEG channels
Fp1
C5
C3
C1
Cz
C2
C4
C6
50
0
-50
1
2
3
4
5
6
7
8
9
10
11
12
13
14
Time(s)


Extracted ICA components
IC1
IC2
IC3
IC4
IC5
IC6
IC7
IC8
10
0
-10
5
-5
1
2
3
4
5
6
7
8
9
10
11
12
13
Time(s)

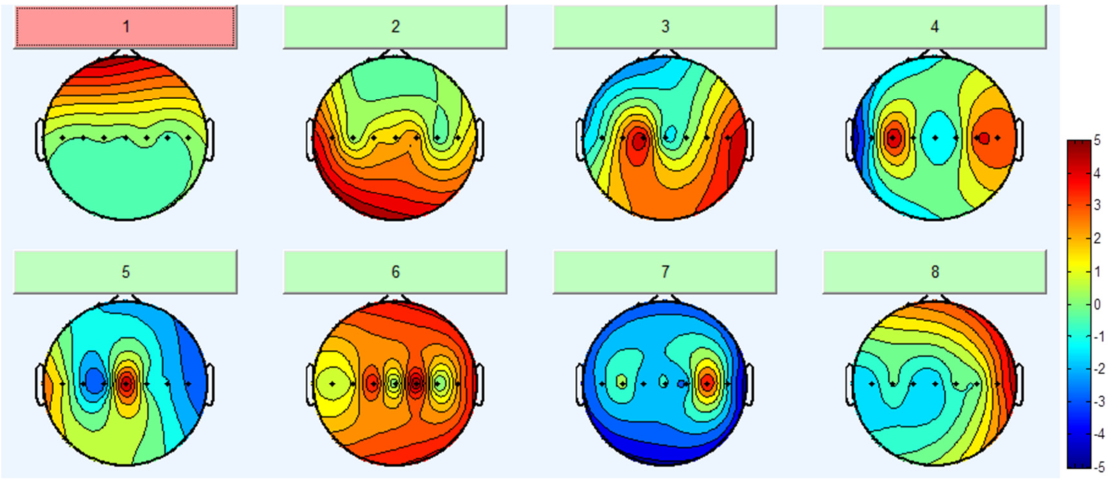


Figure 5.8: Artifact rejection in a 14 s window. (a) Contaminated EEG signal (red line) and artifact free signal (black line). Unit on y-axis is μV and unit on x-axis is s. (b) Artifact components for the same interval (IC1 is recognized as artifact component K=*7.3*), unit on y-axis is μV (c) 2D plot of independent components over the scalp. (The Fp1 site is on the circle that symbolizes the scalp, left from the nose). Scale is minimum-maximum for each plot, with red and blue denoting the two polarity extremes.

5.5.2 Feature extraction and feature selection

Band power features were extracted for each trial from the imagination and relaxation signals acquired from each individual channel excluding Fp1, which was not processed. The discrete Fourier transform of the imagination and the relaxation signals were computed for each trial, with a frequency resolution of 0.086 Hz, given by the analog-digital conversion rate of 128 sample/s and the 5 s length of the analyzed epoch, and the power in positive integer frequencies from 1 Hz to 35 Hz was extracted. The signals were first padded with zeros up from 640 (5×128) samples to 8192 (2^{13}) samples to increase the resolution of the Fourier transform. Magnitudes of the Fourier transform were squared and the logarithm was obtained. Similar research has already been done in extracting these features to distinguish between the three classes we are using and idle (Pfurtscheller et al., 1997). The amplitudes of the EEG signals in selected frequency bands are measured and translated into a device command, in this case right, left, forward, and stop commands. For each channel the magnitudes of the power of the 35 frequencies were normalized in the range of 0 to 1. Prior to the construction of classifiers, the continuous (non-integer) feature values of each class in the six pair-wise comparisons of the 4 classes shown in Figure 5.3 (left arrow, right arrow, foot, relaxation) were input as arguments to the minimum Redundancy Maximum Relevance (mRMR) feature selection algorithm (Peng et al., 2005). Eight features were selected for each channel, and of those, the top five were selected to generate SVM classifiers. The five most informative features selected for each subject from each EEG channel during the training phase of right hand imagination vs. relaxation are listed in Table 5.1. The values are the frequencies in the power band selected out of 35 features (1 Hz – 35 Hz) using mRMR method. As can be seen, the set of selected features is subject- specific and varies over spatial recording sites.

Table 5.1: The five most informative features selected for each subject from each EEG channel during the training phase (right hand imagination vs. relaxation). The values are in Hz.

	EEG channels						
	C5	**C3**	**C1**	**Cz**	**C2**	**C4**	**C6**
S1	2, 10, 12, 13, 34	3, 10, 12, 14, 32	2, 3,10, 12, 17	2, 4, 11, 15, 22	7, 9, 14, 19, 23	5, 10, 12, 16, 24	7, 9, 11, 14, 27
S2	4, 12, 14, 29, 33	2, 12, 15, 14, 33	3, 4,12, 14, 21	3, 4, 13, 16, 27	5, 11, 13, 20, 25	5, 9, 14, 21, 27	5, 9, 12, 16, 31
S3	2, 11, 13, 17, 30	3, 10, 12, 14, 30	4, 9,12, 18, 26	4, 6, 14, 15, 24	5, 10, 12, 16, 24	4, 11, 12, 16, 25	7, 10, 14, 18, 25

5.5.3 Classifier construction, voting algorithm and output to robot

Features selected for each two classes were fed to an algorithm used to generate classifiers. One SVM classifier was generated for each possible comparison of features and for each channel, yielding a total of forty-two classifiers (six comparisons and seven channels). For optimization, a grid search was used to loop over values of the SVM cost parameter C and the gamma parameter. Classifiers, their accuracies, and the selected frequency bands were exported from the offline setup to be used in the online evaluation. The accuracies of SVMs were determined by two-fold cross validation.

To provide biological feedback during the online setup, the miniature e-puck robot described above (5.3) was used. The connection between the robot and MATLAB was established by Bluetooth. After signal processing, features were extracted and classified into one of the four classes. A prediction (either 0 or 1) was first obtained from each binary classifier. Instead of relying on a single "best guess," the output was generated by weighted voting over a whole space of guesses. A threshold was then applied to each of the comparisons across all the channels. To improve the classification accuracy, classifiers that were below 70% accuracy were eliminated; furthermore, classifiers discriminating between relaxed state and any other class were eliminated if less than 90% accurate. Matrices containing the predictions of the classifiers and the accuracies of the classifiers were multiplied element by element. Columns were summed to obtain a prediction that reflected the choice and the accuracy of all the channels for a particular classification.

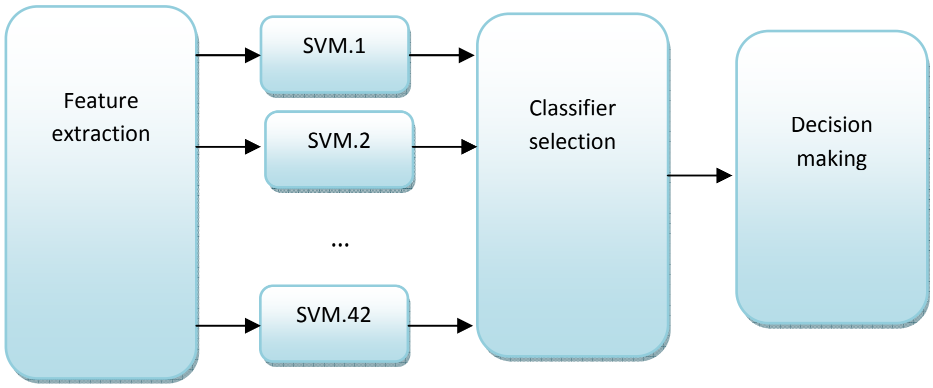


Figure 5.9: The structure of the ensemble classifier applied in the robot control project. One SVM classifier was generated for each possible comparison of features and for each channel, yielding a total of forty-two classifiers (six comparisons and seven channels), called SVM1 to SVM42. To improve the classification accuracy, classifiers that were below 70% accurate were eliminated.

Based on the numerical value contained in each position of the sum vector and the number of classifiers that contributed to that particular value, a final value was obtained that reflected the "strength" of each class. The strengths were ranked in order of magnitude and the class with the highest strength was determined to be the class imagined during that trial.

5.5.4 Alternative classifiers

In an offline study a comparisons between classifiers were done by applying LDA and QDA and RFD classifiers to the same dataset. We adjusted our model to extract the features and to classify the mental patterns using these classifiers.

5.5.4.1. Linear Discriminant Analysis (LDA)

A SVM performs classification by constructing an *N*-dimensional hyper-plane that optimally separates the data into two categories (Steinwart et al., 2008). The LDA approach is based on finding the best linear hyper-plane for separating the data of the different classes (Hastie et al., 2008). Figure 5.10 demonstrates a typical two-dimension boundary solution for the depicted dataset. In a two-class problem the line $\omega_0 + \omega^T x = 0$ forms the boundary. Assume S_b is the between-class variance matrix and S_ω is the within-class variance matrix for classes C_1 and C_2. These parameters can be calculated as follows:

$$S_b = (m_{c_1} - m_{c_2})(m_{c_1} - m_{c_2})^T \tag{5.10}$$

$$S_\omega = \sum_{x \in c_1} (x - m_{c_1})(x - m_{c_1})^T + \sum_{x \in c_2} (x - m_{c_2})(x - m_{c_2})^T \tag{5.11}$$

The optimal solution is based on finding the optimum weight vector ω so that the mean of two classes has maximum distance from each other and the variance between each class is minimized. It is then calculated as the eigenvector corresponding to the maximum eigenvalue of $S_{\omega}^{-1}S_b$ (Fukunaga, 1990).

$$J(w) = \frac{\omega^T S_b \omega}{\omega S_\omega \omega} \tag{5.12}$$

After calculating the training feature vector, each test feature vector can be classified employing a simple decision rule such as a sign function (Alpaydin E., 2004).

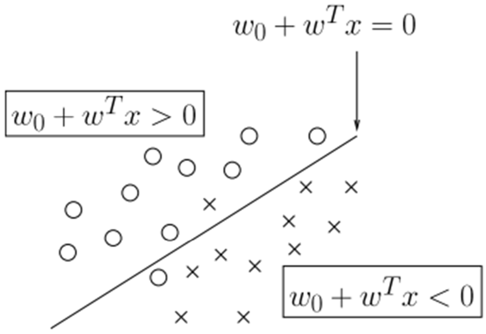


Figure 5.10: Fisher's LDA is a simple linear classifier based on finding the hyper-plane to separate the data representing the different classes.

5.5.4.2. Quadratic Discriminant Analysis (QDA)

LDA is based on the assumption that the classes have a common covariance matrix. It is likely in practice that this assumption does not meet the problem (Zhang, 2010). Decision boundaries in LDA can be easily extended to quadratic boundaries in QDA. By including the square values and cross products of the variable set $X_1^2, X_2^2, \ldots, X_1X_2, \ldots$ linear functions in the augmented space can be mapped to quadratic ones in the original space. In the case of QDA it is assumed that the covariance matrices in each class can be different. Without applying the Gaussianity assumption the decision surface can be estimated in a quadratic shape (Hastie et al., 2008). Figure 5.11 shows decision boundaries for LDA and QDA for the same problem. Quadratic decision boundaries were obtained by finding linear boundaries in the five-dimensional space $X_1, X_2, X_1X_2, X_1^2, X_2^2$.

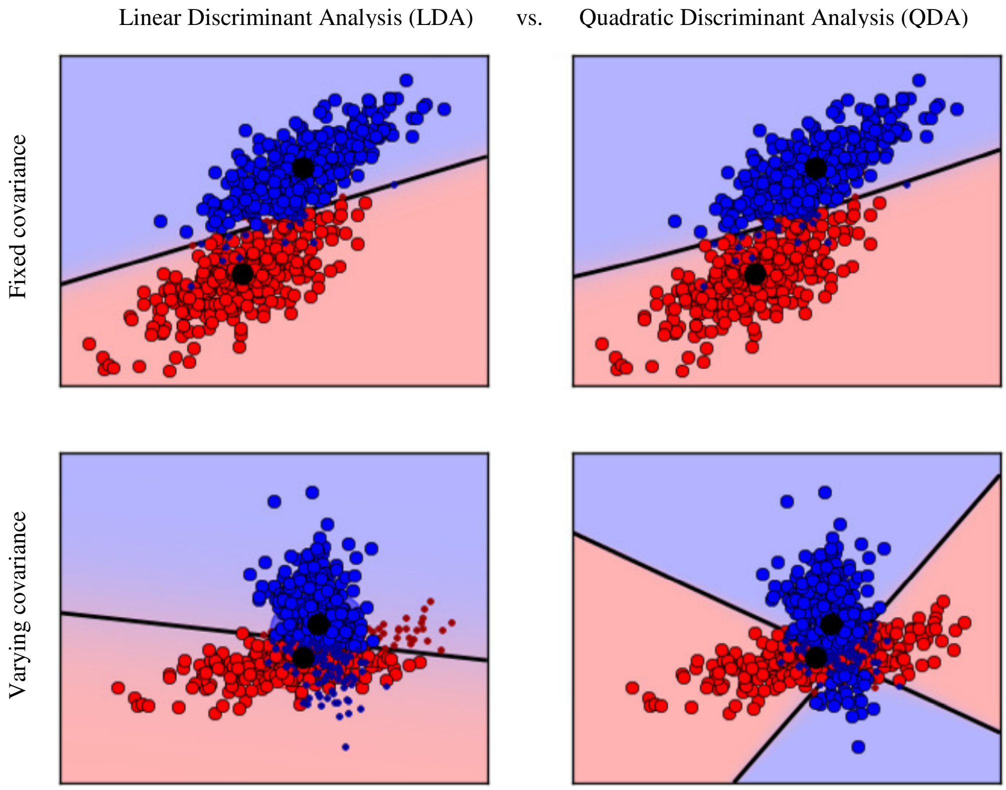


Figure 5.11: Decision boundaries for Linear and Quadratic discrimination. Data from two classes are shown in red and blue circles. Left panel shows linear decision boundaries found by linear discriminant analysis. Left panel shows the boundaries found by LDA method both for data with fixed covariance (Top) and with varying covariance (Bottom). The bottom row shows that the LDA algorithm learns only linear boundaries, but the QDA method will learn quadratic boundaries if the data covariance is not the same and therefore it is more flexible (produced by *Scikit-learn* (Pedregosa et al., 2011)).

5.5.4.3. Regularized Fisher's Discriminant (RFD)

Regularized classifier is a general concept and it is used when the complexity of a classifier is tuned in order to prevent overfitting (Krepkiy, 2008). Regularized LDA proposed by Friedman is a variation of the LDA classifier (Friedman, 1989) with interesting properties. This method allows separate covariances in QDA to be merged to a single covariance matrix. $\hat{\Sigma}$ is the pooled covariance matrix similar to the one used in LDA and $\alpha \in [0,1]$ regulates the transient between two models and is determined practically based on the performance of the model on the test data.

$$\hat{\Sigma}_k(\alpha) = \alpha\hat{\Sigma}_k + (1-\alpha)\hat{\Sigma} \tag{5.13}$$

In order to calculate the classifier parameters ω in RFD a quadratic optimization should be solved mathematically. Mika et al. proposed the following formulation (Mika, et al., 2001):

$$\min \ \frac{1}{2}W^TW + \frac{C}{m}\sum_{i=1}^{m}\xi_i \qquad (5.14)$$

Subject to:

$$y_i(X_i^TW + b) = 1 - \xi_i \text{, } and \ \ \xi_i \geq 0; \quad i = 1, \dots, m$$

Here m is the number of data points that are allowed to be classified incorrectly. The RFD approach searches for an optimized hyperplane as a solution for ω which minimizes the variance within each class and meanwhile makes the maximum distance between means of distributions of the data from each class. This technique has low computational requirements and, thus, is suitable for online BCIs. It has been used successfully in some BCI applications such as motor imagery based BCI (Kalcher et al., 1993) (Blankertz et al., 2006), P300 applications (Kaplan et al., 2013), and asynchronous BCI (Scherer et al., 2004). We applied these three classification methods as alternatives to the SVM classifier that was used in our online experiments.

The exploration of novel methods for improving the classification accuracy in BCI has attracted much attention in recent years. In this application we applied the "average" of multiple classifiers which showed better results than any individual one in the previous applications. Another way to improve classification accuracy is based on the notion of adaptive classification. Recruiting an adaptive intelligent system can highly increase the bit rate and reliability of the whole system. Classifier adaptation can be considered as an effective effort to solve the underlying issue of disparate distributions between the training and testing data (Alpaydin E., 2004).

In the last session of the experiment for each subject, after the offline training, the adaptation of the best classifiers continued during the online test. In these sessions we steered the robot by developing an adaptively trained BCI. The information extracted from the middle window in each trial was used to change the margin parameters. We were inspired by a 2007 paper by Yang et al., who used a set of adaptive support vector machines (ASVM) for the video detection (Yang et al., 2007). If the new sample was classified correctly by the available classifier, it was added to the train data set to change the SVM parameters slightly, i.e., $C_{t+1} = C_t + \varepsilon_c$ and $\gamma_{t+1} = \gamma_t + \varepsilon_\gamma$ where ε_c and ε_γ are less than 1% of the calculated values from the offline session. The applied change was in the direction of maximizing the posterior class probability. Spüler et al. reported a small MATLAB library for adaptive SVM calculation and applied it to a MEG-based BCI. The idea is relatively new for single SVM in BCI applications (Spüler et al., 2012).

5.6. Results

5.6.1 Classification accuracy

Online tests were used to assess the accuracy and the precision of classifiers generated during offline training runs. In order to evaluate the real-time robot navigation, we calculated the classification accuracy in each run. After five offline runs in each session, two types of online tests were performed to determine how precise the classifiers were and how well the user could control the robot. During the first online test, the accuracy was measured based on how many predefined actions the robot was able to perform in sequence. In other words the subject received an instruction consisting of 10 commands before the start of each run of the experiment. These 10 commands included all four possible conditions and led the robot from point A to point B in the 8-shape grid. EEG biofeedback was provided during online experiments. The robot was placed on a table in front of the subject on the 8-shape grid surface which was plotted on cardboard. The goal was to follow the instructed commands in order. The final evaluation was calculated by the number of correctly performed commands. Each of the offline and online sessions consisted of 5 runs and each run consisted of 10 trials. In the second online test subjects were free to select the commands on their own and the general performance was evaluated. Here we only report the results of the first online experiment.

Table 5.2 shows the averaged classification accuracies for each subject in different sessions of the experiment. Experimental evaluation on the three naive subjects demonstrated that an average classification accuracy of 76.6% was obtained during the first experiment session (day) after about 10 min training using the offline BCI setup. For the first online test the average classification accuracy was 71.3%. The evaluated offline results are better than the online ones, which is to be expected because criteria were developed in the "offline" data.

Table 5.2: The average classification accuracy over all experimental runs calculated for each participant and in each session of the experiment using ensemble SVM.

	Experimental sessions	**Classification accuracy %**	
		Training	Test
S1	Session 1	69.66	68.06
	Session 2	76.51	72.70
	Session 3	79.04	75.85
S2	Session 1	79.28	73.80
	Session 2	80.12	75.25
	Session 3	84.05	79.24
S3	Session 1	80.76	72.04
	Session 2	78.46	73.90
	Session 3	79.92	76.17

The left panel in Table 5.2 contains accuracies during the offline experiments, when the commands were sent to robot and biofeedback was provided. The values were averaged over the results achieved from the four classes of interest. If the class label was recognized correctly by the classifier +1 was noted otherwise for the misclassification case 0 was noted. The final percentage was calculated as the classification accuracy for that particular class. The right column shows the average accuracy during the online experiment. The results obtained from 450 trials (3 participants × 3 sessions × 5 runs × 10 trials) on the recorded data show that the proposed algorithm could perform robustly over different sessions and runs of the experiment.

Figure 5.12 shows the average classification accuracy accuracies for each subject in different runs of the experiment separately (averaged across sessions). In general a positive trend can be seen in classification accuracies from the first run to the last run of the offline session. The reason can be the effect of the training on the subject. The values drop by the start of the online experiments, but the positive trend with more variation can be also observed during this set of experiments for all three participants.

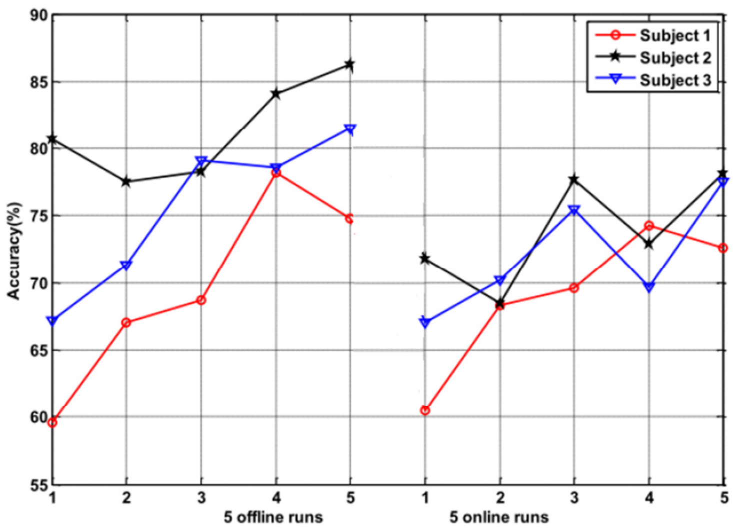


Figure 5.12: Average classification accuracy for the three subjects S1, S2 and S3 over three sessions of the offline and online experiment. Each offline and online session consisted of 5 runs (shown separately) and each run consisted of 10 trials (averaged). The data recorded from offline runs were used to initiate the classifier parameters for the online test.

An online BCI system should be designed in such a way that can cope with subject-to-subject or day-to-day variations. As described, each subject participated in three experiment sessions. We improved the classifier structure to the adaptive scheme in the last (third) session of the experiment. The online performance was slightly affected as can be seen in Table 5.2.

5.6.2 Information transfer rate

The average ITRs of the different experimental sessions for all three participants are shown in Table 5.3. For all subjects there was an increasing trend in ITR across sessions of the experiment. It is also clear that there is a decrease in ITR when using the online BCI paradigm. It can be also observed that applying the adaptive SVM method in the last session improved ITR in particular for the online test.

Table 5.3: The average ITR of the different experimental sessions calculated for each participant and in each session of the experiment.

	Experimental sessions	**Information transfer rate (BITS/min)**	
		Training	Test
S1	Session 1	16	12
	Session 2	20	16
	Session 3	21	18
S2	Session 1	18	12
	Session 2	22	15
	Session 3	23	19
S3	Session 1	15	11
	Session 2	19	14
	Session 3	22	17

In Eq. 5.3 regarding our BCI design, our proposed α_s was set to 10 for all subjects, *V=0.2* trials/second, and *N=4*. The ITR graph in our BCI application can be seen in Figure 5.13. Theoretically, by keeping the same specifications, the maximum ITR in 95% classification accuracy can reach to 32 with maximum SNR. The maximum ITR for this BCI application, as achieved by S2, amounted to 19 during the online experiments. Figure 5.13 illustrates the theoretical ITR curves for three- and four-class BCIs and the empirical values calculated in our experiments.

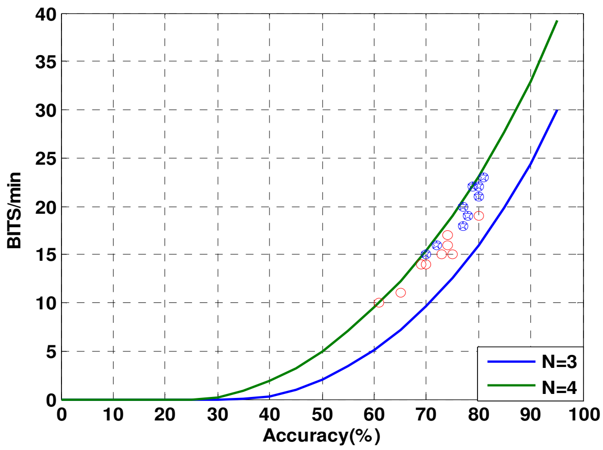


Figure 5.13: Theoretical information transfer rate for three and four class BCIs depicted by the blue and green curves, respectively. Circles show the empirical values from Table 5.3 for ITR and from Table 5.2 for accuracy. Red circles are the averaged values for the test runs and blue circles are the values for the training runs.

5.6.3 Other classifiers: An offline study

We used the recorded data from this experiment and applied other classification methods (LDA, QDA and RFD) in order to investigate the effect of classifier approach for this BCI application. For RFD model α was set to 0.5. Figure 5.14 shows the results achieved after ten-fold cross validation applied to the offline data. Here the same voting structure was applied for each classification method. As an alternative to the SVM classifier, LDA improved the accuracy in the first and third session noticeably, but in the second session had the lower average accuracy. Thus, it seems like a linear threshold may be sufficient in most cases to discriminate between each pair of classes. RFD performance was similar to SVM. This classifier is also the most robust one with the lowest variance in all three sessions. QDA performed well over the third session, but in average it was weaker than LDA. The improvement from SVM to LDA method was not statistically significant over the few studied subjects (df=2, t=2.10, p=.17). Considering the lower load of computation for LDA, we applied this classifier to our next paradigm for navigating an avatar in the virtual world.

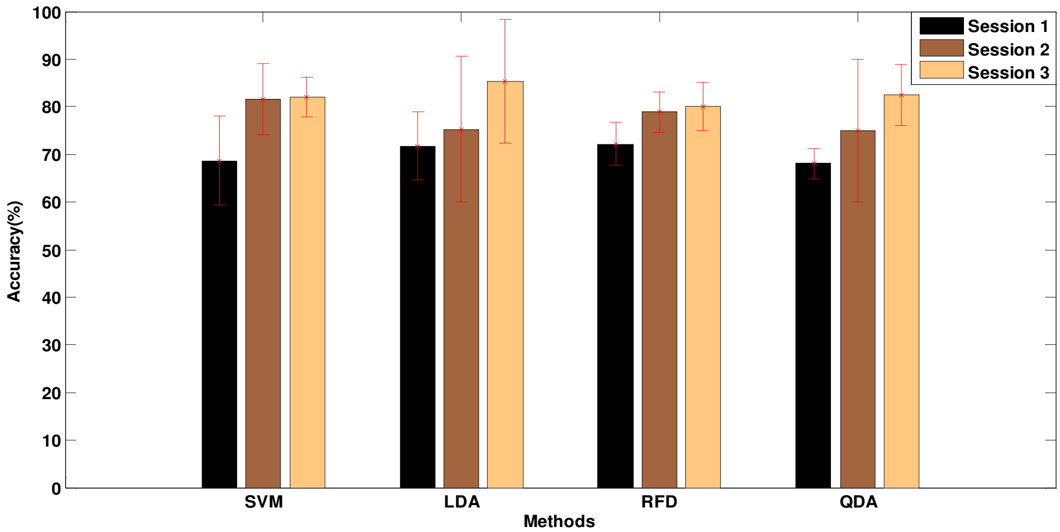


Figure 5.14: Average classification accuracies of the different classification methods over the recorded data of the three participants.

5.7. Discussion

This chapter presented a heuristic online single-trial EEG-based BCI for steering a robot vertically or horizontally in a simple two-dimensional environment. We investigated whether subjects could achieve satisfactory performance online with short-time offline training. The feedback was provided to the subjects all during the controlling endeavor, from the onset of the experiments to the end. Band power features were extracted from the EEG followed by mRMR feature selection to train and optimize an ensemble classifier. A distinguishing characteristic of our work is the creation of a unique classifier for each of seven channels and each of six possible feature space comparisons, yielding a total of forty-two classifiers.

The results after processing were verified as satisfactory. Predictions made by support vector machines whose accuracy was greater than 70% were weighted, and a command was given to the robot after voting. This strategy had a positive effect on the final results, thus, the online computation was limited to the classifiers whose performances were higher during the offline test. Non-stationary characteristic of the brain signal makes it difficult to use the same classifier of the previous session in the next session again, so we decided on a short training phase in every session. The preset 10 min training phase at the beginning of each session makes the classifier noticeably more robust to the non-stationary changes of brain activity. The SVM classifier is sensitive to the selection of parameters and the under-fitting may happen when not sufficient data is available. We hypothesized that a simpler classifier like LDA might results in comparable results with lower computational effort. Using the recorded dataset we investigated whether alternative classification techniques may lead to higher or more robust results. LDA, QDA and RFD were tested. The statistical test suggests that an ensemble LDA can be a suitable option for designing BCI because of its accuracy and simplicity for online calculations. The major problem with LDA compared to SVM classifier is that it is more sensitive to outliers. QDA is more robust and less sensitive than LDA, but may perform weak on particular data because of the misclassification rate, e.g., the first session data of the current application. Both LDA and QDA showed a wide dispersion in the second session, which can be due to outliers. Probably due to the same reason, LDA has also more variation compared to other methods in the last session.

In order to assess the performance of the entire BCI system, several factors should be considered. Among them the number of commands (mental tasks), reliability and accuracy and information transfer rate are the most important parameters. Considering both the hardware and software limitations, in this application, we decided on a four class BCI. Adding the number of mental tasks would have increased the entire training time as well as the online computational time which is not desirable for an online BCI. A general assessment shows that the presented results for this application indeed are better than the classification in the previous applications. Average accuracy in the current application was calculated 75.52 for the offline analysis and 71.61 for the online test. In the first application using fractal features the average accuracy of 81.3 was achieved, in the second application using a combination of power band and higher order features the accuracy of 71.24 in average was obtained. Considering that the chance probability is only 25% for the four-class BCI compared to 50%

for the two-class BCI, the proposed BCI in this chapter with 73.56 % classification accuracy in average outperforms not only the second application (71.24%) but also the first one (81.3%). One interpretation for this rather high performance is the design of the experimental paradigm, including both offline and online steps. The second and more important reason could the structure of the ensemble classifier that was applied on this project. In essence instead of using a single four-class classifier, we implemented six separate binary classifiers for each single channel. Those classifiers which had the high performance on the recorded data were selected for the final implementation. The ultimate decision was made based on the voting strategy.

Chapter 6

Avatar Navigation in "Second Life" using Brain Signals[1]

6.1. Introduction

In this chapter an online single-trial EEG-based BCI linked into an interactive virtual world environment (VWE) is presented. The purpose of this interdisciplinary project was to conduct an innovative approach to establish a connection between a human brain and a virtual world illustration using a simple game paradigm. Development of the interaction technique allows the users to freely decide which action they wish the avatar to perform. Similar to the BCI described in the previous chapter, the system was trained for some predefined movements. The control can be accomplished by online processing of single-trial EEG signal which is recorded using an EPOC headset from Emotiv. Our online BCI was developed as a navigating game in a 3D virtual environment for controlling an avatar. The initial idea was to employ virtual reality for physical rehabilitation. So not only healthy people, but also people with handicaps could benefit from this new technology.

6.2. Virtual Reality for fun and rehabilitation

Virtual social worlds (e.g., "Second Life") are becoming more accessible for people from different generations and social groups. Morton Heilig has originally implemented the initial idea of a “Virtual Reality” (VR) in 1956 in a fully interactive device called Sensorama (Brooks, 1999). VR is a general concept to describe any scenario created virtually by the computer software which is equipped with the user-interaction feature (Brooks, 1999). The

[1] Part of this work was published in: **M. Kh. Hazrati**, and U. G. Hofmann, Avatar Navigation in Second Life using Brain Signals, *8th IEEE international symposium on intelligent signal processing (WISP)*, 2013.

virtual world is usually designed in the three-dimension scheme in such a way that users feel themselves as a part of the scene (Keshner, 2004).

Using the VR design gives the researcher or the therapist the potential to simulate the complex physical problems in a high degree controlled environment without any reductionism (Brooks, 1999). Second Life's virtual world is a well-known web-based interaction environment (Au, 2008). Employing scientific game-based therapies and technologies also constructs another aspect of new BCI generation. Impairment in motor function leads to difficulties in individual and social interactions and will diminish the quality of life in a wide range of patients. According to the literature, VR technology has been applied to many different applications such as entertainment and physical rehabilitation (Gil-Gomez et al., 2011) (Brutsch et al., 2011). Paralyzed people not only suffer from physical inability but also from psychological aspects of inability, which is most of the time the worse issue for them. Neuroplasticity improves with motivation (Jones et al., 2008). Elderly people with stroke and children with sensorimotor disorders may also gain from interactive computer play for rehabilitation purposes (Sandlund et al., 2009) (Rabin et al., 2011). Providing a fun and attractive interface can provoke the user's intention in a more effective way (Schuler et al., 2011). The effectiveness of the Wii system has also been tested for balance rehabilitation in patients with acquired brain injury (Gil-Gomez et al., 2011). A recent research extends the usage of the rehabilitation robot called "Lokomat®" from boring and monotonous physical practice sessions to an exciting game controller for children (Sandlund et al., 2009) (Schuler et al., 2011).

6.3. Project Description

For the current BCI application we combined the VR interface technique with the advancement of wireless electronics technology. Patients as well as healthy people could benefit from increased mobility when the restrictions imposed by cables and wires are removed (Debener S. et al., 2012). This renders the proposed application a possible portable solution for healthcare purposes.

The advent of tiny low cost and low power electronic chips propelled the electronic design forward. It has opened up new doors to the future of BCI. Just in the recent five years several electronic headsets have been presented to the market and the revolution of inexpensive wireless EEG device has enormously affected the field of BCI. Therefore, EEG data used in this application were recorded by commercial EEG amplifiers.

6.3.1. EEG recording Hardware and Software

The EEG acquisition machine was Emotiv EPOC headset from Emotiv (Australia). The EPOC headset is a low-cost revolutionary EEG cap, with dry electrodes. It is easy to wear and use and is equipped with continuous Bluetooth connectivity to the computer. Since 2007, the Emotiv EPOC has been available in the market. Recently it found great attention from researchers for BCI based games (Debener S. et al., 2012) (Kaplan et al., 2013). The sampling rate is 128 Hz (2048 Hz internal) and the resolution is 14 bits, 1 LSB=0.51 μV (16 bit analog to digital amplifier, 2 bits instrumental noise floor discarded). Its dynamic range is up to 8400 μV (peak to peak), which is equivalent to 78 dB ($20\log(V_{out}/V_{in})$). EPOC also includes gyroscopes (X and Y axis). Figure 6.1 shows the headset and positions of electrodes covered by it. The headset is equipped with 14 dry sensors according to the International 10-20 system at AF3, F7, FC5, F3, AF4, F8, FC6, F4, T7, T8, P7, P8, O1, and O2. Two reference channels (CMS and DRL) in addition to 14 EEG channels are approximately positioned at P3 and P4, called CMS and DRL, respectively. The headset is fully wireless and the battery can hold a charge for around 12 hours as announced by the company. Emotiv licensing model is based on providing access to different level of data depending on license. Only a research license provides access to raw EEG data. http://www.emotiv.com/

Before the start of experiment a saline was used to wet the electrodes. The EPOC EEG cap was then fixed on the subject's head, and electrodes were placed in their positions. In the normal use, no sensors are directly placed over the motor cortex, so motor imagery tasks might be a challenge for EPOC.

Figure 6.1: EPOC headset and corresponding electrode locations. The headset is equipped with 14 dry sensors according to the International 10-20 system at AF3, F7, FC5, F3, AF4, F8, FC6, F4, T7, T8, P7, P8, O1, and O2.

To implement the avatar control by BCI, appropriate and optimized computer software was required. In our case, we used a C++ programming interface under Microsoft® Windows Vista for online data acquisition. The Emotiv connection protocol is responsible for recording the raw brain signal. A Bluetooth® driver was used to communicate with the EPOC EEG cap and the computer (Debener S. et al., 2012).

A self-developed C++ graphical user interface (GUI) connected to second life provided the interactive environment. User-written MATLAB scripts were finally converted to stand-alone applications using MATLAB Compiler version 4.11. Therefore, they can be executed independent of the MATLAB environment. MATLAB Compiler Runtime (MCR) version 7.11 should be installed on the computer while executing the interface. It is a standalone set of shared libraries that enable the execution of M-files.

6.3.2. Experimental Paradigm

This experiment was performed in the Computational Neuro-Engineering Laboratory and the Digital worlds institute at the University of Florida. Experimental runs were carried out with three able-bodied volunteer subjects (one female, two males; mean age: 27.3). In preparation for each run, subjects were seated 0.5 meters from a computer monitor with arms in front and palms flat and facing down on a desk in front of them. The experimental paradigm consisted of 4 offline training runs followed by 6 online evaluations / avatar control runs (Figure 6.2). Each offline and online run consisted of 12 trials which yield to the total 48 offline and 72 (6×12) online trials. During the controlling endeavor in online experiments subjects received feedback on the screen. Each online trial lasted ten seconds, the first five devoted to the imagination of movement and the last five to relaxation.

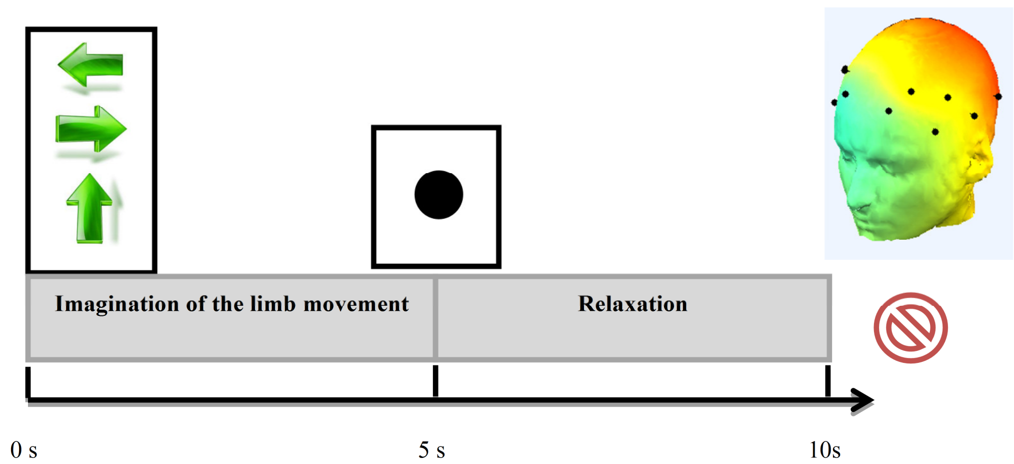


Figure 6.2: Offline experiment (training sessions): The subject is trained by observing right and left and up arrows at the center of the screen. An image of a black circle was continuously displayed for use as a visual fixation point.

During each trial, similar to the previous application (Chapter 5), an image of a black circle was continuously displayed as a visual fixation point for the user, to reduce eye movement artifacts. Since all cues were displayed in the middle of the screen for all choices, it is assumed that possible induced eye movements were independent from the indicated targets. In the offline session visual stimuli indicated which of the following three motor imageries the subject should perform. One of three images – a left arrow, a right arrow, and an upward arrow – was shown for five seconds, and the subject was asked to perform a corresponding brain activity, i.e., imagination of moving the left hand, right hand, or foot (either left or right one). The type of movement was not specified beforehand, however, the subjects reported that they preferably imagined sideward movements. After five seconds, the picture disappeared, and a screen with only the focal point was shown, indicating that the subject should relax. The black circle was visible on the screen for 5 s. The black circle was extinguished for periods of random length, between 2 s and 3 s, and then the next trial started with presentation of an arrow. Before the start of the online experiment, the parameters of the classifier were initialized using information derived from offline experiments. The goal was to investigate whether subjects could achieve satisfactory performance online with short-time offline training followed by adaptive online training.

Here also two types of online tests were performed. In the first setup, the subjects were given a visual cue (i.e., one of the three arrows or the black circle) to begin imagining one of the actions (or relaxation) for which they had been trained. The subjects were then given five seconds to imagine the required mental state. Processing of signals was performed online and two commands were sent to the avatar during each trial (after 2.5 s and 5 s of the onset of the trial); the execution of the avatar action took place accordingly. The subjects were then given random length between 2 to 3 s to relax. The online performance was recorded to determine how accurate the training was and how well the user could control the avatar. In the second setup similar to our previous experiment (Chapter 5), the subjects were asked to navigate from the start point to the end point as shown in Figure 6.3. They were shown a black focal point and were given five seconds to imagine the mental state corresponding to the action they wished the avatar to perform. The avatar status was updated after 2.5 s had passed. Here we will only consider the data recorded from the first online test for the evaluation.

Figure 6.3: A snapshot of the user interface for the online experiment. The subject was trained to send a command to the avatar by thinking about the direction commands: Left, right and forward. During the test section users should navigate their avatar from point A to point B.

6.3.3. Second Life interaction software

Users can explore the Second Life's (SL) virtual world willingly and steer their avatar to the target point normally with mouse or arrow keys. Recently BCIs have been used to control Second Life, but for a different navigation design (Kaplan et al., 2013). Here we developed a sophisticated new brain-computer interface based on the functionality of this environment. MathGL and OpenGL were used for handling the graphics and computation. Instead of arrow keys on a keyboard for controlling the avatar, commands were sent directly to the interface. It let users explore the virtual world and steer their avatar within the 3D environment.

The interface was developed in C++ including Emotiv SDK Developer Edition. It passes the arguments to the main interface using *keybd_event()* Win32 API function which synthesizes a keystroke. The keyboard driver's interrupt handler calls the *keybd_event* function. Using this function one can send a virtual key *(VK_XXX)*, which has been already defined in the header file *winuser.h*. The function returns a flag for denoting the key type, e.g. KeyDown, KeyUp. It can also be set to return a state for indicating an extended key. The first input of this function is another function called *VkKeyScan()*. The latter is used to translate normal characters with type *Char* into virtual keys with type *Word* denoting a *VK*. The *KEYEVENTF_KEYUP* flag should be updated to terminate the key pressed command. A part of the code can be seen in the next page:

```
void MyKeyEvent(std::string text)
{
  for(int i = 0; i < text.length(); ++i)
  {
//Press key
    keybd_event(VkKeyScan(text[i]),0x9e, 0 , 0);
…
//Stop pressing key
    keybd_event(VkKeyScan(text[i]),0x9e, 0 ,KEYEVENTF_KEYUP);
  }
}

#include <iostream>
#include <windows.h>
using namespace std;

int main()
{
    cout << "Simulated key Pressed " << endl;
    getchar();
    Sleep(1000);
    POINT p;
    GetCursorPos(&p);
    HWND Right_Arrow=WindowFromPoint(p);
    cout<<"HWND: "<<Right_Arrow<<endl;
    PostMessage (Right_Arrow,WM_KEYDOWN, 0, 'H');
    Sleep(1000);
}

//LeftArrow  E0 4B
//RightArrow E0 4D
//UpArrow    E0 48
//DownArrow  E0 50
```

In order to record sufficient data a preset 10 min offline phase was assigned. Here we selected six classifiers (six pair-wise comparisons of the 4 classes) using the information of all channels. After the training phase ten-fold cross-validation was used to assess each classifier. For the first online test with a cue-paced interface, the subjects were restricted to perform predefined (trained) mental tasks in intervals and tried to steer an avatar on the screen. In the second online test the paradigm was similar to the first online test, but this time the subject was free to select the type of mental task. In the latter study subjects steered their avatar to a final destination through an arbitrary path only by imagining four movements.

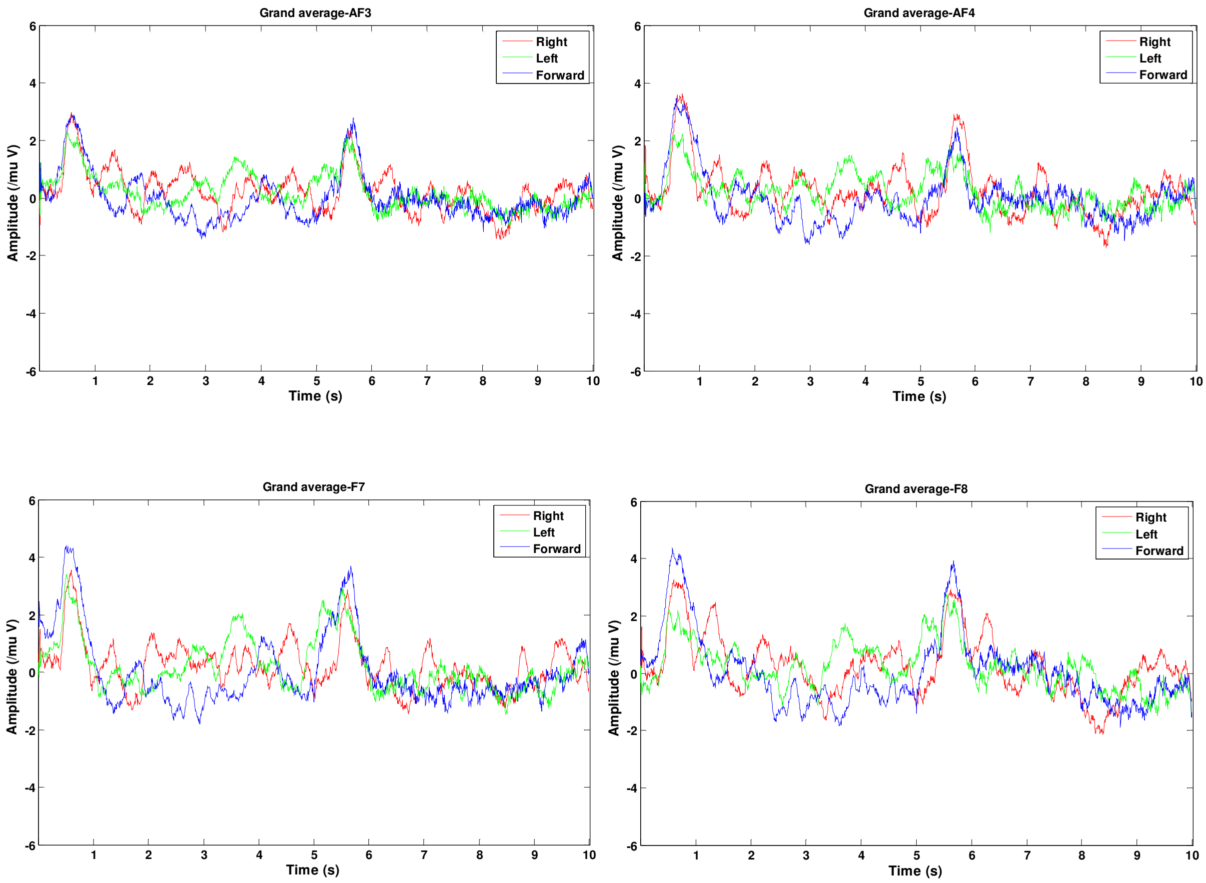
6.4. Evaluation of the Data

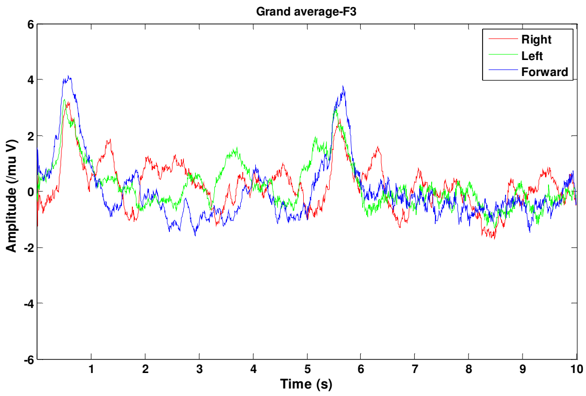
6.4.1. Evoked potentials

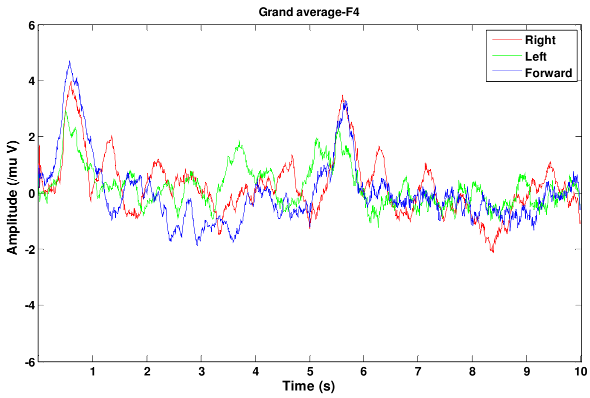
In order to evaluate the recorded data, some further time and frequency analyses have been accomplished. Figure 6.4 presents grand average ERPs for all subjects recorded from AF3, AF4, F4, F3, FC5, FC6, P7, P8, O1 and O2 for trials associated to the right, left and forward arrow and the following stop (idle state). Units on x-axis are *s* which represent the entire trial starting from the point when the arrow appeared on the screen at the start of each trial till 10 s. At time point 5 s the relaxation phase started when the arrow image disappeared from the screen.

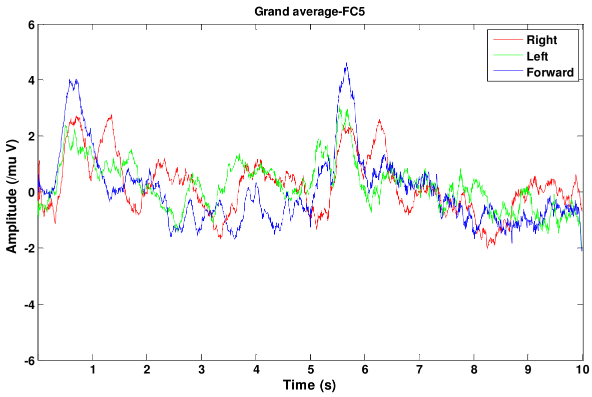
The first deflection is the positive peak of the evoked potential, consisting of a 5.8 μV at about 480 ms for forward command trials maximum at O1 and O2 and 4.9 μV and 4.2 μV at the same time for right and left commands. The maximum amplitude decreases from occipital to frontal sites and reached a minimum of 3 μV at AF3. Another positive peak with the amplitude of +4.1 μV can be seen at 5560 ms after the start of the trial. The disappearance of the arrow and the display of the black dot on the screen might have caused this evoked potential.

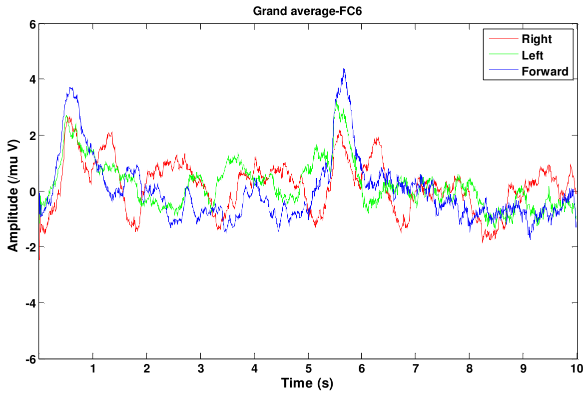
The positive peaks differ in shape and duration for different imagination trials. Especially for the upward arrow the topography is more similar to a P300 component with maximize at O2. In general there are more fluctuations in the ERPs calculated for this data set, which can be due to the low number of the trials and the specification of the recording system. But it is pleasing to see that there were no gross side differences between AF3 and AF4 (which are near to the eyes) for right- vs. left-pointing arrows, which confirms the expectation that no systematic eye movements were induced by arrow direction.

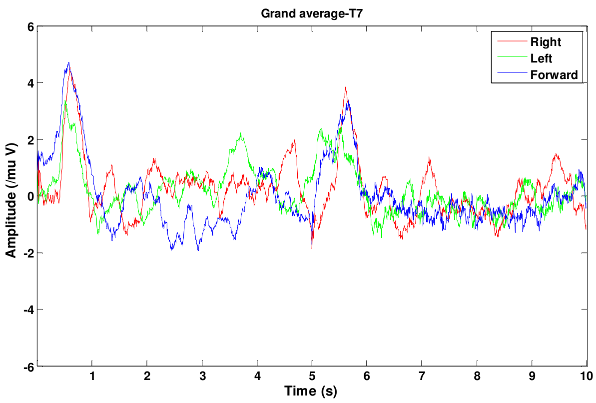


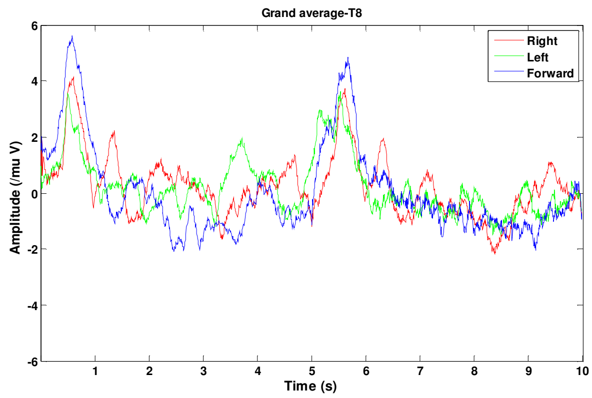

Grand average-F3
Right
Left
Forward
Amplitude (/mu V)
Time (s)

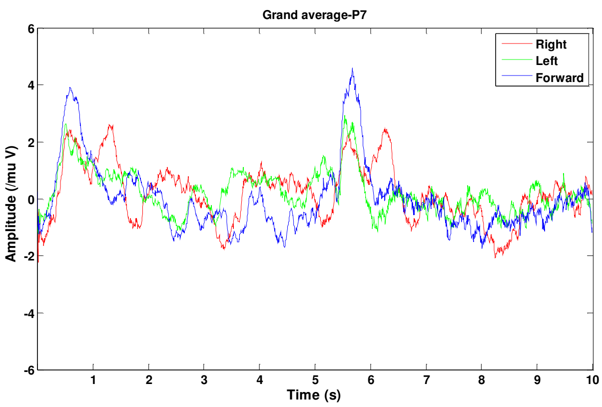

Grand average-F4
Right
Left
Forward
Amplitude (/mu V)
Time (s)

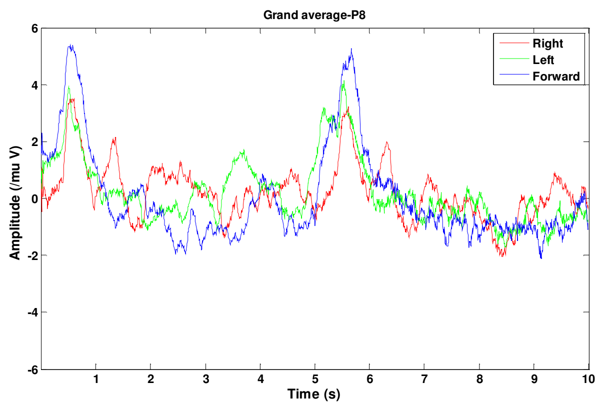

Grand average-FC5
Right
Left
Forward
Amplitude (/mu V)
Time (s)


Grand average-FC6
Right
Left
Forward
Amplitude (/mu V)
Time (s)


Grand average-T7
Right
Left
Forward
Amplitude (/mu V)
Time (s)


Grand average-T8
Right
Left
Forward
Amplitude (/mu V)
Time (s)


Grand average-P7
Right
Left
Forward
Amplitude (/mu V)
Time (s)


Grand average-P8
Right
Left
Forward
Amplitude (/mu V)
Time (s)

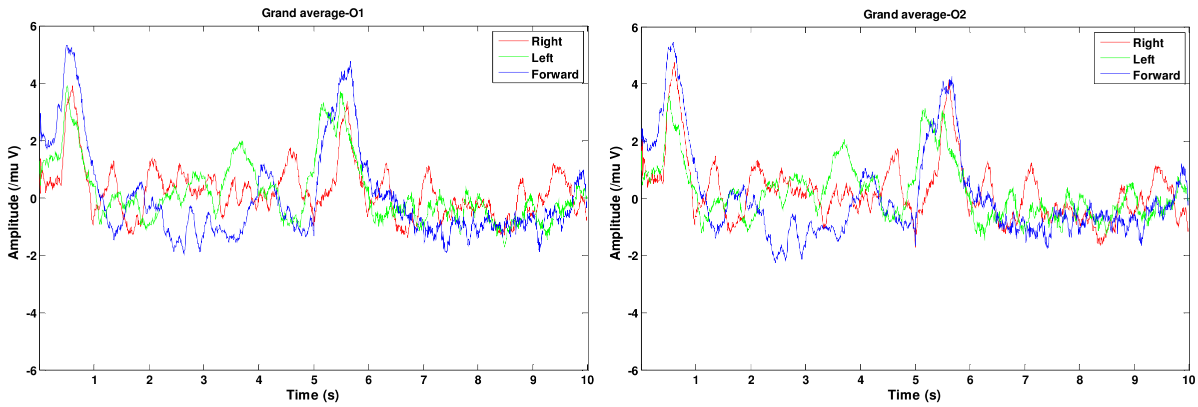
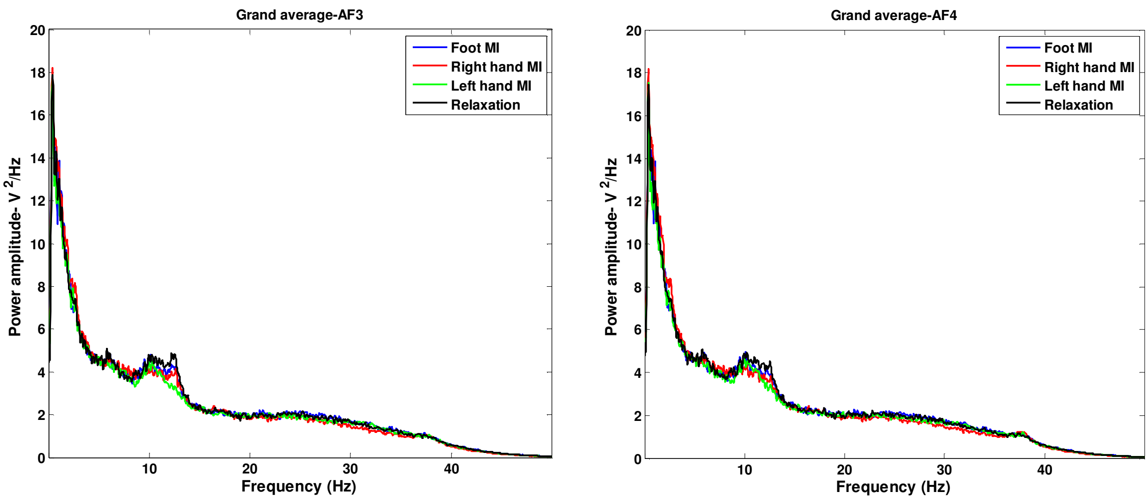
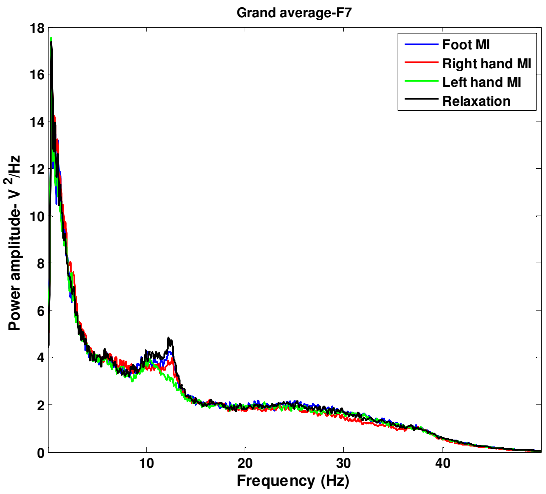


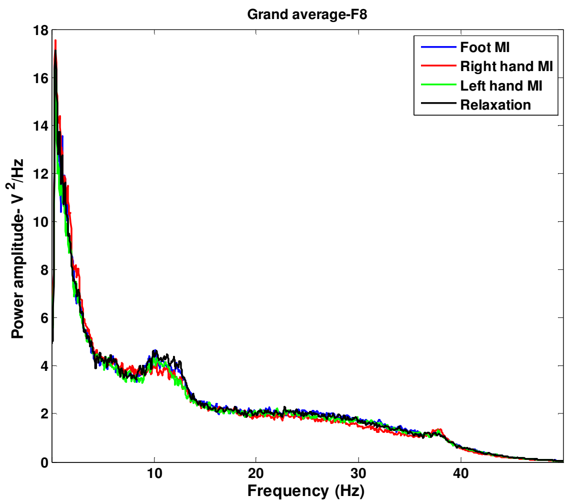
Figure 6.4: Grand average ERP calculated for the current data set. Data were averaged over all trials and subjects separately for AF3, AF4, F7, F3, F4, F8, FC5, FC6, T7, T8, P7, P8, O1, and O2. At onset of the trial an arrow appeared on the screen showing randomly one of the images corresponding to the imagination of movement. At 5 s a black dot appeared on the screen as a sign for the relaxation interval which lasted 5 s (5 s – 10 s). Amplitude units are μV, x-axes range from 0 s (left) to 10 s (right).

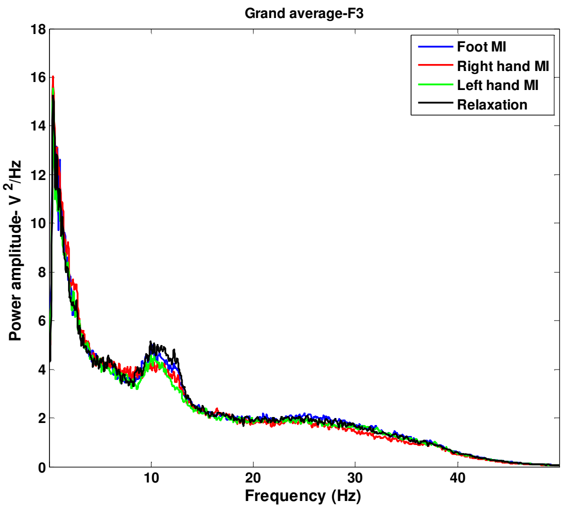
6.4.2. Power band spectrum

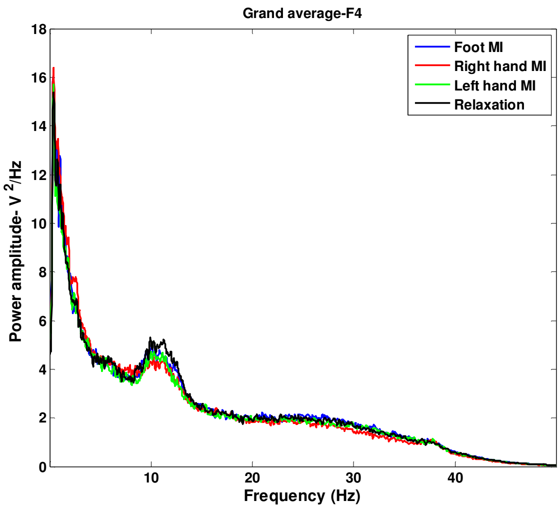
Figure 6.5 demonstrates the average power band spectrum for the current dataset. The amplitude is calculated in 10×log10 (V^2/Hz) unit. Data from relaxation states are presented in black and data from movement imagination states are presented in color (blue, red and green). Due to the lower number of trials available in this dataset (N=3 subjects × 48 offline trials), some local fluctuations can be seen in Figure 6.5.

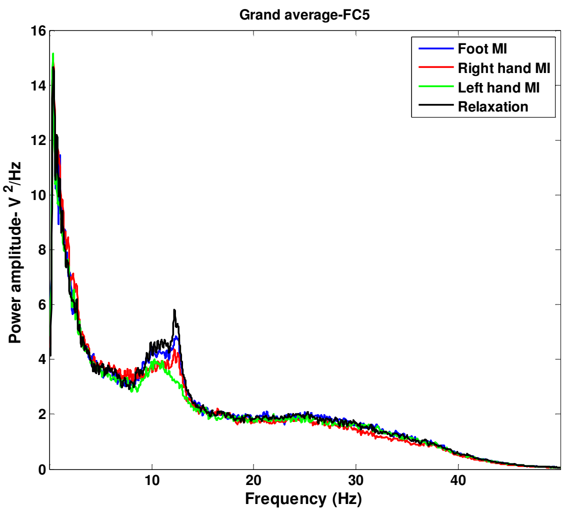


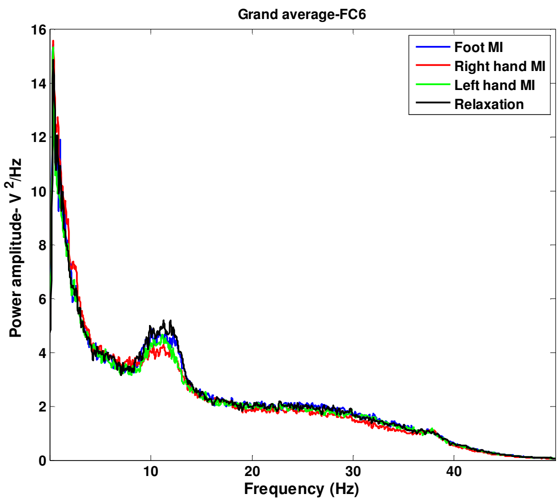

Grand average-F7
Foot MI
Right hand MI
Left hand MI
Relaxation
Power amplitude- V ^{2}/Hz
Frequency (Hz)


Grand average-F8
Foot MI
Right hand MI
Left hand MI
Relaxation
Power amplitude- V ^{2}/Hz
Frequency (Hz)


Grand average-F3
Foot MI
Right hand MI
Left hand MI
Relaxation
Power amplitude- V ^{2}/Hz
Frequency (Hz)


Grand average-F4
Foot MI
Right hand MI
Left hand MI
Relaxation
Power amplitude- V ^{2}/Hz
Frequency (Hz)


Grand average-FC5
Foot MI
Right hand MI
Left hand MI
Relaxation
Power amplitude- V ^{2}/Hz
Frequency (Hz)


Grand average-FC6
Foot MI
Right hand MI
Left hand MI
Relaxation
Power amplitude- V ^{2}/Hz
Frequency (Hz)

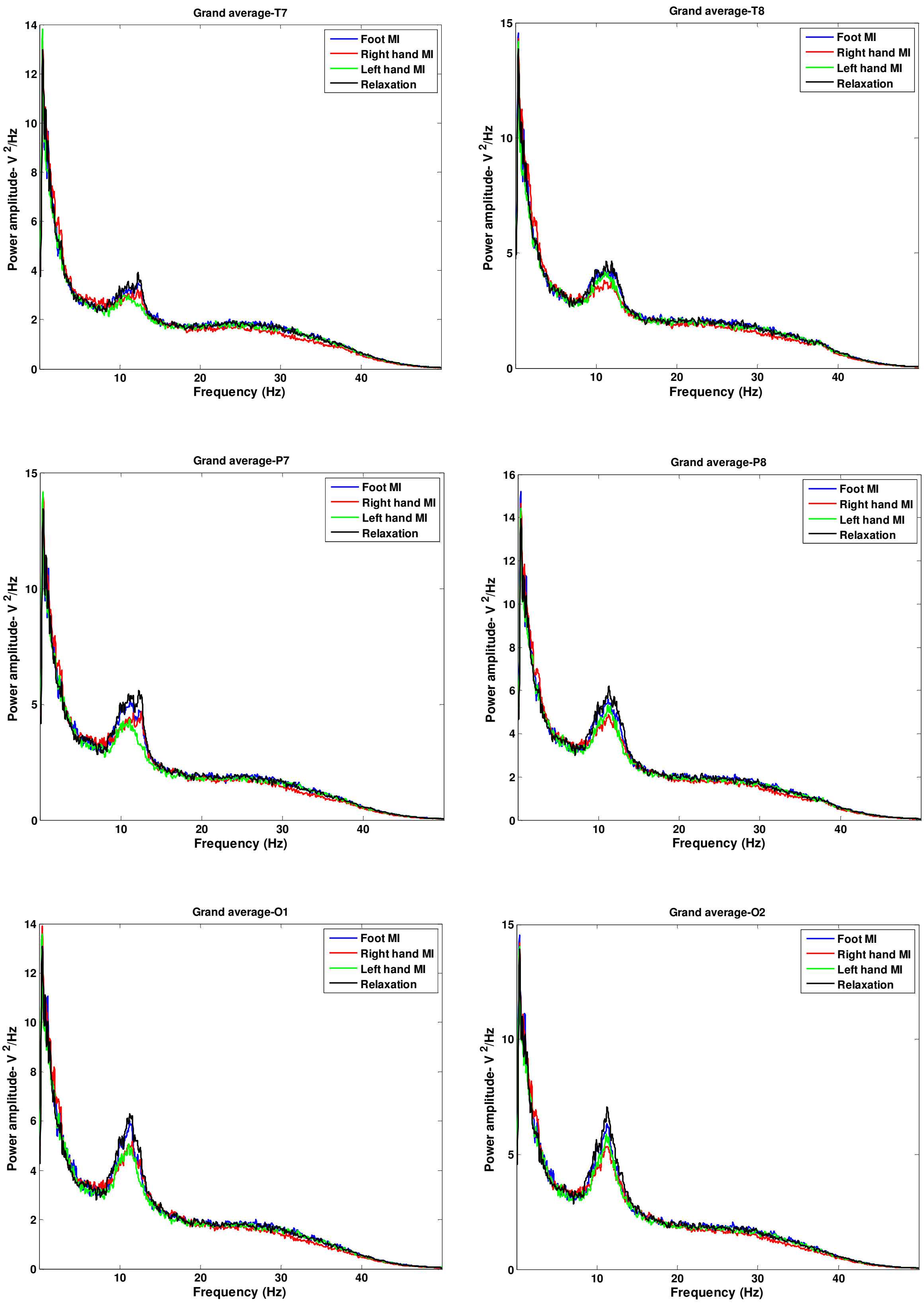


Figure 6.5: Average power band over the entire dataset for two different brain states demonstrated over AF3, AF4, F7, F3, F4, F8, FC5, FC6, T7, T8, P7, P8, O1, O2. - Unit on x-axis is Hz, unit on y-axis is log10 (V^2/Hz).

The alpha peak can be observed from 8 Hz to 12 Hz and is stronger during the relaxation and foot imagination trials. In the power spectrum calculated from right hand imagination trials another peak exists in gamma band, which is stronger at F8 and AF4 sites. In AF3, F7 and FC5 a late alpha peak at 13 Hz can be observed. The alpha peak in left hand imagination trials is smaller in comparison to other imagination states and the alpha peak can be seen earlier compared to the other states. The power spectrum computed for the left hand imagination (green) is larger than the one for the right hand imagination (red) over the electrodes positioned on the right hemisphere. For the left-side electrodes the green curve has smaller values than the red one. Thus the side specific ERD occurred.

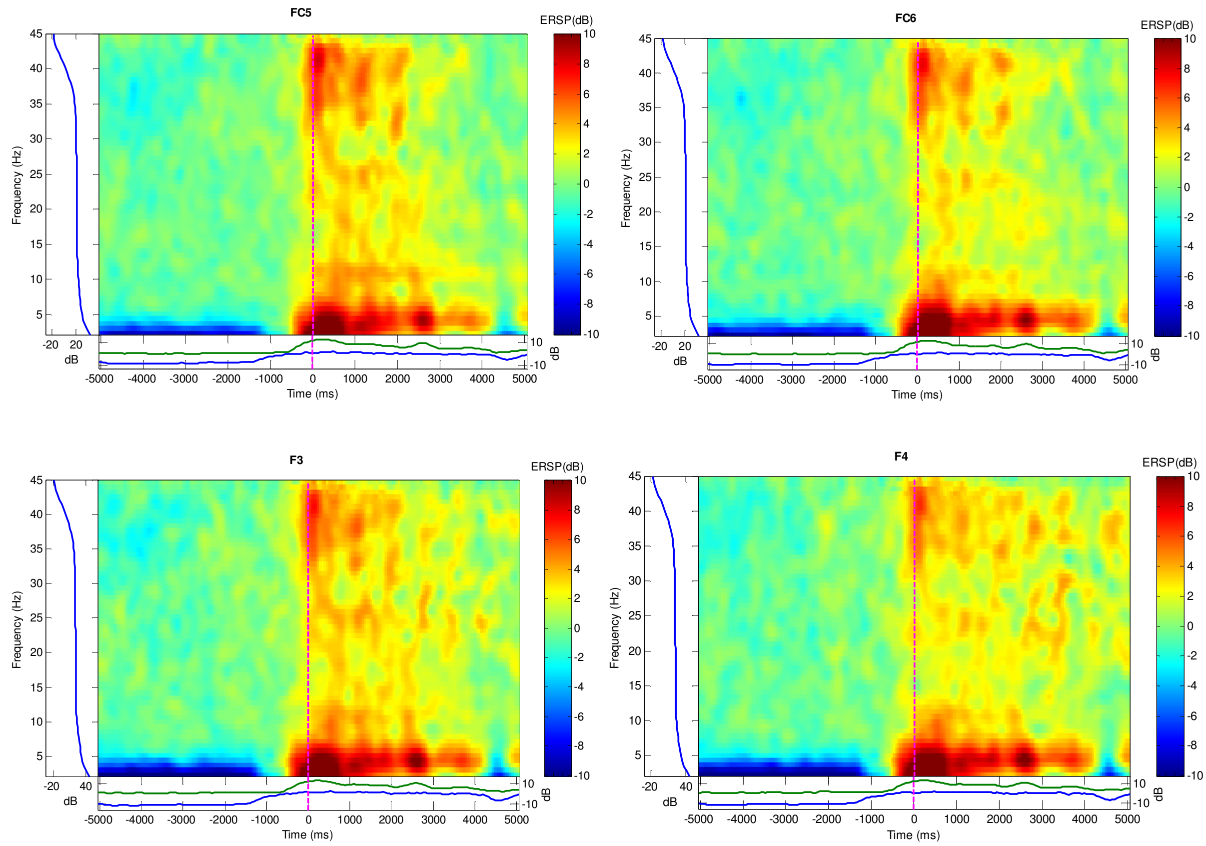
6.4.3. ERSP time-frequency analysis

Figure 6.6 shows some time-frequency analysis using ERSP approach on this dataset. Lacking electrodes over the motor cortex, here the averaged ERSP over subjects is depicted for two selected channels as close to the motor cortex as possible. Three groups of ERSP plots are depicted in Figure 6.6; each is calculated using the trials for the imagination of different limb movement. In the first group of subplots the ERSP analyses for imagination of right hand movement can be seen. A clear theta ERD can be seen over all four plotted figures.

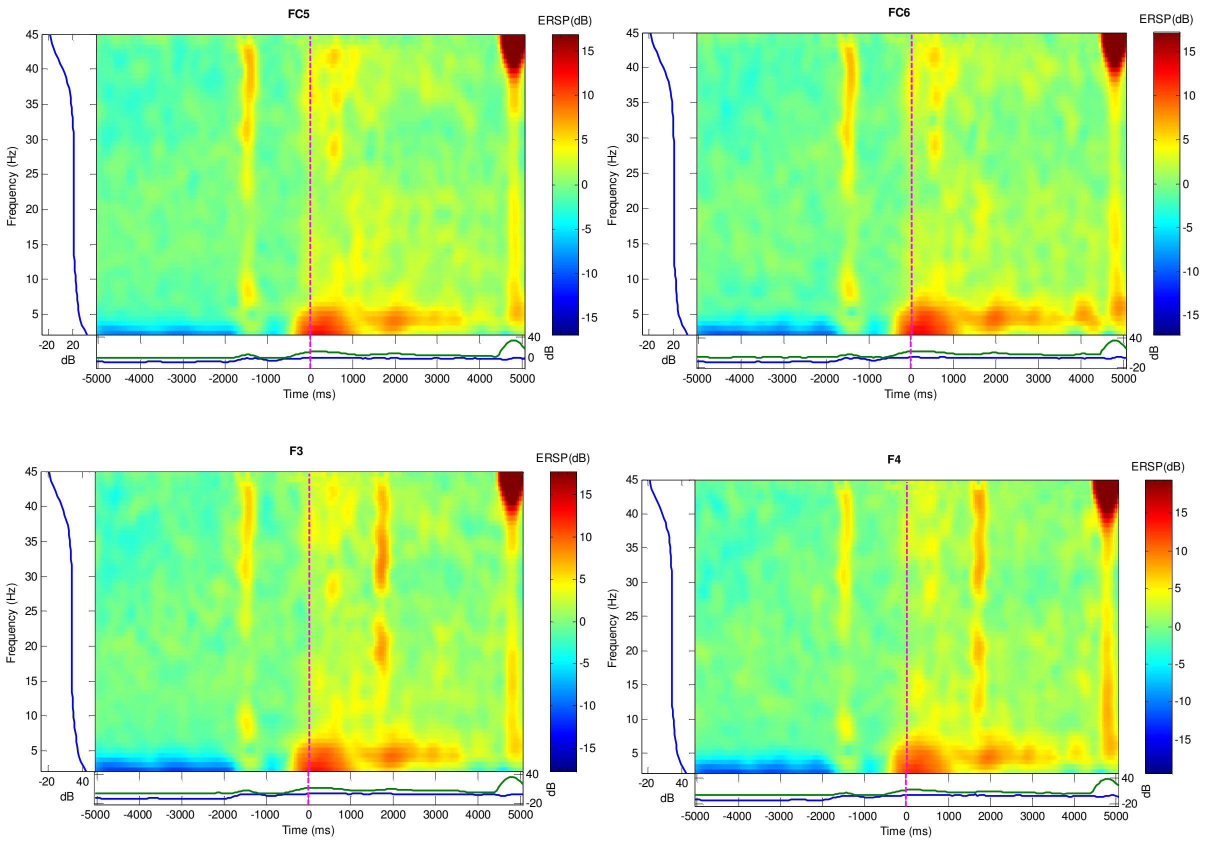
Like Figure 6.4, this Figure 6.6 shows the time range from time point 0 s, when the arrow was displayed on the screen until 10 s , the end of the relaxation phase. However, the first five seconds were defined as baseline for ERSP computation, so the time axis is shifted by five s, with -5 s in Figure 6.6 denoting the start of the imagination phase, and 0 s denoting start of the relaxation phase. In the first 5 s, a noticeable decrease in delta band over all conditions in FC5 and FC6 (overlying right and left hand-motor cortex) can be observed. At 0 s the cue disappeared and a black dot remained on the screen for another 5 s. An increase in power of theta and alpha bands exists over all four plotted channels. Increases in gamma band can also be seen at this period too. The second four subplots in Figure 6.6 show the similar analyses for the imagination of left hand movement. Here the ERS in FC6 is clearer than FC5, in accordance with the contralateral organization of movement control. Gamma ERS in this period might be due to artifacts.

The third group of figures is related to foot movement imagination. Some sparse gamma and theta ERS can be seen during the imagination of movement phase from the start of the trial till the onset of the relaxation phase at 5 s. Gamma ERS exist in trials for imagination of foot which is probably due to residue effect of artifacts. A sparse beta and gamma ERS starts at 2 s and lasts till 5 s. A strong theta ERS at zero point can be the effect of ERPs in all subplots. Interestingly some sparse gamma ERD can be observed in FC6.

Imagination of right hand movement



Imagination of left hand movement



Imagination of foot movement

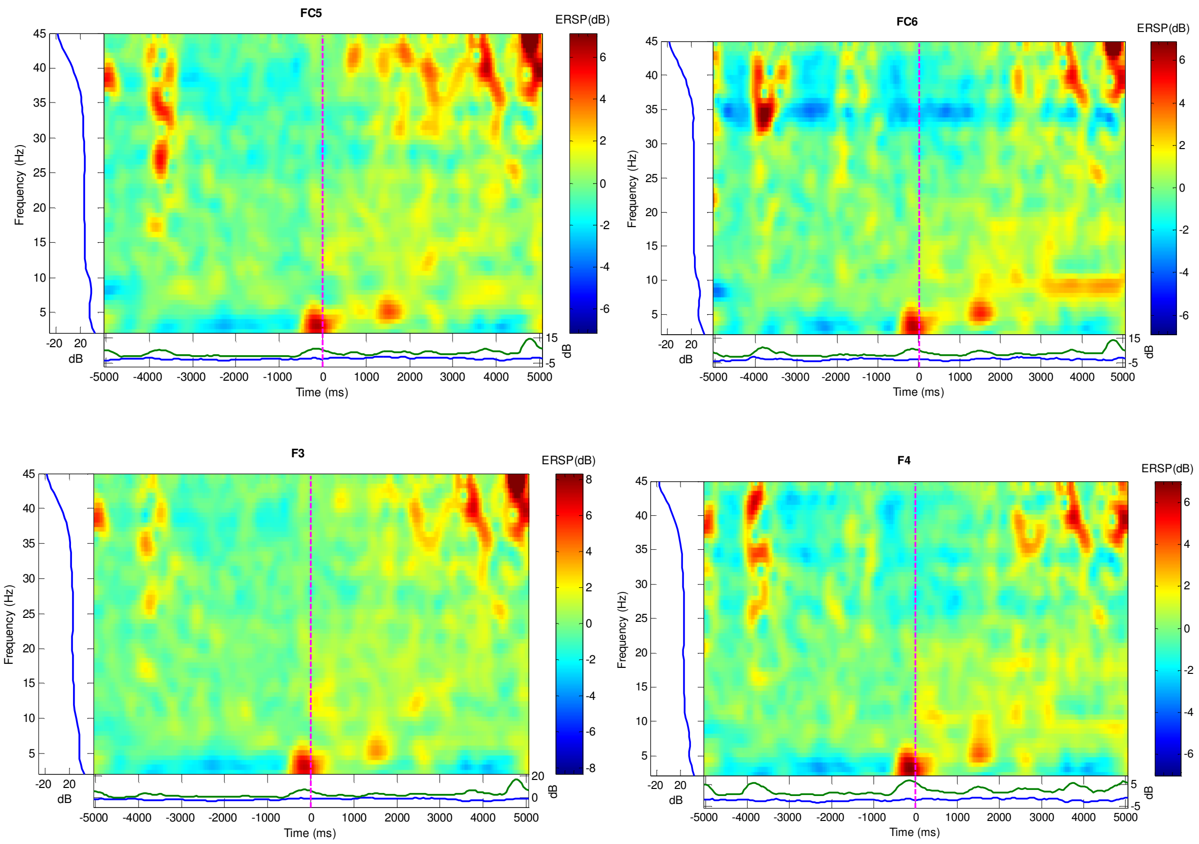


Figure 6.6: Grand average- - ERSP averaged over all trials for imagination of the limb movements calculated for FC5 (left) and FC6 (right)- At onset of the trial an arrow appeared on the screen showing randomly one of the images corresponding to the imagination of movement. Each four subplots correspond to the imagination of right hand, left hand and foot movements, respectively. At 0 s a black dot appeared on the screen as a sign for the relaxation interval which lasted 5 s (0 s – 5 s). Amplitude units are µV, x-axes range from 0 ms (left) to 10000 ms (right). Unit on x-axis is *ms*, unit on y-axis is *Hz*. The color bar represents the absolute values of ERSP in dB.

By the start of the relaxation phase in this group a delta ERS (0 s – 5 s) can be seen over all presented channels and another alpha ERS (in the last second) at FC6. (The timing of the delta ERS appears peculiar, starting before the event at 0 s. This is most probably an effect of windowing in the ERSP method where time denotes the start of the window rather than its midpoint). There is no clear difference between left and right scalp sites in this case, among others simply for the reason that participants were free in imagining movements of either their left or their right foot.

6.5. Materials and Methods for BCI

6.5.1. Preprocessing the data

Preprocessing consisted of filtering and artifact removal. The hardware is equipped with a band-pass filter between 0.2 Hz and 45 Hz and two notch filters at 50 Hz and 60Hz to remove power line noise. In each trial the mean values over time of the imagination and relaxation signals were calculated and subtracted from each respective signal.

6.5.2. Artifact Suppression: Adaptive Filter

In this application an adaptive artifact reduction approach was used which in principle is similar to the one introduced in Chapter 5. In the current application the positions used for artifact reduction were recorded from AF3 and AF4 which is in contrast to previous applications and enabled us to measure horizontal eye movements, which is assumed to be essential in these BCI setups where arrows point left and right. AF3 and AF4 are not ideal, each one being situated above one eye. More lateral sites would have been a much better choice, e.g., AF7 and AF8, or, best, electrodes placed at the outer rims of the eyes, but the commercial hardware used in this setup restricted our choice. Data from AF3 and AF4 were used as inputs to a series of adaptive filters in order to eliminate the effect of artifacts recorded in these channels. Figure 6.7 shows a sample of the analysis on this dataset for channel F4. In Table 6.1 the calculated parameters are reported.

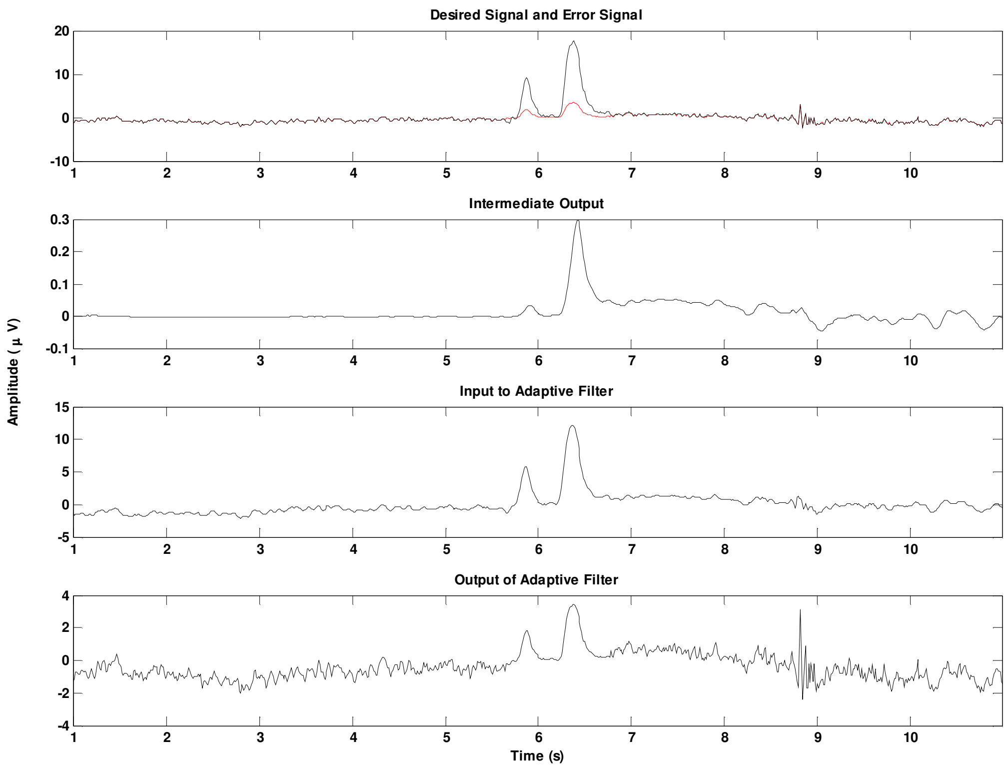


Figure 6.7 Artifact reduction using information based adaptive filter with Gabor filter with the sigma update over the time applied in 10 s ongoing EEG data. The output of the adaptive filter is the artifact-free EEG. The upper panel shows the measured signal at F4 in black, and the corrected signal in red (output of the adaptive filter). The reference signal (input to the adaptive filter) is the signal recorded from AF3 and AF4 smoothed by a moving-average filter.

Table 6.1: Averaged correntropy values over 12 channels (all except AF3 and AF4) for the three subjects using adaptive filter method for the current dataset.

Gabor kernel filter with $\boldsymbol{\sigma}$ adaptation	**AF- artifact reduction evaluation**		
	Correntropy signal	**Correntropy artifact**	**Performance**
S1	0.7969	0.1044	7.633
S2	0.9007	0.1201	7.499
S3	0.8114	0.1125	7.212

6.5.3. Feature Extraction using the Matching Pursuit approach

Matching pursuit (MP) was initially introduced by Mallat and Zhang and to date there are several extensions to the original formulation (Mallat, et al., 1993) (Durka PJ., 2006). MP algorithm is rather intuitive and simple. It is, in fact, a greedy decomposition of signals into a set of basic waveforms which are selected from a large and redundant dictionary of functions (Sanei, 2007). The robustness of the algorithm depends on the complexity of the combination of the dictionary's functions (Durka, et al., 2001)(Sanei, 2007). It forms an adaptive time-frequency representation of the signal and is based on an over-complete dictionary (Mallat, et al., 1993). Each iteration of the algorithm searches for the most informative projection, yielding an estimate that is basically free from arbitrary settings.

In other words, the MP algorithm tried to find the optimum projection of the input data onto an over-complete dictionary which matches best to the original data. The signal will be then presented as weighted sum of functions (atoms) g_{γ_n} selected from the dictionary. A Common choice for functions g_{γ_n} is the Gabor function, which is a sinusoidal modulated by a Gaussian.

$$\frac{1}{2\pi\sigma_x\sigma_y}\exp\left(-\frac{x^2}{2\sigma_x^2}-\frac{y^2}{2\sigma_y^2}\right)\cos(kx-\varphi) \tag{6.2}$$

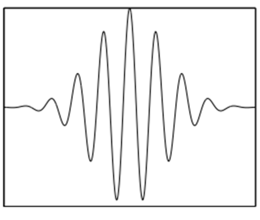


Figure 6.8: A sample of atoms used in the matching pursuit method. It is generated by changing the Gabor function parameters in order to generate a complete dictionary.

Mallat proposed that any signal $f(t)$ can be decomposed as following (Mallat, et al., 1993):

$$f(t) = \sum_{n=0}^{+\infty} a_n . g_{\gamma_n}(t) \tag{6.3}$$

here the *n* index represents the chosen atom, and a_n is the corresponding weight of that atom. Within a fixed dictionary, the MP algorithm starts to search for the atom having the greatest inner product with the input signal. In the next step the contribution of the first atom $(a_1 . g_{\gamma_1}(t))$ is subtracted from the signal. In an iteration scheme, the process continues until the point that the residue can be ignored (Durka, et al., 2001). We are interested in using a dictionary $D= \{g_{\gamma_n}\}$ *n=1…N* that provides the lowest entropy of the weights $\{a_n\}$, which is equivalent that the most intrinsic properties are extracted from the signal. Unfortunately, no universal recipe is available to date for such a prior choice.

The algorithm can be summarized in the following table. $f(t)$ and dictionary D are inputs, and the output is the list of coefficients (a_n, g_{γ_n}).

Table I: Algorithm of Matching Pursuit
1.Initialization: $R_1 \leftarrow f(t)$ $n \leftarrow 1$
2. Find $g_{\gamma_n} \in D$ with maximum inner product $\langle R_n, g_{\gamma_n}\rangle$
3. Let $a_n \leftarrow \langle R_n, g_{\gamma_n}\rangle$
4. Let $R_{n+1} \leftarrow R_n - a_n g_{\gamma_n}$
5.Let $n \leftarrow n + 1$
4. If not converged ($\|R_n\| < Threshold$), go back to 2

Several signal analysis approaches search for a linear expansion of the unknown signal $f(t)$ in terms of summation of basic functions. Sinusoidal basis functions construct the smallest possible complete dictionary and the algorithm results in the Fourier series representation of the signal. Fourier analysis reveals only global features of signals which does not adopt completely for the signal. A set of Gabor functions is a commonly used dictionary for the biological signals and in particular for the brain signals (Durka, et al., 2001). Real and imaginary features were extracted in time-frequency domain using the Gabor transform to visualize the differences between two motor tasks (Vuckovic et al., 2008). Real and imaginary parts of the Gabor coefficients were employed in a similar research (Miwakeichi, 2004).

It is claimed that the by employing the redundant dictionary, the multichannel matching pursuit approach can find an adaptive suboptimal solution to the problem of signal approximation (Mallat, et al., 1993). This approach might provide a sparse representation of the coefficients which is desirable for reduction of the signal analysis as well as coding and compression (Tipping M., 2001)

In the current application we used the multichannel matching pursuit algorithm in order to decompose the EEG recordings into a sum of atoms, each being the product of spatial (topographic) signature and waveforms having determined time–frequency localization.We consider the calculated weights $\{a_n\}$ and center frequency of the functions $\{g_{\gamma_n}\}$ as features for the BCI system. Figure 6.9 shows a set of features extracted using this method. The top plot presents the atoms used in the matching pursuit method. It is generated by changing the Gabor function parameters in order to generate a complete dictionary. Middle figure: The recorded and the approximated signal calculated by MP decomposition for one EEG channel (F4-5s) recorded from S2. Bottom figure: the feature set consists of values of 50 non-zero weights of used Gabor atoms. In order to decrease calculation time during the online experiment we applied a feature selection method to increase both the speed and the accuracy of the system.

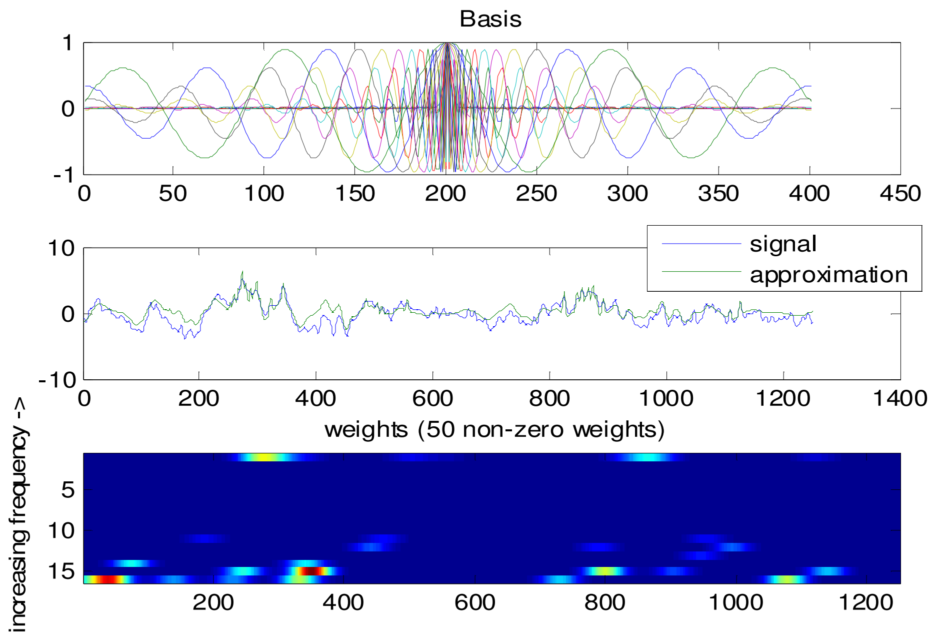


Figure 6.9: Top plot: Atoms used in the matching pursuit method. They are generated by changing the Gabor function parameters in order to generate a complete dictionary. Middle figure: The recorded and the approximated signal calculated by matching pursuit decomposition for one EEG channel (F4 – 5 s) recorded from S2. Bottom figure: the feature set consists of values of 50 non-zero weights of used Gabor atoms.

6.5.4. Feature selection: mRMR

We used the mRMR method described in chapter 4 for this BCI paradigm in order to decrease the dimension of the feature space. The method calculates the mean value of all mutual information values between any individual feature and any class over the offline dataset. The features are then selected if they fulfill the minimal redundancy and maximal relevance criterion at the same time. In the current BCI application, we used the mRMR method in order to decrease the feature space to 10 features as the input of each classifier. We treated all the features extracted from all recorded channels (excluding AF3 and AF4) equivalently. Using mRMR the best five features (weighting coefficients) extracted from each channel were used as features in BCI, excluding AF3 and AF4, yielding 5 x 12 = 60 features.

6.5.5. Classification: Ensemble LDA

Based on the offline analysis applied on the previous datasets and the lower computational effort we used LDA classifiers in this application. Similar to the previous applications we exploited the concept of combination of several classifiers on this BCI. Features extracted from each channel were fed to a group of classifiers. The final decision was made based on the combination of the outputs. Three LDA classifiers for each channel were estimated (Left, Right, Forward vs. Relax).

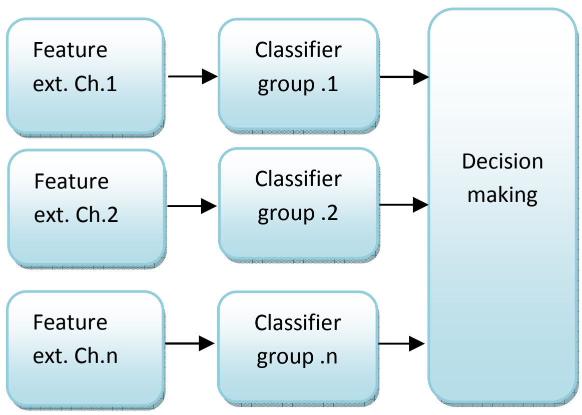


Figure 6.10: The idea of ensemble classification for different channel and decision making based voting

Time-frequency features with combination of an ensemble LDA classifier were used for decision making. The command is extracted from all channels (except AF3 and AF4) each second. A boosting algorithm was implemented for the decision making section.

6.5.5.1. Boosting

Overtraining happens when more data is fed to the classifier than is needed to construct the classifier. It usually leads to decreasing the generalization accuracy for classifiers. In the boosting method, later classifiers focus on the samples that were misclassified by earlier classifiers and each classifier is weighted with the error that it made. Boosting is a popular classifier for ensembling and it is unlikely to result in overtraining. The algorithm repeats in M iterations (Soria Frisch, 2012).

Table I - Boosting algorithm
1. Initialize example weights $w_i = {}^1/_N, i = 1, \dots, N$
2. For m= 1 to M a. Learn a classifier C_m using the current example weights b. Compute a weighted error estimate $$err_m = \frac{\sum w_i \ all\ incorrectly\ classified}{\sum_{i=1}^{N} w_i}$$ c. Compute the current classifier weight $$\alpha_m = \frac{1}{2} log\ ((1 - err_m)/err_m)$$ d. For all correctly classified examples e_i: $w_i \leftarrow w_i e^{-\alpha_m}$ e. For all incorrectly classified examples e_i: $w_i \leftarrow w_i e^{+\alpha_m}$ f. Normalizes the weights $\sum_{i=1}^{N} w_i = 1$
3. For each test example a. Calculate the output of all classifiers (C_1 to C_m) for the test sample b. Predict the class that has the largest sum of weights α_m

6.6. Results

6.6.1. Classification accuracy

After the offline phase, we applied the trained classifiers during the online experiments. Training was repeated for 4 runs, each run containing 12 trials for three movement imaginations. An average classification accuracy of 79.06% was obtained. For the online test the average classification accuracy was 72.71%. Table 6.2 summarizes the average of error rate and accuracy of the classification in training and test sessions for the three subjects.

Table 6.2: The average error rate using the test and the training data while applying different feature extraction techniques for S1. Also compiled are sensitivity, specificity, and Matthew's correlation coefficient using the offline data.

	Ensemble LDA				
Subjects/ Evaluation technique	**Sensitivity %**	**Specificity %**	**MCC**	**Training Accuracy% (offline)**	**Test Accuracy% (online)**
S1	84.09	72.88	0.604	77.62	76.45
S2	76.00	76.00	0.520	78.51	64.57
S3	78.43	75.59	0.580	81.05	77.04
Average	79.50	74.82	0.568	79.06	72.71

As the table shows, averaged sensitivity, specificity and MCC values were maximized for offline commands by using the ensemble classifier. Figure 6.11 illustrates the average classification accuracy during the online test for each subject. Training was repeated for 4 runs, each containing 12 trials for three movement imaginations. As can be seen, performance increased in the first three runs of the training phase for S2 and S3. The average classification accuracy is higher for all three subjects in this phase. During the online test, S1 and S3 show a positive trend. This may indicate that the interactive virtual reality had a positive effect on subjects' performance.

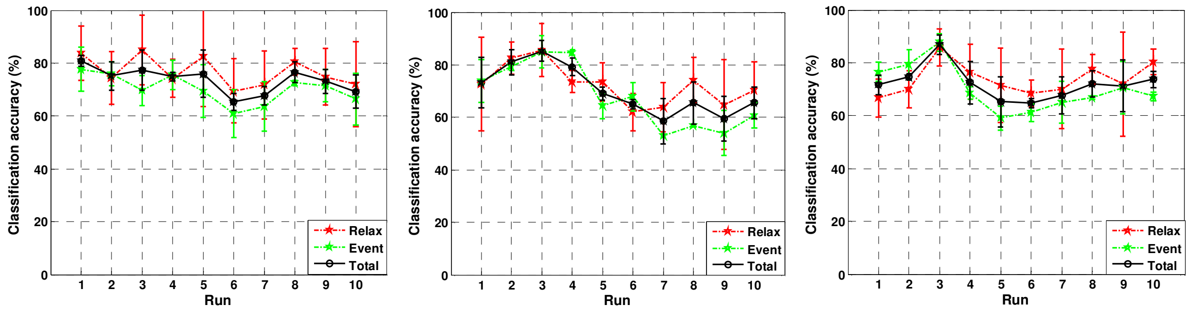


Figure 6.11: Performance of S1, S2 and S3 in several runs of the offline and online experiment. The red line corresponds to the stop command and the green line shows the classification accuracy in average for move commands (Right, Left or Forward). The Black line demonstrates the average performance for all four classes. The classification accuracy follows a smooth graph. This can be the proof for the robustness of the proposed classification method.

In the previous chapter the formula to measure the performance of a BCI system based on ITR was presented. Figure 6.12 depicts the theoretical ITR curves for three- and four-class BCIs and the empirical values of the ITR graph for this BCI application. For a four-class BCI problem, the maximum ITR in 95% classification accuracy can reach to 32 with maximum SNR. Empirically, in the avatar control online test, S3 achieved the highest ITR, amounting to 15.

Table 6.3: The average of information transfer rate in training and test session for three subjects.

LDA Method	ITR (Bit/min)	
	Training	Test
S1	13	11
S2	11	9
S3	15	12

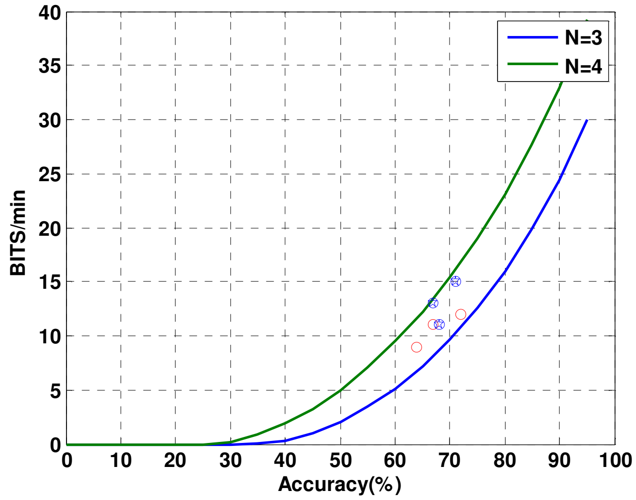


Figure 6.12: Theoretical information transfer rate for three and four class BCIs depicted by the blue and green curves, respectively. Circles show the empirical values from Table 6.3 for ITR and from Table 6.2 for accuracy. Red circles are the averaged values for the test runs and blue circles are the values for the training runs.

6.6.2. An offline study

Hitherto, it was unclear to what extent the accuracy of a BCI depends on the applied feature set and the classification approach. We applied an offline investigation on this dataset using all the methods suggested in this dissertation in order to evaluate the classification accuracies. The offline evaluation of different classifiers is reported for the training and test sessions for the three subjects. Figure 6.13 compares these methods with respect to average classification accuracy while applying different feature extraction and classification techniques.

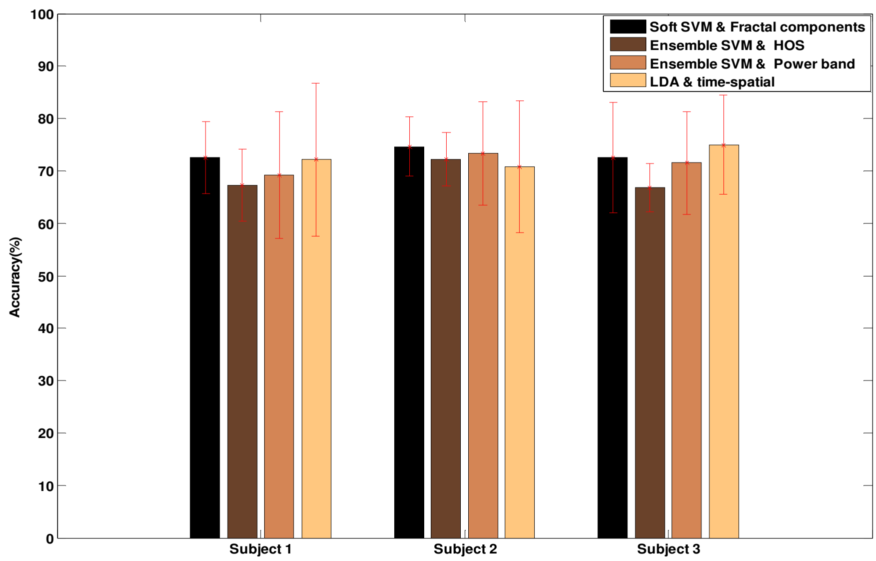


Figure 6.13: A general comparison between different integration methods using LDA, QDA and SVM classifiers and three different feature spaces extracted from the current dataset.

The MP method combined with LDA resulted in higher classification accuracy in comparison with band power and HOS features. In general Soft SVM with combination of fractal components shows stable and high classification accuracies. Both HOS features and power band features fed to ensemble SVM show lower performance; however, it varies over subjects. LDA has higher variance, similar to the previous application (Chapter 5) but its mean accuracy is comparable to the first group (soft SVM+ fractal components).

6.7. Discussion

In this project heuristic interaction techniques were developed that allow the user to freely control the robot or their avatar in the Second Life virtual environment by using mental activities. Systems are trained for some predefined movements. The training phase can be repeated until the desired classification accuracy is achieved. The simulated online implementation can help to improve the machine learning elements of the experiments.

Control was accomplished by online processing of single-trial EEG signals which were recorded using the EPOC headset from Emotiv. This is one of the first times the latest BCI technology and web-based virtual reality technology and online gaming are merged together. Our proposed method used a commercial EEG cap to pick up brain signals, and translated them into commands that were relayed to control the virtual avatar in a simple gaming scenario. It can be combined with a real world environment too: turning on and off the computer and opening other web pages or answering the phone are potential applications.

The quality of the recorded signal was lower than in the previous application. This can be due to the differences between recording systems, i.e., g.tec vs. Emotiv. A recent paper compared the performance of the Emotiv Epoc headset for P300 applications and a medical-grade system, the ANT device. The results suggest that the Emotiv headset performs significantly worse than the medical device (Duvinage et al., 2013).

We applied several criteria to evaluate our BCIs during the offline and online tests. The evaluation or comparison of the different BCI paradigms is not directly possible, because of enormous variety of factors that affected performance. For instance, the subjects participating in each experiment were not identical. However, the proposed evaluation methods like MCC and ITR can assess the system in a more advanced and general sense. In general, these principles can be applied to figure out which tasks or channels are optimal for conveying the desired message in a BCI setup. A long term study and more number of subjects are required to prove the generalization of the proposed methods.

Chapter 7
Conclusion and outlook

7.1. Perspective

The ultimate goal of the BCI system is to form a direct, reliable and robust communication channel between a human brain and a computer, bypassing the natural muscular and nervous pathways (McFarland, et al., 1988). Controlling a computer or communicating with external devices without error in daily situations purely by using thoughts seems yet a futuristic and fictional notion. Over the past four decades, several research groups worldwide have sought a reliable and robust methodology to realize the idea (Pfurtscheller et al., 1993) (Birbaumer et al., 2000) (Wolpaw, 2012). The vast research enthusiasm may enable us to reconnect the brain to a paralyzed limb or a robotic arm without surgery in order to add a new dimension and enhance the communication ability of humans (Blankertz et al., 2007) (Iturrate, et al., 2009). Such a possibility seemed remote when people originally started to apply noise like EEG signals from massive amplifier devices to run a binary control system with nearly random results (Berger, 1933)(Sanei, 2007). However the fast progress both in electromechanical engineering and machine learning science all over the world provoked the desire once again (Graimann et al, 2010). The original and still prevalent motivation for developing neural interfaces and neuro-prosthetics is to help disabled people and patients for rehabilitation and restoration of their lost functions (Allison, 2012). The very recent impetus behind developing BCIs is not only to help disabled individuals to recover or substitute their motor functions substitution, but also to provide a gaming modality to entertain healthy users (Brunner et al., 2011). We are witnessing certain technical trends that might affect drastically the fashion we interact with each other and also with our environment. Diverting the goal to design an easy wearable headset which remotely sends commands to an external device and offering the technology to the public for gaming and entertainment purposes opened up a new era of brain-machine interface research (Debener, 2012).

The proliferation of sophisticated communication and electronic devices may pave the way for a renaissance in brain computer interfacing (Wolpaw, 2012). In the near future BCIs may become publicly available machines which enhance the quality of life for everybody and in particular bring benefit to the elderly and disabled.

7.2. General discussion

Brain computer interface is a fascinating theme of research. During my PhD research I pursued both hardware and software investigations in this scope. In our team two alternative electrodes for brain signal recording were developed; near infrared recording and capacitive sensors (see the publication list in my resume). However, presented results in the dissertation are related to the EEG signals recorded with commercial devices.

I have proposed and evaluated several state-of-the-art algorithms for constructing the BCI elements in this dissertation. After building the theoretical foundation, I then designed prototype BCI scenarios. In essence, all proposed BCI paradigms have a uniform foundation which consists of imagination of movement as the main mental strategy. In the first application group (chapter 3 and 4) hand grasp control in both real and virtual environments were evaluated. In the gaming part (chapter 5 and 6), robot control and avatar control were tried.

I investigated several mathematical and advanced signal processing techniques to find the appropriate set of features in order to distinguish imaginations of limb movements from each other and from the relaxation state of the brain. The combinations of these features were evaluated with well-known classification techniques such as LDA and SVM but in an ensemble scheme. The distinguished idea of ensemble classification was implemented for several applications and the evaluation showed a significant increase both in the accuracy and in the speed of the entire system over several datasets. Self-recorded EEG data were exploited to compare the proffered schemes and to evaluate the performance of the entire system. Nonlinear mathematical analyses have been applied to both offline and online BCI studies.

With future refinement, this work can serve as a low-cost research tool that can be controlled by complex interfaces. Continued development yields to increase the range of movements the system can detect. I tried to critically evaluate the applied methods and the achieved results, thereby making clear whether the present approach should be further pursued, where the approach might be corrected etc.

7.3. Modifications

Current BCI research strives for increased accuracy and enhanced information transfer rate. There are several ways to achieve this goal. However, meeting one challenge may raise other challenges. The following solutions can be considered:

1- Long-term training of BCI users up to 300 hours to improve the classification accuracy
2- Improving advanced signal processing and machine learning algorithms
3- Extending the BCI system to multi classes
4- Designing faster trial-based BCIs
5- Applying invasive approaches for the recording of brain signals

Essential modifications and developments in both software and hardware technologies are required in order to make the BCI a daily practical technology:

Switch on/off strategy: In order to achieve a robust and reliable system on command, I suggest the closed eye strategy for more than 5s. Two characteristics, 1- no eye blink and 2- a strong alpha band power are measured to insure the command. This is a toggle command and is used to switch off the system.

Error correction is an interesting feature that can be implemented in the BCI paradigm and helps to improve the accuracy and the robustness of the entire system.

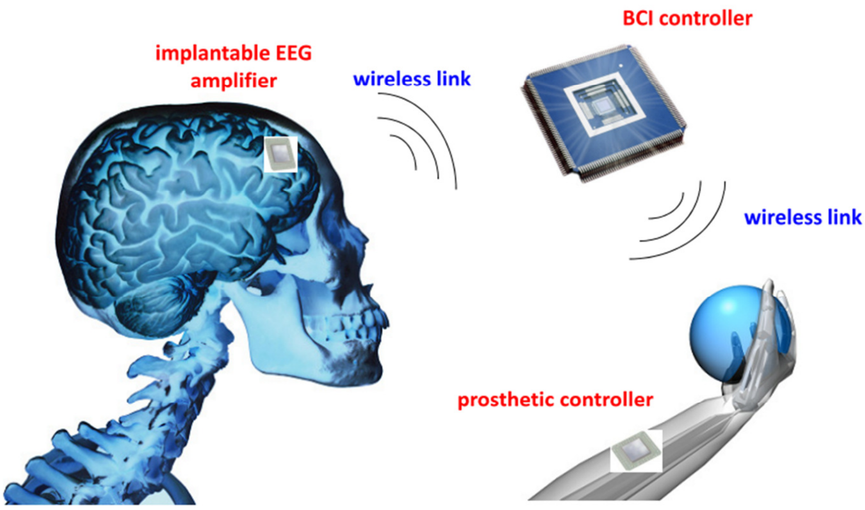
Since gaining BCI experience is a tedious, time consuming and error prone process, an easy to wear EEG cap is definitely preferred. A wireless connection is also very attractive because the wired system is more prone to noise. A comb shaped easy to wear EEG cap equipped with dry electrodes would be an ideal choice.

Improvement in the speed of the BCI system (bit/min) and long-term validation of results are needed. BCI technology should be validated in long-term studies for stability and from this point of view it is yet in a rudimentary stage of development.

The average classification accuracy for BCIs based on imagination of movement crossed the chance level, but is still far from being totally reliable. This fact limits the use of BCI in real-life without using robot intelligence. Error correction techniques based on mental task are not 100% accurate. It can be, however, an excellent option for gaming applications. A solution can be network BCIs when more than one person would try to use a single BCI system simultaneously, by generalizing the idea of the ensemble decision making.

7.4. Epilogue

David H. Hubel, the Nobel Prize winner, in 1982 stated, “in short, the brain can be studied, just as the kidney can.” After around twenty years, integrating technology with our biological systems seems more like an inevitable progression but the human brain remains as a mystery. The medias sometimes claim that scientists can already read the human mind. It is to some extent true but mostly absurd at this stage of development. By considering the current technical achievements, we catch a glimpse of what the future might hold. It is possible now to decipher certain functions, for instance moving arms or feet, with rather satisfactory accuracy in some subjects. However reading what one’s thinking about is not possible … yet. If this happens, one can imagine numerous applications. I anticipate the day that each electronic device around us has an identification label and we will need a password to work with them through our brains. It would solve the current problem and establish the ultimate interface, but would also bring along lots of sophisticated issues.



You never start completely from scratch!

References

Adlakha A. 2002. Single trial EEG classification. *Tech. Rep., Swiss Federal Institute of Technology.* 2002.

Allison B. Z., McFarland D. J., Schalk G., Zheng S. D., Jackson M. M., and Wolpaw J. R. 2008. Towards an independent brain-computer interface using steady state visual evoked potentials. *Clin Neurophysiol.* 2008, pp. 399-408.

Allison Brendan Z. 2012. *Towards Practical Brain-Computer Interfaces: Bridging the Gap from Research to Real-World Applications.* s.l. : Springer, 2012.

Alpaydin E. 2004. *Introduction to Machine Learning.* Cambridge, MA : The MIT Press, 2004.

Altman D. G and Bland, J. M. 1994. Diagnostic tests. 1: Sensitivity and specificity. *BMJ.* 1994, vol. 308, no. 6943.

Anderer A. 1999. Artifact processing in computerized analysis of sleep EEG - a review. *Neuropsychobiolog.* Sep 1999, vol. 40, no.3, pp. 150-157.

Argunsah A. O. and Cetin M. 2010. A brain-computer interface algorithm based on Hidden Markov models and dimensionality reduction. *Signal Processing and Communications Applications Conference, IEEE 18th.* 2010, pp. 93-96.

Arndt C. 2001. *Information measures, Information and its description in science and engineering.* s.l. : Springer, 2001.

Attwood H. L., and MacKay W. A. 1989. *Essentials of Neurophysiology.* Canada : B. C. Decker, Hamilton, 1989.

Au Wagner James. 2008. *The Making of Second Life.* New York : Collins. ISBN978-0-06-135320-8, 2008.

Baker K. S., Piriyapunyaporn T., and Cunnington R. 2012. Neural activity in readiness for incidental and explicitly timed actions Neuropsychologia. vol. 50, no. 5, pp. 715-22.

Ball T., et al. 2004. Towards a brain-machine interface based on epicortical field potentials. *Biomedical Engineering (Berlin).* 2004, 49 (Suppl. 2), pp. 756-759.

Ball T., et al. 2009. Signal quality of simultaneously recorded invasive and non-invasive EEG. *Neuroimage.* 2009, vol. 46, pp. 708-716.

Barbati G., Porcaro C., Zappasodi F., Rossini P. M., and Tecchio F. 2004. Optimization of an independent component analysis approach for artifact identification and removal in magneto encephalographic signals. *Clinical Neurophysiology.* 2004, pp. 1220-1232.

Barlow H. B. 1958. Sensory Mechanisms, the Reduction of Redundancy, and Intelligence. *The Mechanization of Thought Processes.* 1958, pp. 537-559. National Physical Laboratory Symposium.

Barlow J. S. 1986. Artifact processing (rejection and minimization) in EEG data processing. *Handbook of Electroencephalography and Clinical Neurophysiology (Revised Series Ed.).* Amsterdam : Elsevier, 1986, pp. 15-62.

Bear Mark, and Paradiso M. 2007. Exploring the Brain. *Neuroscience.* 2007.

Bell C. J. et al. 2008. Control of a humanoid robot by a noninvasive brain-computer interface in humans. *Journal of Neural Engineering.* 2008, vol 5, no. 2, pp. 214-220.

Berg P., and Scherg M. 1994. A multiple source approach to the correction of eye artifacts. *Electroenceph. Clinical Neurophysiology.* 1994, pp. 229-241.

Berg P., and Davies, M. B. 1988. Eyeblink-related potentials. *Electroencephalography and Clinical Neurophysiology* vol. 69, pp.1-5

Berger H. 1930. On the Electroencephalogram of Man. *Journal fur Psychologie und Neurologie.* 1930, vol. 40, pp. 60-179.

Berger H. 1933. Über das Elektrenkephalogramm des Menschen. *Archiv für Psychiatrie und Nervenkrankheiten.* 1933, vol. 99, no. 1, pp. 555-574.

Berger T. 2003. Living information theory: The 2002 Shannon Lecture. *IEEE Information Theory Society Newsletter.* March 2003, vol. 53, no. 1.

Bin G., Gao X., Yan Z., Hong B., and Gao S. 2009. An online multi-channel SSVEP-based brain computer interface using a canonical correlation analysis method. *Joural of Neural Engineering.* August 2009, vol. 6.

BinAb R., Rani, M.S., bt.Mansor, W. 2009. Detection of eye blinks from EEG signals for home lighting system activation. *ISMA '09. 6th International Symposium on Mechatronics and its Applications.* March 2009, vol. 1, no. 4, pp. 23-26.

Binnie C. D., Dekker E., Smit A. and Van der Linden G. 1982. Practical considerations in the positioning of EEG electrodes. *Electroenceph. Clin. Neurophysiol.* 1982, vol. 53, pp. 453-458.

Birbaumer N., Kübler A., Ghanayim, N., Hinterberger T., Perelmouter J., Kaiser J., Iversen I., Kotchoubey B., Neumann N., Flor H. 2000. The Thought Translation Device (TTD) for Completely Paralyzed Patients. IEEE Transactions on Rehabilitation 2000, vol. 8, no. 2, pp. 190-193.

Bishop C. M. 1995. *Neural Networks for Pattern Recognition.* Oxford : Oxford University Press, 1995.

Blakely T, Miller K. J., Zanos S. P., Rao R. P., Ojemann J. G. Robust 2009. long-term control of an electrocorticographic brain-computer interface with fixed parameters. *Neurosurg Focus.* 2009.

Blankertz et al. 2006. The Berlin brain-computer interface: EEG-based communication without subject training. *IEEE Trans. on Rehabilitation Engineering.* 2006, pp. 147-152.

Blankertz B., Curio G. , Müller K. R. 2002. Classifying single trial EEG: towards brain computer interfacing. *Adv. Neural Inf. Process. Syst. (NIPS 01).* 2002, pp. 157-164.

Blankertz B. , Dornhege G., Krauledat M., Müller K., and Curio G. 2007. The non-invasive Berlin Brain-Computer Interface: fast acquisition of effective performance in untrained subjects. *NeuroImage.* August 2007, vol. 37, pp. 539-550.

Blume W. T., Kaibara M. 1999. *Atlas of Pediatric Electroencephalography.* Philadelphia:2nd ed. : Lippincott-Raven, 1999.

Bogdanov A. V. 2008. Neuroinspired architecture for robust classifier fusion of multisensory imagery. *IEEE Trans Geosci. Remore Sens.* 2008, pp. 1467-1487.

Boostani R., and Moradi M. 2004. A new approach in the BCI research based on fractal dimension as feature and Adaboost a classifier. *Journal Neural Eng.* 2004, pp. 212-217.

Boser B. E., Guyon I. and Vapnik, V. 1992. A training algorithm for optimal margin classifiers. *In Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory.* 1992, ACM, Madison, WI, pp. 144-152.

Bougrain Laurent and Liang Nanying. 2009. Band-specific features improve Finger Flexion Prediction from ECoG. *Jornadas Argentinas sobre Interfaces Cerebro Computadora, JAICC.* 2009.

Bradley, Ronald J. and Harris, Adron R. 2009. *International review of neurobiology.* London : Elsevier, 2009. vol. 86.

Breiman L. 1998. Arcing classifiers. *The Annals of Statistics.* 1998, pp. 801-849.

Breiman L. 2001. Random Forests. *Machine Learning.* 2001, p. 32.

Brodu Nicolas, Lotte Fabien and Lécuyer Anatole. 2012. Exploring Two Novel Features for EEG-based Brain-Computer Interfaces: Multifractal Cumulants and Predictive Complexity. *Neurocomputing 79.* 2012, pp. 87-94.

Brooks Jr., F. P. 1999. What's Real About Virtual Reality? *IEEE Computer Graphics And Applications.* 1999, vol. 19, no. 6, 16.

Brunner P., Bianchi L., Guger C., Cincotti F., and Schalk G. 2011. Current trends in hardware and software for brain–computer interfaces (BCIs). *Journal of Neural Eng.* 2011, vol. 8, p. 025001 7 pages.

Brutsch et al., 2011. VR for enhancement of robot-assisted. *Journal of Rehabilitation Med.* vol. 43. pp. 493-499, 2011.

Bundhoo, V. 2005. Design of an artificial muscle actuated finger towards biomimetic prosthetic hands. *IEEE.* 2005.

Campos Viola, F., Thorne, J., Edmonds, B., Schneider, T., Eichele, T., and Debener, S. 2009. Semi- automatic identification of independent components representing EEG artifact. *Clinical Neurophysiology*, vol. 120, pp. 868-877.

Cannan James, and Huosheng Hu. 2010. *Human machine interaction (HMI): A survey,.* s.l. : Technical report:CES-508, 2010.

Carmena J . M., Lebdev M. A., Crist R.E., O'Doherty J. E., Santucci D. M., Dimitrov D. F., Patil P. G., Henriques C., and Nicolelis M. A. 2003. Learning to control a brain-machine interface for reaching and grasping by primates. *PLOS Biology.* 2003, vol. 1, no. 2, pp. 193-208.

Carrol J. M. 1997. Human-Computer Interaction: Psychology as a Science of Design. *Annu. Rev. Psyc.* 1997, 48, pp. 61-83.

Cauchy, Pradhan, et al. 2012. Higher-Order Spectrum in Understanding Nonlinearity in EEG Rhythms. *Computational and Mathematical Methods in Medicine.* 2012, Article ID 206857.

Chang C. C., Lin C. J. 2001. LIBSVM: A Library for Support Vector Machines. [Online] 2001. www.csie.ntu.edu.tw/~cjlin/libsvm .

Chao Z. C., Nagasaka, Y. and Fujii, N. 2010. Long-term asynchronous decoding of arm motion using electrocorticographic signals in monkeys . *Frontiers Neuroengineering.* 2010, vol. 3, no. 3.

Chapin J. K. 1999. Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex. *Nature Neuroscience.* 1999, vol. 2, pp. 664 -670.

Cheron G., et al. 2007. Pure phase-locking of beta/gamma oscillation contributes to the N30 frontal component of somatosensory evoked potentials. *BMC Neuroscience.* 2007, pp. 8-75.

Chomsky N. 2006. *Language and mind.* s.l. : Massachusetts Institute of Technology,Third edition, 2006.

Clausen J. 2009. Man, machine and in between. *Nature.* 2009, vol. 457, pp. 1080-1081.

Clausius Rudolf. 1865. *Mechanical theory of heat.* 1865.

Coates Thomas D., et al. 2008. *Neural Interfacing: Forging the Human-Machine Connection.* s.l. : Morgan & Claypool Publishers series, 2008.

Corralejo R., Hornero R. and Alvarez D. 2011. Feature selection using a genetic algorithm in a motor imagery-based Brain Computer Interface. *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE* . 2011, pp. 7703-7706.

Cortes C. and Vapnik V. 1995. Support vector networks. *Machine Learning.* 1995, vol. 20, pp. 1-25.

Coyle, S., et al. 2004. On the suitability of near infrared (NIR) systems for next-generation brain-computer interfaces. *Physiological Measures.* 2004, vol. 25, no. 4, pp. 815-822.

Coyle, Shirley, Ward, Tomás, Markham, Charles. 2010. *Brain computer interfaces- A review.* 2010.

Crespo-Garcia M., Atienza M., LCantero J. 2008. Muscle artifact removal from human sleep EEG by using independent component analysis. *AnnBiomed Eng.* 2008, pp. 467-475.

Croft R. J., Barry R. J. 2000. Removal of ocular artifact from EEG: a review. *Neurophysiol Clin.* Feb. 2000, vol. 30, no.1, pp. 5-19.

Coles Michael G. H. 1989. Modern Mind-Brain Reading: Psychophysiology, Physiology and Cognition, Psychophisiology, vol. 26, no. 3, 1989.

Curran E., Sykacek P., Stokes M., Roberts S. J., Penny. W., Johnsrude I. 2004. Cognitive tasks for driving a brain–computer interfacing system: a pilot study. *IEEE Trans Neural Syst Rehabil Eng.* 2004, pp. 48-54.

Daly I., Nasuto S.J. and Warwick K. 2012. Brain computer interface control via functional connectivity dynamics. *Pattern Recognition.* 2012, pp. 2123-2136.

Dash M., and Liu H. 1997. Feature Selection for Classification. *Intelligent Data Analysis.* 1997, 1, pp. 131-156.

Dayan Peter, Abbott L. F. 2001. *Theoretical neuroscience: Computational and Mathematical Modeling of Neural Systems.* s.l. : Massachusetts Institute of Technology, 2001.

Debener Stefan, Minow Falk, Emkes Reiner, Gandras Katharina, Maarten de Vos 2012. How about taking a low-cost, small, and wireless EEG for a walk? *Psychophysiology, PMID 23013047.* 2012, vol. 49, no. 11, pp. 1617-21.

Delorme A., Makeig S., Sejnowski T. 2001. Automatic artifact rejection for EEG data using high-order statistics and independent component analysis. *Proceedings of the 3rd International Workshop on ICA.* Dec. 2001, pp. 457-462.

Dennett D. C. 1992. *Consciousness explained.* s.l. : Back Bay Books, Lippincott Williams & Wilkins, 1992.

Diez Pablo F., et al. 2009. Application of the Empirical Mode Decomposition to the Extraction of Features from EEG Signals for Mental Task Classification. *31st Annual International Conference of the IEEE EMBS.* 2009, pp. 2579-2582.

Diniz V. 2008. *Adaptive Filtering.* s.l. : Springer Science+Business Media, LLC, 2008. DOI: 10.1007/978-0-387-68606-6_1.

Dispenza J. 2008. *Evolve Your Brain: The Science of Changing Your Mind.* 2008.

Donoghue, John P. 2007. Assistive technology and robotic control using motor cortex ensemble-based neural interface systems in humans with tetraplegia. *Journal of Physiol 579.3 , pp, 2007.* 2007, vol. 579, no. 3, pp. 603-611.

Dornhege G., Blankertz B., Curio G., and Muller K.-R. 2003. Increase information transfer rates in BCI by CSP extension to multiclass. *Advances in Neural Inf. Proc. Sys. (NIPS03).* 2003.

Dornhege G., Millan J. del R., Hinter-berger T., McFarland D. J., Muller K.-R. 2007. *Toward Brain-Computer Interfacing.* s.l. : The MIT Press, Massachusetts Institute of Technology, 2007.

Durka P. J., Dobieslaw, I. and and Blinowska, K. J. 2001. Stochastic time–frequency dictionaries for matching pursuit. *IEEE Trans. Signal Process.* March 2001, vol. 49, no. 3.

Durka P. J. 2006. Time–frequency microstructure and statistical significance of ERD and ERS. *Prog Brain Res, 159.* 2006, 159, pp. 121-133.

Duvinage Matthieu, Castermans Thierry, Petieau Mathieu, Hoellinger Thomas,Guy Cheron, and Thierry Dutoit, 2013. Performance of the Emotiv Epoc headset for P300-based applications Biomed Eng Online. 2013; vol. 12, no. 56.

Elbert T., Ray W. J. Kowalik Z.J., Skinner J. E., Graf K. E. and Birbaumer N. 1994. Chaos and Physiology: Deterministic Chaos in Excitable Cell Assemblies. *Phys.Rev.* 74 1994, pp. 1-47.

Erfanian A., and Mahmoudi B. 2005. Real-time ocular artifact suppression using recurrent neural network for electro-encephalogram based brain–computer interface. *Med. Biol. Eng. Comput.* 2005, pp. 296-305.

Esteller R., et al. 2001. A Comparison of Waveform Fractal Dimension Algorithms. *IEEE Trans. On Circuits and Systems-I: Fundamental Theory and Application.* February 2001, vol. 48, no. 2, pp. 177-183.

Farwell L.A. and Donchin E. 1986. The brain detector: P300 in the detection of deception. *Psychophysiology* . 1986, vol. 24, no. 434.

Farwell L. A. and Donchin E. 1988. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical.* 1988.

Fatourechi M., Bashashati A., Ward R. K., Birch G. E. 2007. EOG and EMG Artifacts in Brain Interface Systems: a Survey. *Clinical Neurophysiology.* March 2007, vol. 118, no. 3, pp. 480-494.

Fausett Laurene. 2010. *Fundamentals of Neural Networks: Architecture، Algorithms and Application.* 2010?

Fetz E. E. and Finicchio D. V. 1975. Correlations between activity of motor cortex cells and arm muscles during operantly conditioned response patterns. 1975, , vol. 23, no. 3, pp. 217-240.

Fisher R. A. 1936. The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics.* 1936, vol. 7, no. 2, pp. 179-188.

Fitzgibbon S. P., Powers D. M. W., Pope K. J., and Clark C. R. 2007. Removal of EEG Noise and Artifact Using Blind Source Separation. *Clinical Neurophysiology.* 2007, pp. 232-243.

Flamary R., and Rakotomamonjy A. 2012. Decoding finger movements from ECoG signals using switching linear models. *Frontiers in Neuroscience.* 2012, vol 6, no. 29.

Fleischl V., Marxow E. 1890. Mittheilung, betreffend die Physiologie der Hirnrinde. *Central bl Physiol.* 1890, 4, pp. 537-540.

Frank, R.M., and Frishkoff, G. 2007. Automated protocol for evaluation of electromagnetic component separation (APECS): Application of a framework for evaluating statistical methods of blink extraction from multichannel EEG. Clinical Neurophysiology 118, pp. 80-97.

Fraser A. M. and Swinney H. L. 1986. Independent coordinates for strange attractors from mutual information. *Physics. Review* 1986, A 33, p. 1134.

Friedman J. H. K. 1997. On bias, variance, 0/1-loss, and the curse-of-dimensionality. *Data Mining and Knowledge Discovery.* 1997.

Friedman J. H. 1989. Regularized Discriminant Analysis. *Journal of the American Statistical Association (American Statistical Association).* 1989, vol. 84, no. 405, pp. 165-175.

Friman O., Volosyak I., and Gräser A. 2007. Multiple channel detection of steady-state visual evoked potentials for brain-computer interfaces. *IEEE Trans Bio-Med Eng.* April 2007, vol. 54, pp. 742-750.

Fukunaga K. 1990. *Statistical Pattern Recognition.* New York: : Academic, 2nd edition, 1990.

Gao X., Xu D., Cheng M., Gao S. 2003. A BCI-based environmental controller for the motion-disabled. *IEEE Trans Neural Syst Rehabil Eng.* June 2003, vol. 11, pp. 137-140.

Geng T., Dyson M., Tsui CS., Gan JQ..2007. A 3-class asynchronous BCI controlling a simulated mobile robot. *Conf Proc IEEE Eng Med Biol. Soc.* 2007, pp. 2524-7.

Geng T., et al. 2008. A Novel Design of 4-Class BCI Using Two Binary Classifiers and Parallel Mental Tasks. *Computational Intelligence and Neuroscience.* 2008, pp. 1-5.

Georgiev S., et al. 2009. EEG Fractal Dimension Measurement before and after Human Auditory Stimulation. *Bio Automation.* 2009, vol. 12, pp. 70-81.

Gilden D., Thornton, T. and Mallown, M. 1995. 1/f Noise in Human Cognition. *Science, 267.* 1995, pp. 1837-1839.

Gil-Gomez et al. 2011. Effectiveness of Wii balance board system for balance rehabilitation in patients with acquired brain injury. *Journal of NeuroEngineering and Rehabilitation.* 2011, vol. 8 no. 30.

Goncharova I. I., McFarland D. J., Vaughan T. M. and Wolpaw J. R. 2003. EMG contamination of EEG: spectral and topographical characteristics. *Clinical Neurophysiology.* Sep. vol. 114, no.9, 2003, pp. 1580-1593.

Gonzalez B., and Sanei, S. and and Chambers, J.,. 2003. Support vector machines for seizure detection. *Proceedings of the IEEE ISSPIT.* December 2003, pp. 126-129.

Graimann B., Allison Brendan, Pfurtscheller, Gert. 2010. *Brain-Computer Interfaces:Revolutionizing Human–Computer Interaction.* s.l. : Springer-Verlag Berlin Heidelberg, 2010, 2010.

Grandvalet Y. 2004. Bagging equalizes influence. *Machine Learning.* 2004, vol. 55, no. 3, pp. 251-270.

Greb U. and Rusbridge M. G. 1988. The interpretation of the bispectrum and bicoherence for non-linear interactions of continuous spectra. *Plasma Phys. Control. Fusion.* 1988, vol. 30, no. 5, pp. 537-549.

Greco A., Mammone N., Morabito F. M. , Versaci M. 2005. Semi-Automatic Artifact Rejection Procedure based on Kurtosis, Renyi's Entropy and Independent Component Scalp Maps. *World Academy of Science, Engineering and Technology.* 2005, pp. 22-26.

Greiner M., Pfeiffer, D and and Smith, R. D. 2000. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Preventive Veterinary Medicine.* 2000, vol. 45, pp. 23-41.

Guger C., et al. 2009. How many people are able to control a P300-based brain-computer interface (BCI)? *Neurosci Letter, 462.* Oct. 2009, pp. 94-98.

Halder S., et al. 2013. Prediction of auditory and visual P300 brain-computer interface aptitude. *PLoS ONE.* 2013, vol. 8.

Hartley R. V. L. 1928. Transmission of Information. *Bell System Technical Journal.* July 1928, p. 535.

Hastie T., Tibshirani R., Friedman J. 2008. *The Elements of Statistical Learning.* 2008.

Haufe S., Treder M. S., Gugler M. F., Sagebaum M., Curio G., Blankertz B. 2011. EEG potentials predict upcoming emergency brakings during simulated driving. *Journal of neural engineering.* 2011, vol. 8, no. 5.

Haykin, Simon. 1996. *Adaptive Filter Theory.* s.l. : Prentice-Hall, Eaglewood Cliffs, 3rd edition, 1996.

Haykin S., et al. 2007. *New Directions in Statistical Signal Processing: From systems to brains.* Cambridge, Massachusetts : The MIT Press, Massachusetts Institute of Technology,, 2007.

Hazrati M. Kh. , Erfanian A. 2008. An On-line BCI for Control of Hand Grasping Sequence and Holding Using Adaptive Probabilistic Neural Network. 2008.

Hazrati M. Kh., Subramanian V., Hofmann U. G. 2011. A Four-Class Brain Computer Interface for Robot Control. *ISSNIP.* 2011.

Hazrati, M. Kh., and Erfanian A. 2008. An On-line BCI for Control of Hand Grasping Sequence and Holding Using Adaptive Probabilistic Neural Network. 2008, pp. 1009-12.

Heussen, Yana, Ferdinand Binkofski, Jacob Jolij. 2010. The semantics of the lying face - An EEG Study. *Internation Journal of Psychophysiology.* Jan 2010, vol. 77, no. 3. 206.

Higuchi T. 1988. Approach to an Irregular Time Series on the Basis of the Fractal Theory. *Physica D: Nonlinear Phenomena,.* 1988, vol. 31, pp. 277-283.

Hinterberger et al. 2003. Brain-Computer Communication and Slow Cortical Potentials. *TBME-00279.* 2003.

Ho T. K. 1988. The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.* 1988, vol. 20, no. 8, pp. 832-844.

Hochberg Leigh R. 2012. Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature.* 2012, 485, pp. 372-375.

Hochberg L. R, et al. 2006. Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature.* 2006, 442, pp. 164-171.

Howell, David. 2002. Statistical methods for Psychology. *Duxbury.* 2002, pp. 324-325.

Huan, Liu and Motoda, Hiroshi. 2008. *Computational methods of feature selection.* s.l. : Taylor & Francis Group, LLC, 2008.

Huang. 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. of the Royal Society London,.* 1998, Ser. A, vol. 454, pp. 903-995.

Hubel D. 1979. The Brain. *Scientific American.* 1979, vol. 241, no. 3, pp. 45-53.

Hyvärinen A. 1999. Fast and Robust Fixed-Point Algorithms for Independent Component Analysis. *IEEE Transactions on Neural Networks.* 1999, pp. 626-634.

Iturrate, Inaki and Antelis, Javier, A. K. and Minguez, J. 2009. A Noninvasive Brain-Actuated Wheelchair Based on a P300. *Neurophysiological Protocol and Automated Navigation.* 2009.

Iwasaki, M, Kellinghaus C, Alexopoulos AV. 2005. Effects of eyelid closure, blinks, and eye movements on the electroencephalogram. *Clinical Neurophysiology.* 2005, pp. 878-885.

Jackson, Andrew. 2012. Neuroscience: Brain-controlled robot grabs attention. *Nature.* 2012, 485, pp. 317-318.

Jain A. K., and Chandrasekaran B. 1982. *Dimensionality and sample size considerations in pattern recognition practice, Handbook of Statistics.* 1982.

James C. J., and Gibson O. J. 2003. Temporally constrained ICA: An application to artifact rejection in electromagnetic brain signal analysis. *IEEE Trans. Biomed. Eng.* 2003, pp. 1108-1116.

Jian-Wu, Xu, et al. 2003. Recursive Renyi's entropy estimator for adaptive filtering. *Proceedings of the 3rd IEEE International Symposium on Signal Processing and Information Technology.* Dec. 2003, pp. 134-137.

Johansen, J. W. and Sebel P. S. 2000. Development and clinical application of electroencephalographic bispectrum monitoring. *Anesthesiology.* No. 2000, vo. 93, no. 5, pp. 1336-1344.

Jones, Kleim. 2008. *Neuroplasticity.* 2008.

Joyce C. A., Gorodnitsky I. F. and Kutas M. 2004. Automatic removal of eye movement and blink artifacts from EEG data using blind component separation. *Psychophysiology.* 2004, pp. 313-325.

Jung T-P., Makeig S, Humphries C, Lee T-W, Mckeown M J, Iragui V and Sejnowski T J. 2000. Removing electroencephalographic artifacts by blind source separation. *Psychophysiology.* 2000, pp. 163-78.

Jung T-P., Humphries C., Lee M., Iragui V., Makeig S., Sejnowski T. 1998. Removing electroencephalographic artifacts: Comparison between ICA and PCA. *IEEE International Workshop on Neural Networks for Signal Processing.* 1998, pp. 63-72.

Kalcher J., Flotzinger D., Pfurtscheller G. 1993. Graz brain-computer interface: an EEG-based cursor control system. 1993, pp. 1264-1265.

Kaplan, Alexander Y. 2013. Adapting the P300-based brain-computer interface for gaming: a review. *, IEEE Transactions on Computational Intelligence and AI in Games (Special Issue on Brain/Neuronal-Computer Games Interfaces and Interaction).* 2013, in press.

Karat, John,Vanderdonckt, Jean. 2010. *Brain-Computer Interfaces: Applying our minds to Human-Computer Interaction.* London : Springer-Verlag, 2010.

Kato P. M., Cole S. W., Bradlyn A. S., Pollock B. H. 2008. Pediatrics. A video game improves behavioral outcomes in adolescents and young adults with cancer: a randomized trial. 2008, vol. 122, no. 2, pp. 305-317.

Kelly S. P., Lalor E. C., Reilly R. B., Foxe J. J. 2005. Visual spatial attention trackingusing high-density SSVEP data for independent brain-computer communication. *IEEE Transactions on Neural Systems and Rehabilitation Engineering.* 2005, vol. 3, no. 2, pp. 172-178.

Kempe L., 2010. Serious Games für die Gesundheit: Spielen für eine ernste Sache. *Dtsch Arztebl.* 2010, vol. 107, no. 38.

Kennedy, P.R., Bakay, R.A.E., Moore, M.M., Adams,K., Gold waithe, J.,. 2000. Direct Control of a Computer from the Human Central Nervous System. *IEEE Transactions on Rehabilitation Engineering.* 2000, vol. 8, no. 2, pp. 198-202.

Keren, A.S., Yuval-Greenberg, S., & Deouell, L.Y. 2010. Saccadic spike potentials in gamma-band EEG: Characterization, detection and suppression. NeuroImage vol. 49, pp: 2248-2263.

Keshner, Emily A,. 2004. *Virtual reality and physical rehabilitation: a new toy or a new research and rehabilitation tool?, Journal of Neuroengineering and rehabilitation.* 2004, vol. 1, no. 8.

Klass, DW. 1995. The continuing challenge of artifacts in the EEG. *Journal of EEG Technology.* 1995, pp. 239-269.

Kobayashi, M., Musha, T. 1982. 1/f fluctuation of heart beat period. *IEEE Transactions on Biomedical Engineering.* 1982, 29, pp. 456-457.

Kołodziej, Marcin, Majkowski, Andrzej and and Rak, Remigiusz J. 2011. A New Method of EEG Classification for BCI with Feature Extraction Based on Higher Order Statistics of Wavelet Components and Selection with Genetic Algorithms. *ICANNGA'11 Proceedings of the 10th international conference on Adaptive and natural computing* . 2011, pp. 280-289.

Koprinska, Irena. 2009. Comparison of Feature Selection Methods for Classification of Brain-Computer Interface Data. *Workshop on Advances and Issues in Biomedical Data Mining* . 2009, pp. 41-50.

Kornhuber H. H., and Deecke L. 1965. Changes in the brain potential in voluntary movements and passive movements in man: Readiness potential and reafferent potentials Pflugers Archiv fur die gesamte Physiologie, vol 284, no. 1.

Krasner, S. 1990. The ubiquity of chaos. *American Association for the Advancement of Science.* 1990.

Krepkiy, Roman. 2008. *Brain-Computer interface: Design and implementation of an online BCI system for the control in gaming application an virtual limb.* s.l. : VDM Verlag Dr Muller, 2008.

Kronegg J., Voloshynovskiy S. and Pun T. 2005. Analysis of bit-rate definitions for Brain-Computer Interfaces. *presented at Int. Conf. on Human-computer Interaction (HCI'05).* 2005.

Kronegg Julien, Voloshynovskiy Sviatoslav and Pun Thierry. 2005. Information-transfer rate modeling of EEG-based synchronized brain-computer interfaces. *Technical Report 0503,.* December 20 2005. http://vision.unige.ch/publications/.

Krusienski Dean J., SellersEric W. and Cabestaing Francois. 2006. A comparison of classification techniques for the P300 Speller . *Journal of Neural Engineering.* December 2006, vol. 3, no. 4., pp. 299-305.

Kübler A. , Neumann N. 2005. Brain-computer interfaces- the key for the conscious brain locked into a paralyzed body. *Progress in Brain Research.* 2005, pp. 513-525.

Kuebler A. , Kotchoubey B., Kaiser J., Wolpaw J. R., and Birbaumer N. 2001. Brain-computer communication: unlocking the locked in. *Psychology Bulletin.* 2001, pp. 358-375.

Kuncheva, L. I. 2004. *Combining pattern classifiers: Methods and algorithms.* Hoboken, New Jersey : Wiley, 2004.

Lagerlund T. D., Sharbrough F. W. and and Busacker N. E. 1997. Spatial filtering of multichannel electroencephalographic recordings through principal component analysis by singular value decomposition. *J. Clin. Neurophysiol.* 1997, vol. 14, no. 1, pp. 73-82.

Lange et al. 2011. PrimeSense camera and game for upper extremity function. 2011.

Lawrence N. D., Herbrich R. 2001. A sparse Bayesian compression scheme - the informative vector machine. *Neural Information Processing Systems Workshop on Kernel Methods,.* Dec. 2001.

Lebdev M. A., and Nicolelis, M. A. L. 2006. Brain-Machine Interfaces: past, present and future. *Trends in Neuroscience.* 2006, vol. 29, no. 9, pp. 536-545.

Leuthardt E. C., et al. 2004. Brain–computer interface using electrocorticographic signals in humans. *Journal of Neural Engineering.* 2004, pp. 63-71.

LeVan P., Urrestarazu E. and Gotman J. 2006. A system for automatic artifact removal in ictal scalp EEG based on independent component analysis and Bayesian classification. *Clinical Neurophysiology.* 2006, pp. 912-927.

Liang Yizeng and Xu Qing-Song. 2011. *Support vector machines and their application in chemistry and biochemistry.* s.l. : CRC press, 2011.

Liu Hailong, Wang Jue and Zheng Chongxun. 2005. Using Self-organizing Map for Mental Tasks Classification in Brain-Computer Interface. *Advances in Neural Networks – ISNN 2005 , Lecture Notes in Computer Science.* 2005, vol. 3497, pp. pp 327-332.

Liu, Y., Zhou, Z., and Hu, D. 2011. Gaze independent brain-computer speller with covert visual search tasks. Clinical Neurophysiology 122, pp. 1127-1136

LooAlfred W. 2003. The future of peer-to-peer computing. *Commun. ACM 46.* September 9, 2003, pp. 56-61.

Loong T. 2003. Understanding sensitivity and specificity with the right side of the brain. *BMJ 327* . 2003, 7417, pp. 716-719.

Lotte F., Congedo M., Lécuyer A., Lamarche F., Arnaldi B. 2007. A review of classification algorithms for EEG-based brain-computer interfaces. *J Neural Eng.* 4 2, 2007, pp. R1-R13.

Lutz H. Hamel. 2009. *Knowledge discovery with support vector machines.* s.l. : John Wiley& Sons, 2009.

Lytton William W. 2002. *From Computer to Brain: Foundations of Computational Neuroscience.* New York Inc., : Springer-Verlag, 2002.

Makeig S., Bell A. J., Jung T-P, and Sejnowski T. J. 1996. Independent component analysis of Electroencephalographic data. *Advances in Neural Information Processing Systems.* 1996, pp. 145-151.

Makeig Scott. 1993. Auditory Event-Related Dynamics of the EEG Spectrum and Effects of Exposure to Tones. *Electroencephalography and Clinical Neurophysiology.* 1993, vol. 86, pp. 283-293.

Mallat S. G. and Zhang, Z. 1993. Matching Pursuits with Time-Frequency Dictionaries. *IEEE Transactions on Signal Processing.* December 1993, pp. 3397-3415.

Mandelbrot B. B. 1982. *The fractal Geometry of Nature.* New York : Freeman, 1982.

Mandelbrot B.B. and and Van Ness J. W. 1968. Self-affine fractals and fractal dimension. *SIAM Rev., 10.* 1968, pp. 422-436.

Manocha, S., and Girolami M. A. 2007. An empirical analysis of the probabilistic k-nearest neighbor classifier. *Pattern Recognition Letters.* Oct. vol. 28, no. 13, 2007, pp. 1818-1824.

Mason S. G. , Bashashati A., Fatourechi M., Navarro K. F., Birch G. E. 2005. A Comprehensive Survey of Brain Interface Technology Designs. *Annals of Biomedical Engineering.* 2005.

Mason S. G. andBirch G. E. 2000. A brain-controlled switch for asynchronous control applications. *IEEE Transactions on Biomedical Engineering,.* 2000.

Mason S. G., Kronegg J., Huggins J., Fatourechi M., Navarro K., Birch G. E. 2006. valuating the performance of self-paced brain computer interface technology. 2006.

Matsuoka, Yoky and P. A. and OH., M. 2006. On the design of robotic hands for brain machine interface. 2006.

McFarland et al. 1997. Design and operation of an EEG-based brain-computer interface with digital signal processing technology. *Behav. Res. Meth. Inst. Comp.* 1997, vol. 29, pp. 337-345.

McFarland D. J., Anderson C. W., Muller K.-R., Schloegl A., and Krusienski D. J. 2006. BCI meeting 2005-workshop on BCI signal processing: feature extraction and translation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering.* 14 2006, pp. 135-138.

McFarland D. J., McCane L. M., David S. V., and Wolpaw J. R. 1997. Spatial filter selection for EEG-based communication. *Electroenceph. Clin. Neurophysiol.* Sept. 1997, vol. 103, no. 3, pp. 386-394.

McFarland D. J., Sarnacki W. A., Vaughan T. M. and Wolpaw J. R. 2006. Brain-computer interface (BCI) operation: signal and noise during early training sessions . *Clinical Neurophysiology.* Jan. vol. 116, no.1, 2006, pp. 56-62.

McFarland D, et al. 2000. Mu and beta rhythm topographies during motor imagery and actual movements. *Brain Topogr.* 2000, vol. 12, no. 3, pp. 177-186.

McFarland D. J., McCane, L. M. and and Wolpaw J. R. 1988. EEG-Based communication and control: short-term role of feedback. *IEEE Transactions on Neural Systems and Rehabilitation Engineering.* March 1988, vol. 6, no. 1, pp. 7-11.

McFarland D. J. and Krusienski D. ,W. S. and Wolpaw J. R. 2008. Emulation of computer mouse control with a noninvasive brain computer interface. 2008.

McFarland Dennis J., Sarnacki William A., and Wolpaw Jonathan R. 2003. Brain computer interface (BCI) operation: optimizing information transfer rates. *Biological Psychology 63.* 2003, pp. 237-251.

Mehring C. and Ball T. 2003. Decoding of movement direction from electrocorticographic (ECoG) recordings in human sensorimotor cortex - a potential basis for a brain-machine interface. *Ladislav Tauc Conference in Neurobiology.* 2003, 49.

Meinicke P, Kaper M., Hoppe F., Heumann M., Ritter H. 2003. Improving transfer rates in brain computer interfacing: a case study. 2003, pp. 1131-1138.

Mendel J. M. 1991. Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications. *Proceedings of the IEEE.* 1991, vol. 79, no. 3, pp. 278-305.

Meyer D., Leisch F., andHornik K. 2003. The support vector machine under test. *Neurocomputing.* Sept. vol. 55, no. 1-2, 2003, pp. 169-186.

Mika S., Smola A. J. and Schölkopf B. 2001. An improved training algorithm for kernel fisher discriminants. *In Proceedings AISTATS, Morgan Kaufmann.* 2001.

Mika Sebastian, Rätsch Gunnar and Müller Klaus-Robert. 2001. A Mathematical Programming Approach to the Kernel Fisher Algorithm. 2001.

Millan J. R., Renkens F., Mourino J., and Gerstner W. 2004. Noninvasive brain-actuated control of a mobile robot by human EEG. *IEEE Transactions on Biomedical Engineering.* 2004, vol. 51, no. 6, pp. 1026-33.

Molina G.G., Tsoneva T. and Nijholt A. 2009. Emotional brain-computer interfaces. *3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, ACII 2009.* Sep. 2009, pp. 1-9.

Mondada F., Bonani M., Raemy X., Pugh J., Cianci C., Klaptocz A., Magnenat S., Zufferey J. C., Floreano D., and Martinoli A. 2009. The e-puck, a Robot Designed for Education in Engineering. *Proceedings of the 9th Conference on Autonomous Robot Systems and Competitions.* 2009, vol. 1, pp. 59-65.

Moore-Jackson M. M., Mason S. G., and Birch G. E. 2005. Analyzing Trends in Brain Interface Technology: A Method to Compare Studies. *Annals of Biomedical Engineering.* 2005.

Müller K. R., Anderson C. W., and Birch G. E. 2003. Linear and nonlinear methods for brain-computer interfaces. *IEEE Trans. on Rehabilitation Engineering.* 2003, pp. 165-169.

Münßinger, J. I., et al. 2010. Brain Painting: first evaluation of a new brain-computer interface application with ALS-patients and healthy volunteers. *Frontiers in Neuroscience.* 2010, vol.4, no. 182, pp. 1-11.

Naeem, M, et al. 2006. Separability of four-class motor imagery data using independent component analysis. *Journal of Neur Eng.* March 2006, vol 3, pp. 208-216.

NaitAli A. 2009. *Advanced signal processing.* Berlin Heidelberg : Springer-Verlag, 2009.

Ng S.C., and Raveendran P. 2008. Removal of EOG Artifacts Using ICA Regression Method. 2008, pp. 226-229.

Nicolas-Alonso, Luis Fernando and Gomez-Gil, Jaime. 2012. Brain Computer Interfaces, a Review. *Sensors.* 2012, vol. 12, pp. 1211-1279.

Nicolelis M. A. L. 2001. Actions from thoughts. *Nature 2001*, vol. 409, pp. 403-407 .

Niedermeyer E., and Lopes da Silva F. H. 1999. *Electroencephalography: Basic principles, clinical applications and related fields.* Baltimore, MD, : Williams and Wilkins, 1999.

Nijboer. 2008. A P300-based brain-computer interface for people with amyotrophic lateral sclerosis. *Clin Neurophysiol.* 119(8) 2008, pp. 1909-1916.

Nunez P. L., Srinivasan R. 1981. *Electric fields of the brain: The neurophysics of EEG.* s.l. : Oxford University Press, 1981.

Ochoa, C.J., Polich, J. 2000. P300 and blink instructions. *Clinical Neurophysiology 2000.* vol. 111, pp. 93-98.

O'Doherty Joseph E. et al., 2011. Active tactile exploration using a brain–machine–brain interface. *Nature.* Nov. 2011, 479, pp. 228-231.

O'Doherty Joseph E. , Lebedev Mikhail A., Ifft Peter J., Zhuang Katie Z., Shokur Solaiman, Bleuler Hannes, Miguel Nicolelis A. L. 2011. Active tactile exploration using a brain–machine–brain interface. *Nature letter.* 2011, 10489.

Obermaier, Bernhard , Neuper Christa, Guger Christoph, and Pfurtscheller Gert. 2001. Information Transfer Rate in a Five-Classes Brain–Computer Interface. *IEEE Transactions on neural systems and rehabilitation.* September 2001, vol. 9, no. 3, pp. 283-287.

O'Doherty J. E., Lebdev M. A., Hanson T. L., Fitzsimmons N. A., Nicolelis M. A. L. 2009. A brain-machine interface instructed by direct intracortical microstimulation. *Frontiers in Integrative Neuroscience.* 2009, vol. 3, no. 20.

Onton J., Makeig S. 2006. Information-based modeling of event-related brain dynamics. *Prog Brain Res.* 2006, pp. 99-134.

Opitz D. and Maclin R. 1999. Popular ensemble methods: An empirical study. *Journal of Artificial Intelligence Research 11.* 1999, pp. 169-198.

Oskoei, M. A., Gan, J. Q., and Huosheng Hu 2009. "Adaptive schemes applied to online SVM for BCI data classification," *Engineering in Medicine and Biology Society, IEEE EMBS, 2009,* pp: 2600-2603.

Oung, S., Pohl, B. M. and U. G. Hofmann. 2012. Preliminary Design of a Tendon-Based Anthropomorphic Robotic Hand. BMT, 2012.

Palmini, Andre. 2006. The concept of the epileptogenic zone: a modern look at Penfield and Jasper's views on the role of interictal spikes. *Epileptic Disorders.* 2006, 8, pp. 10-15.

Parasuraman Raja, Rizzo Matthew. 2007. *Neuroergodics: The brain at work.* s.l. : Oxford university press, 2007.

Park Hyung-Min and Oh Sang-Hoon , and Lee Soo-Young. 2002. On adaptive noise cancelling based on independent component analysis. *Electronics Letters.* 2002, vol. 38, no. 15, pp. 832-833 .

Pedregosa F. et al., .Scikit-learn: Machine Learning in Python, JMLR 12, pp. 2825-2830, 2011

Peng H, Long F., and Din Ch. 2005. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 2005, vol. 27, no. 8, pp. 1226-1238.

Pereda E., et al. 1998. Non-linear behavior of human EEG: fractal exponent versus correlation dimension in awake and sleep stages. *Neuroscience Letter.* 1998, 250, pp. 91-94.

Petropulu A. P. 2000. Higher-Order Spectral Analysis. [book auth.] Boca Raton Ed. Joseph D. Bronzino. *The Biomedical Engineering Handbook Second Edition.* s.l. : CRC Press LLC, 2000.

Pfurtscheler G. 2003. "Thought" – Control of Functional Electrical Stimulation to Restore Hand Grasp in a Patient with Tetraplegia. *Neuroscience Letters.* 2003, vol. 351, pp. 33-36.

Pfurtscheller G. Neuper C., Flotzinger D., and Pregenzer M. 1997. EEG-based discrimination between imagination of right and left hand movement. *Electroencephalography and Clinical Neurophysiology.* 1997, vol. 103, pp. 642-651.

Pfurtscheller G, and Neuper C. 2001. Motor Imagery and Direct Brain– Computer Communication. *IEEE Proc.,, pp. , Jul. 2001.* Jul. vol. 89, no.7, 2001, pp. 1123-1134.

Pfurtscheller G. , Flotzinger Doris, Kalcher Joachim. 1993. Brain-Computer Interface—a new communication device for handicapped persons. *Journal of Microcomputer Applications.* 1993, pp. 293-299.

Pfurtscheller G, et al. 2000. Brain oscillations control hand orthosis in a tetraplegic. *Neuroscience Letters.* 2000, vol. 292, no. 3, pp. 211-214.

Pfurtscheller G, et al. 2010. The hybrid BCI Front. *Frontiers Neuroscience.* 2010, vol. 4, no. 42.

Pfurtscheller G., and Lopes da Silva, F. H. 1999. Event related EEG/EMG synchronization and desynchronizations: Basic principals. *Invented review in clinical neurophysiology.* 1999, 110, pp. 1842-1857.

Pfurtscheller G. 1992. Event-related synchronization (ERS): an electrophysiological correlate of cortical areas at rest. *Electroenceph. clin. Neurophysiol.* 1992, vol. 83, no.4, pp. 62-69.

Pfurtscheller G.,and Neuper C. 2006. Future prospects of ERD/ERS in the context of brain–computer interface (BCI) developments. *Progress in Brain Research.* 2006, vol 159, pp. 433-437.

Pham, T. T. H., Croft, R. J., Cadusch, P. J., and Barry, R. J. 2011. A test of four EOG correction methods using an improved validation technique. International Journal of Psychophysiology 79:203-210

Phothisonothai M. and and Nakagawa M. 2007. Fractal-Based EEG Data Analysis of Body Parts Movement Imagery Tasks. *J. Physiol. Science.* August 2007, vol. 57, no. 4, pp. 217-226.

Pierce J. R. 1980. An Introduction to Information Theory. *Dover, New York.* 1980, pp. 145-165.

Pistohl T., Ball T., Schulze-Bonhage A., Aertsen A., Mehring C. 2008. Prediction of arm movement trajectories from ECoG-recordings in humans. *Journal of Neuroscience Methods.* 2008, vol. 167, no. 1, pp. 105-114.

Plöchl M, Ossandón JP, König P. 2012. Combining EEG and eye tracking: identification, characterization, and correction of eye movement artifacts in electroencephalographic data. Front Hum Neuroscience. 6:278, 2012.

Poree F., et al. 2006. Blind source separation for ambulatory sleep recording. *IEEE Trans. InformationTechnol. Biomed.* 2006, vol. 10 no. 2, pp. 293-301.

Principe Jose C. 2010. *Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives.* New York : Springer, 2010.

Qian S. and Chen D. 1996. *Joint Time-Frequency Analysis: Methods and Applications.* s.l. : Prentice-Hall, 1996.

Rabin et al. 2011. Integrative rehabilitation of eldery stroke survivors. 2011.

Rakotomamonjy A., Guigue V., Mallet G., and Alvarado V. 2005. Ensemble of svms for improving brain computer interface p300 speller performances. *International Conference on Artificial Neural Networks.* 2005.

Rakotomamonjy A. 2003. Variable selection using SVM-based criteria. *Journal of Machine Learning Research.* 2003, vol. 3, pp. 1357-1370.

Ramoser H., Muller-Gerking J. and Pfurtscheller G. 2000. Optimal Spatial Filtering of Single Trial EEG during Imagined Hand Movement. *IEEE Trans. Rehab. Eng.* Dec. vol. 8, no.4, 2000, pp. 441-446.

Rapp P. E., et al. 1993. Filtered noise can mimic low-dimensional chaotic attractors. *Physics Rev. E; 47.* 1993, pp. 2289-2297.

Raudys S. J. and Jain, A. K. 1991. Small sample size effects in statistical pattern recognition: Recommendations for practitioners. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 1991, vol. 13. no. 3, pp. 252-264.

Renard Y., et al. 2010. OpenViBE: an open-source software platform to design, test and use brain–computer interfaces in real and virtual environments. *Presence Teleoperators Virtual Environment.* 2010.

Renyi A. 1960. On measures of entropy and information. 1960, p. 547.

Rodgers J. L., Nicewander W. A. 1988. Thirteen ways to look at the correlation coefficient. *The American Statistician.* February 1988, vol. 42, no. 1, pp. 59-66.

Romero, S, Mañanas M. A. and Barbanoj M. J. 2008. A comparative study of automatic techniques for ocular artifact reduction in spontaneous EEG signals based on clinical target variables: A simulation case. *Comput. Biol. Med.* 2008, pp. 348 -360.

Rosenbaum M., Raz D. Denial. locus of control and depression among physically disabled and nondisabled men. *Psychodynamics and Psychopathology.* vol. 33, no. 3, pp. 672-676.

Rothwell J. C. 2001. First studies of the organization of the human motor cortex. [book auth.] Eds. M. L. Latash and V. M. Zatisiorsky. *Classics in Movement Science.* Champaign, IL : Human Kinetics, 2001, pp. 273-288.

Sandlund M., McDonough S., Häger-Ross C. 2009. Interactive computer play in rehabilitation of children with sensorimotor disorders: a systematic review. *Dev Med Child Neurol.* Epub, March 2009, vol. 51, no. 3, pp. 173-179.

Sanei Saeid and Chambers J. A. 2007. *EEG signal processing.* s.l. : John Wiley & Sons Ltd., 2007.

Scarabino T., Salvolini U., Di Salle F., Duver H. 2003. *Atlas of morphology and functional anatomy of the brain.* s.l. : Springer, 2003.

Schalk G., Kubanek, J., Miller, K.J., Anderson, N.R., Leuthardt, E.C., Ojemann, J.G., Limbrick, D., Moran, D.W., Gerhardt, L.A., and Wolpaw J. R. 2007. Decoding Two-Dimensional Movement Trajectories Using Electrocortico-graphic Signals in Humans. *Journal of Neural Enginering.* 2007, 4, pp. 264-275.

Schalk G., Leuthardt E. C. 2011. Brain-computer interfaces using electrocorticographic signals. *IEEE Rev Biomed Eng.* 2011, vol. 4, pp. 140-154.

Schalk G., Mellinger J. 2010. *A Practical Guide to Brain–Computer Interfacing with BCI2000.* 1st edition, Berlin : Springer, 2010.

Schalk G. 2010. Can electrocorticography (ECoG) support robust and powerful brain–computer interfaces? *Frontiers in Neuroengineering.* June 2010, vol. 3, no. 9.

Scherer R., Müller G. R., Neuper C., Graimann B., and Pfurtscheller G. 2004. An asynchronously controlled EEG-based virtual keyboard: improvement of the spelling rate. *IEEE Trans. Biomed. Eng.* Jun. 2004, pp. 979-984.

Schlogl A. et al. 2002. Estimating the mutual information of an EEG-based brain-computer interface. *Biomed Tech.* 2002, pp. 3-8.

Schlogl A, Lee F., Bischof H., and Pfurtscheller G. 2005. Characterization of four-class motor imagery EEG data for the BCI-competition. *Journal of Neural Engineering.* 2005.

Schoelg A, Neuper C., Pfurtscheller G. 2002. Estimating the Mutual Information of an EEG-based Brain-Computer Interface. *BioMed technique.* 2002, vol. 47, pp. 3-8.

Schopflocher T. P., and Sullivan P. J., 2005. The Relationship between Skewness and Kurtosis of a diffusing Scalar, Boundary-Layer Meteorol. vol 115, no. 3 pp. 341-358.

Schouenborg Jens, Garwicz Martin, Danielsen Nils. 2011. *Brain Machine Interfaces: Implications for Science, Clinical Practice and society,* Elsevier, 2011.

Schuler T., Brutsch K., Muller R., Hvan Hedel U. J., Meyer-Heim A. 2011. Virtual realities as motivational tools for robotic assisted gait training in children: A surface electromyography study. *Neuro Rehabilitation.* 2011, vol. 28, no. 4, pp. 401-411.

Schwartz A. B., Taylor, D. M., and Helms Tillery, S. I. 2001. Extraction Algorithms for cortical control of arm prosthetics. *Current Opinions in Neurobiology.* 2001, 11, pp. 701-707.

Schwartz A.B. 2007. Cortical Neural Prosthetics. *Annual Review of Neuroscience.* 2007, 27, pp. 487-507.

Semlitsch H. V., Anderer P. 1986. A solution for reliable and valid reduction of Ocular Artefacts, applied to the P300 ERP. *Psychophisiolohgy.* 1986, pp. 695-703.

Sepulveda F., Meckes M. and Conway B. A. 2004. Cluster separation Index suggests usefulness of EEG non-motor channels in detecting wrist movement direction intention. 2004, pp. 943-947.

Sethi N., Sethi P., Torgovnick J., Arsura E. 2007. Physiological and non-physiological EEG artifacts. *The Internet Journal of Neuromonitoring.* 2007.

Shannon and Weaver. 1963. The mathematical theory of communication. *UrbanaIL: Univ of Ill Press.* 1963.

Shannon C. E. 1948. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* July and October 1948, vol. 27, pp. 379-423, 623-656.

Shawe-Taylor J., and Cristanini N. 2000. *An introduction to Support Vector Machine.* Cambridge : Cambridge University Press, 2000.

Shibasaki H., and Hallett M. 2006. What is the Bereitschaftspotential? Clinical Neurophysiology. vol. 117. pp: 2341-2356.

Shoaie Shirehjini Z. Bagheri Shouraki S. Esmailee M. 2009. Variant combination of Multiple Classifiers Methods for Classification the EEG signals in Brain-Computer Interface. s.l. : Springer, 2009, Vol. 9, 59, pp. 477-484.

Shoker L., Sanei S. and Chambers J. 2005. Artifact removal from electroencephalograms using a hybrid BSS-SVM algorithm. *IEEE Signal Process. Lett.* October 2005, vol.12, no. 10.

Shoker L., Sanei S. and Sumich A. 2005. Distinguishing between left and right finger movement from EEG using SVM. *Proceedings of the IEEE EMBS.* September 2005, pp. 5420-5423.

Silbert P. L., Radhakrishnan K., Johnson J., and Class D. W. 1995. The significance of the phi rhythm. *Electroenceph. Clin. Neurophysiol.* 1995, vol. 95, pp. 71-76.

Society of Neuroscience. 2008. *Brain facts: A primer on the brain and nervous system.* 2008.

Soria Frisch Aureli. 2012. A critical review on the usage of ensembles for BCI. [book auth.] Brendan Z. Allison. *Towards Practical Brain-Computer Interfaces: Bridging the Gap from Research to Real-World Applications.* Berlin,New York : Springer, 2012, pp. 41-65.

Spüler 2012. Adaptive SVM-based classification increases performance of a MEG-based BCI, 2012

Steinwart I., Christmann A. 2008. *Support vector machines.* Springer 2008.

Stern, J. A., Walrath, L.C., and Goldstein, R. 1984. The endogenous eyeblink. *Psychophysiology* vol. no. 21, pp. 22-33

Stern J. M. and Engel J. 2004. *An Atlas of EEG Patterns.* Philadelphia: Lippincott : Williams & Wilkins, 2004.

Sun Shiliang, Zhang Changshui and Zhang Dan. 2007. An experimental evaluation of ensemble methods for EEG signal classification, . *Elsevier, Pattern recognition letters.* 2007.

Syed A. 2012. Flex Sensor Based Robotic Arm Controller Using Micro Controller. *Journal of Software Engineering and Applications.* 2012, vol. 5, pp. 364-366.

Tangermann Michael, et al. 2011. Data Driven Neuroergonomic Optimization of BCI Stimuli. *Pattern Analysis Statistical Modelling & Computational Learning,* . 2011.

Tatum W. O., Dworetzky B., Schomer D. 2011. Artifact and Recording Concepts in EEG. *Journal Clin Neurophysiol.* 2011, pp. 252-263.

Taylor D. M., Helms Tillery S. I., and Schwartz A. B. 2002. Direct cortical control of 3D neuroprosthetic devices. *Science.* 2002, vol. 296, pp. 1829-1832.

Theiler J. and Rapp P. 1996. Re-examination of the evidence for low-dimensional, nonlinear structure in the human electroencephalogram. *Electroenceph. clin. Neurophysiologty.* 1996, vol. 98, pp. 213-222.

Thomas D., Coates Jr. 2008. *Neural Interfacing: Forging the Human-Machine Connection.* s.l. : Morgan & Claypool Publishers series, 2008.

Tipping M. 2001. Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning Research.* Sept. vol. 1, 2001, pp. 211-244.

Tomasello M. 2008. *Origins of Human Communication.* s.l. : Massachusetts Institute of Technology, 2008.

Torrance Steve. 1984. *The mind and the machine, philosophical aspects of artificial intelligence.* s.l. : Elis Horwood series Artificial Intelligence, 1984.

Torrance S., Horwood E. 1984. *The mind and the machine, philosophical aspects of artificial intelligence.* s.l. : series Artificial Intelligence, 1984.

Touw W. G. and Bayjanov J. R. 2012. Data mining in the Life Sciences with Random Forest: a walk in the park or lost in the jungle. *Brief. Bioinformatics.* 2012.

Townsend G., Graimann B. and Pfurtscheller G. 2006. A comparison of common spatial patterns with complex band power features in a four-class BCI experiment. *IEEE Transactions on Biomedical Engineering.* 2006, vol. 53, no. 4, pp. 642-651.

Townsend G., et al. 2010. A novel P300-based brain-computer interface stimulus presentation paradigm: moving beyond rows and columns. *Clinical Neurophysiolog*y. 2010, vol. 121, no. 7, pp. 1109-1120.

Turing, Alan M. 1950. Computing machinery and intelligence. *Mind.* 1950, pp. 433-460.

Vallabhaneni A., Wang T., and He B. 2005. Brain-Computer interface. *Neural engineering, bioelectricengineering,Springer.* 2005, p. 85:121.

Vansteensel M. J., et al. 2010. Brain-computer interfacing based on cognitive control. *Annals of Neurology.* vol. 67, no, 6.

Vapnik, V. 1999. *The nature of statistical learning theory* . New York : springer, 1999.

Vaughan T. M., Heetderks W. J., Trejo L. J., Rymer W. Z., Weinrich M., Moore M. M., Kubler A., Dobkin B. H., Birbaumer N., Donchin E., Wolpaw E. W. and Wolpaw J. R. 2003. Brain-computer interface technology: a review of the Second International Meeting", . *IEEE Trans. Neural Syst. Rehabil. Eng.* 11 2, 2003, pp. 94-109.

Velliste M., Perel S., Spalding M. C., Witford A. S., and Schwartz A. B. 2008. Cortical control of a prosthetic arm for self-feeding. *Nature.* 2008, vol. 453, no. 7198, pp. 1093-1101.

Verleger R. 1991. The instruction to refrain from blinking affects auditory P3 and N1 amplitudes. *Electroencephalogr. Clin. Neurophysiol.* 1991, pp. 240-251.

Vidal J. 1973. Toward Direct Brain-Computer Communication. *Annual Review of Biophysics and Bioengineering.* 1973, pp. 157-180.

Vidaurre C., Blankertz B. 2010. Towards a Cure for BCI Illiteracy. *Open Access, Brain Topogr.* 2010, pp. 194-198.

Vidaurre C., et al. 2005. Adaptive online classification for EEG-based brain computer interfaces with AAR parameters and band power estimates. *Biomedizinische Technik* . 2005, vol. 50, no. 11, pp. 350-354.

Vidaurre C., et al. 2011. Co-adaptive calibration to improve BCI efficiency. *Journal of neural engineering 8.* 2011, no.2, Art. 025009.

Vuckovic A., Sepulveda F. 2006. EEG-Based Eight-Class, Single Trial Classification of Imaginary Wrist Movements. *IEEE.* 2006.

Vuckovic A., and Sepulveda F. 2008. Quantification and visualisation of differences between two motor tasks based on energy density maps for brain–computer interface applications. *Clinical Neurophysiology.* 2008, vol. 119, pp. 446-458.

Walter W. G., Cooper R., Aldridge V. J. and McCallum W. C. 1964. Contingent negative variation: An electric sign of sensorimotor association and expectancy in the human brain, *Nature*.

Waszak F., Hommel B., Allport A. 2005. Interaction of task readiness and automatic retrieval in task switching: negative priming and competitor priming, Mem Cognit. vol. 33, no. 4, pp: 595-610

Waterink W. and Boxtel A. van. 1994. Facial and jaw-elevator EMG activity in relation to changes in performance level during a sustained information processing task. *Biol. Psychol.* Jul. vol. 37, no.3, 1994, pp. 183-198.

Watson, Andrew. 1997. Why can't a computer be more like a brain? *Science 26.* September 1997, vol. 277 no. 5334, pp. 1934-1936.

Weifeng Liu, Pokharel P. P., Principe J. C. 2006. Correntropy: A Localized Similarity Measure. *Neural Networks, IJCNN '06. International Joint Conference on Neural Networks.* 2006, pp. 4919- 4924.

Weinberg H., Walter W. G. , and Crow J. H. 1970. Intra cerebral events in Humans related to imaginary stimuli, Electroencephalography and Clinical Neurophysiology, vol. 29 pp. 1-9.

Wessburg J., Stambaugh, C. R., Kralik, J. D., Beck, P. D., Srinivasan, M. A., and Nicholelis, M. A. L. 2000. Real-time predition of hand trajectory by ensembles of cortical neurons in primates. *Nature.* 2000, 408, pp. 361-365.

White J. R., Levy T., Bishop W., and Beaty J. D. 2010. Realtime decision fusion for multimodel neural prosthetic devices. *PLoS ONE.* 2010.

Wickelgren, Ingrid. 2003. Tapping the mind. *Science 24.* 2003, vol. 299 no. 5606, pp. 496-499.

Williams RW, Herrup K. 1988. The control of neuron number. *Annual Review of Neuroscience.* 1988, vol. 11, no. 42353.

Winkler I., Haufe S., and Tangermann M. 2011. Automatic Classification of Artifactual ICA-Components for Artifact Removal in EEG Signals. *Behavioral and Brain Functions.* 2011.

Woestenburg J. C., Verbaten M. N. and Slangen J. L. 1983. The removal of the eye-movement artifact from the EEG by regression analysis in the frequency domain. *Biol. Psychol.* 1983, pp. 127-147.

Wolpaw J. R., McFarland, D. and Pfurtscheller G. 1998. EEG-based communication: Improved accuracy by response verification. *IEEE Transactions on Rehabilitation Engineering.* 1998, vol. 6, no. 3, pp. 326-333.

Wolpaw J. R. , Birbaumer N., Heetderks W. J., McFarland D. J., Hunter Peckham P., Schalk G., Donchin E., Quatrano L. A., Robinson C. J., and Vaughan T. M. 2000. Brain-Computer Interface Technology: A Review of the First International Meeting. *IEEE Transactions on Rehabilitation Engineering.* 2000, vol. 8, pp. 164-173.

Wolpaw J. R., McFarland, D. J., and Vaughan, T. M. 2000. Brain–computer interface research at the Wadsworth Centre. *IEEE Trans. Neural System Rehabil. Engng.* 2000, vol. 8, pp. 222-226.

Wolpaw J. R., Birbaumer N., McFarland D. J., Pfurtscheller G., Vaughan T. M. 2002. Brain-computer interfaces for communication and control. *Clinical Neurophysiolgy.* 113 2002, pp. 767-792.

Wolpaw J. R., McFarland, D. J. 2004. Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *Proc. Natl Acad. Sci.* 2004, vol. 101, pp. 17849-17854.

Wolpaw J. R., Wolpaw E. W. 2012. *Brain-Computer Interfaces: Principles and Practice.* New York : Oxford University Press, 2012.

Yamamoto Y., and Hughson R.L. 1991. Coarse-graining spectral analysis: new method for studying heart rate variability, *The American Physiological Society* 1991, vol. 71, no. 1143.

Yamamoto Y. and and Hughson, R. L. 1993. Extracting fractal components from time series. *Physica.* 1993, vol. 68, pp. 250-264.

Yang Y., Yan R., Hauptmann A. 2007. Cross-Domain video concept detection using Adaptive SVMs. *ACM Multimedia*.

Zarzoso V., Phlypo R., Meste O., Comon P. 2009. Signal extraction in multisensory biomedical recordings. *in Advances in Biomedical Engineering.* Oxford, UK : Elsevier BV, , 2009, pp. 95-143.

Zecca M., Micera S., Carrozza M. C., Dario, P. 2002. Control of multifunctional prosthetic hands by processing the electromyographic signal. *Crit. Rev. Biomed. Eng.* 2002, pp. 459-485.

Zhang J. W., Zheng, C. X. and Xie, A. 2000. Bispectrum Analysis of focal Ischemic Cerebral EEG Signal Using Third-Order Recursion Method,. *IEEE Trans. Biomed. Eng.* 2000, vol. 47, no. 3.

Zhang Yagang. 2010. *Application of Machine Learning.* InTech 2010.

Zhao Chunyu, Tianshuang Qiu. 2011. An Automatic Ocular Artifacts Removal Method Based On Wavelet-Enhanced Canonical Correlation Analysis. *IEEE EMBS.* 2011.

Zhou-Yan Feng. 2003. Analysis of Rat EEG during Sleep Using Wavelet Transform. *ACTA BIOCHIMICA et BIOPHYSICA SINICA,.* 2003, vol. 35, no. 8, pp. 741-746.

Zweig M. H. and Campbell G. 1993. Receiver-Operating Characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry.* 1993, vol. 39, no. 4, pp. 561-577.

List of Abbreviations

ADC	Analogue-to-digital converter
AE	Approximate entropy
AEP	Auditory evoked potential
AF	Adaptive filter
Ag–AgCl	Silver–silver chloride
AIC	Akaike information criterion
ALS	Amyotrophic lateral sclerosis
ANOVA	Analysis of variance
AP	Action potential
AR	Autoregressive model
BCI	Brain computer interface
BMI	Brain machine interface
BOLD	Blood oxygenation level dependent
BSS	Blind source separation
CNS	Central nervous system
CSP	Common spatial patterns
ECG	Electrocardiogram, electrocardiography
ECoG	Electrocorticogram, electrocorticography
EEG	Electroencephalogram, electroencephalography
EMG	Electromyogram, electromyography
EOG	Electrooculogram
EP	Evoked potential
ERD	Event-related desynchronization
ERP	Event-related potential
ERS	Event-related synchronization
fMRI	Functional magnetic resonance imaging
fNIR	Functional near infrared
FP	False positive
FN	False negative
HCI	Human computer interface
ICA	Independent component analysis
ITR	Information transfer rate
KL	Kullback–Laibler
LDA	Linear discriminant analysis

LFP	Local field potential
LIS	Locked-in state
MCC	Matthew correlation coefficient
MEG	Magnetoencephalogram, magentoencephalography
MEP	Movement-evoked potential
MI	Motor imagery
MRI	Magnetic resonance imaging
NIRS	Near-infrared spectroscopy
PCA	Principal component analysis
PET	Positron emission tomography
SCP	Slow cortical potential
SMA	Supplementary motor area
SMR	Sensorimotor rhythm
SNR	Signal-to-noise ratio
SSVEP	Steady state visual evoked potential
VEP	Visual evoked potential
VR	Virtual reality

List of Symbols

f(x)	Function of *x*
g'	First derivative of g
Hz	Hertz; cycles per second
I(s)	Entropy of s
Im(.)	Imaginary part
i, j	$\sqrt{-1}$
J_n	Cost function
mV	millivolt
$x \in Rd$	x belongs to the d-dimensional space of real values
Z	Z-transform
Z^{-1}	Inverse Z-transform
α	Alpha brain rhythm
α	Penalty term
α_j and β_j	Nonlinear model coefficients
α_k	Damping factor
α	Forward rate function
β	Beta brain rhythm
β	Backward rate function
γ	Gamma brain rhythm
γ	Learning rate
δ	Delta brain rhythm
ζ	Learning rate
η	Performance index
θ	Theta brain rhythm
θ_j	Initial phase in radians
κ	Kappa brain activity
λ	Lambda brain activity