

Translational Algorithms: the heart of a Brain Computer Interface

Harsimrat Singh¹, Ian Daly²

¹ Department of Medical Physics and Bioengineering, University College London, UK

² Brain Embodiment Lab, University of Reading, UK

{harsimrat.singh@ucl.ac.uk; i.daly@reading.ac.uk}

Abstract

Brain computer Interface (BCI) development encapsulates three basic processes: data acquisition, data processing, and device control. Since the start of the millennium the BCI development cycle has undergone a metamorphosis. This is mainly due to the increased popularity of BCI applications in both commercial and research circles. One of the focuses of BCI research is to bridge the gap between laboratory research and commercial applications using this technology.

A vast variety of new approaches are being employed for BCI development ranging from novel paradigms, such as simultaneous acquisitions, through to asynchronous BCI control. The strategic usage of computational techniques, comprising the heart of the BCI system, underwrites this vast range of approaches. This chapter discusses these computational strategies and translational techniques including dimensionality reduction, feature extraction, feature selection, and classification techniques.

1. INTRODUCTION

Brain-computer interfacing (BCI) is a highly challenging multidisciplinary area of research. Since the start of the millennium, research groups exploring this area have made increasingly impressive progress. The core BCI research has opened up hundreds of avenues for its applications. There have been interesting revelations in the field of rehabilitation ((Birbaumer, 2006); (Ang et al., 2010)), gaming ((Nijholt and Desney, 2009)), composing music ((Miranda et al., 2011)), and other biophysics applications

based on BCI. However, there are still a lot of unanswered questions facing the BCI community, which impedes the launch of mainstream commercial BCI applications. The first five international meetings of BCI community (Wolpaw et al., 2000), (Vaughan et al., 2003), (Vaughan and Wolpaw, 2006) (Vaughan and Wolpaw, 2011), and (Huggins et al., 2014)) have evolved sufficient consensus on the BCI terminology, signals used and the computational techniques but questions relating to user variability, session variability and optimal training remain a major research issues. The inherent complexity and enormity of the neural data from measurement techniques such as (electroencephalogram (EEG), Electrocorticogram (ECoG), functional near infra-red spectroscopy (fNIRS), and functional Magnetic resonance Imaging (fMRI)) warrants sophistication in computational strategies for meaningful interpretation.

Research areas of BCI evolved from work of Hans Berger, a German Psychiatrist who first recorded EEG in 1929. Since then EEG has served as a standard Clinical diagnostic tool for a range of neurological complexities (Tudor et al., 2005). Among several definitions, an EEG is expressed as sustained fluctuations of electric potential, recorded from the human scalp which can be used to decipher corresponding variations in the cortex of the brain. Our ability to feel, think and act can be attributed to these variations of electrical activity. Farwell & Donchin, 1988 first demonstrated that the ability of EEG response to change to an externally accentuated event can be developed as a non-muscular communication channel for sending messages and commands to the outside world for control purposes; this is popularly known as the brain computer interface. Subsequently, clinical BCI applications such as a speller have underlined the popularity of speller based event related potential (ERP) as the basis of EEG based BCI (Sellers and Donchin, 2006).

The principles underlying of this opening up of a communication channel between the brain and the computer are based on the classification of the changes in the EEG spectra which relate to, for example, the imagination of movements (Pfurtscheller and Lopes da Silva, 1999). The heart of the BCI is the translational algorithm which converts the

Pre-Print version

electrophysiological measurements from the user into output that controls external devices (Wolpaw et al., 2003). We use the term EEG here to typify the signal which is used as the vehicle for development of the BCI; EEG being the most popular and easy to record – other brain-related signals include electrocorticogram (ECoG), magnetoencephalogram (MEG), and functional magnetic resonance (fMRI) images. These will be discussed in the section - ‘Sources of Information for a BCI’. Ideally an interface may suggest a two way communication channel but the present state of the art in BCI purposes is one way communication i.e. from the brain to the computer. Information from the brain in the form of signals or images has been sufficient to realise a plethora of applications (Guger et al., 1999), (Perelmouter and Birbaumer, 2000), and (Birbaumer, 2006) for medical rehabilitation and for gaming by companies like Emotiv (Emotiv, Australia) and Neurosky (Neurosky, USA).

This chapter proceeds as follows. Section 2 gives an overview of sources of information for BCI. Section 3 considers the BCI development process. Section 4 is concerned with the types of BCI. Section 5 discusses nomenclature for feature extraction and classification. Section 6 and 7 cater to evaluation criteria, conclusion and future work.

2. SOURCES OF INFORMATION FOR A BCI

A brain computer interface is developed on the basis of the knowledge that can be extracted from various sources of information from the brain, mapped to our abilities to feel, think, and act. These sources are in the form of signals or images. Scalp EEG has been one of the most popular signals used in BCI research (Hwang et al., 2013). Over the years, the efforts of neuroscientists to investigate EEG (Donchin et al., 1984), (Handy, 2005) has proved be a real benefit to the BCI development process as it has provided more avenues to implement the mapping of our cognitive processes to the characteristics of the EEG Signal.

2.1 Electroencephalogram (EEG)

EEG is clinically defined as the mean electrical activity of the brain recorded as the summed action potentials of thousands of neurons firing together (Niedermeyer and Silva, 2005). It is composed of electrical rhythms and transient discharges which are distinguished by location, frequency, amplitude, form, periodicity, and functional properties (Schomer and Lopes de Silva, 2011). EEG activity has been classified on the basis of these attributes. The most widely accepted basis of classification of EEG activity is done using frequency segregation into prototypical bands, referred to as delta, theta, alpha, beta, gamma, and mu rhythms. Table 1 lists the various EEG rhythms, their frequency and amplitude range, the brain regions in which they are most prominent, and the events most often related to each of them.

Table 1. Classification of frequency components in Rhythmic Brain Activity.

Frequency Band	δ (Delta)	θ (Theta)	α (Alpha)	μ (Mu)	β (Beta)	γ (Gamma)
Freq Range (Hz)	0.5 – 4	4 – 8	8 – 13 dev > 0.5	Approx 10	13 – 22	22 – 30
Amp Range (μV)	50 -100	<100	<10	<50	<20	<2
Prominent Related brain region	Central region	Central region	Posterior brain	Somato-sensory Cortex	Central region	
Examples of corresponding events	Infants (Age: 2 months), Adults in Sleep	Children, Adults: Drowsiness, Sleep	Opening the eyes & focussing attention	Movement	High state of wakefulness	Sensory stimulation & attention

2.1.1 Slow cortical potentials

Slow cortical potentials (SCPs) are related to the emergence of a BCI application known as a thought translational device (TTD) (Birbaumer, 2003). SCPs are understood to be the result of shifts in the depolarization level of the upper cortical dendrites, caused by the intracortical and thalamocortical afferent inflow to neocortical layers I and II. The TTD

has been designed for completely paralyzed patients and has been tested on patients with ALS (Amyotrophic Lateral Sclerosis) (Birbaumer, 2006). However, the use of SCPs has recently declined in EEG based BCI research in favour of other features such as ERPs or sensorimotor rhythm activations (Hwang et al., 2013).

2.1.2 Event related potentials

The event - related potential (ERP) is a common title for the potential changes in the EEG that occur in response to a particular “event” or a stimulus (Luck, 2005). These changes occur at very small amplitudes. Therefore, in order to reveal them, EEG samples have to be averaged over many repetitions. This removes the “random” fluctuations of the EEG, which are not stimulus-locked. ERPs can be divided into exogenous and endogenous. Exogenous ERPs occur up to about 100ms after the stimulus onset. They depend on the properties of the physical stimulus (intensity, loudness etc.). The potentials from 100ms post-stimulus onward are called endogenous (Ebersole and Pedley, 2003). They depend largely on psychological and behavioural processes related to the event. The most commonly studied ERP is the P300

A popular example of a BCI based upon ERPs is the P300 speller, which is underwritten by the neuroscientific concept that a slow, large neural response is elicited after 300ms of a rarely occurring stimulus in a train of consecutive, continuously occurring stimuli (O’Brien, 1982). A P300 BCI speller operates by presenting users with a matrix of alphanumeric letters. Each row and column of the matrix flashes and a higher amplitude peak in the EEG (a P300 ERP) may be observed to occur 300ms after the target letter flash is presented. Interestingly it has been found that the greater the number of letters the higher number of flashes in the matrix and, subsequently, the better is the P300 response (Nam et al., 2010).

2.1.3 Sensorimotor rhythm activations

Sensorimotor rhythm activations are changes in activation levels that may be observed over the sensorimotor cortex during a range of cognitive events (Pfurtscheller and Lopes da Silva, 1999). Sensorimotor rhythm changes are referred to as event related

Pre-print version

desynchronisation (ERD) in the event of a decrease in cortical activity and event-related synchronisation (ERS) in the event of an increase in motor cortical activity. They are most often associated with movement planning and execution (Pfurtscheller and Lopes da Silva, 1999), but may also be observed during tasks such as mental arithmetic (Friedrich et al., 2012), and mental rotation (Chen et al., 2013).

Unlike the ERP the ERD/S is not phase locked to a stimulus presentation and, therefore, may not be identified via averaging of EEG amplitudes. Instead band-power is measured in frequency bands of interest (typically the alpha and beta bands) relative to a pre-stimulus baseline period. Significant decreases or increases in band power indicate the presence of an ERD/S and the cortical region at which it's observed identifies the corresponding cognitive process. For example, ERD over the left primary motor cortex hand representation area may indicate movement or planning of movement in the right hand.

Finally, ERD/S may also be observed during motor imagery. Thus, BCI users may attempt to control a BCI by imagining the feeling of moving their body (kinesthetic motor imagery) to control a BCI without actually performing overt movement (Neuper et al., 2005).

Case study: ERD/S detection

An example of the ERD/S phenomena is described using a classical 8s BCI paradigm for left and right hand motor imagery (Pfurtscheller et al., 2006). The paradigm consisted of a random repetition of cue-based trials. The subject was seated in a relaxing chair with armrests and was instructed to perform imagery movements prompted by a visual cue. Each trial started with an empty black screen; at time point $t = 2s$ a short beep was presented and a cross '+' appeared on the screen to arouse the subject's attention. Then at second 3 ($t = 3s$) an arrow appears pointing either to the left or right. Each position indicated by this arrow prompts the subject to imagine either a left hand or a right hand movement. The respective movement imagination should be performed until the cross

disappears at $t = 8s$. The next trial started after a very short resting period, during which the EEG was continuously recorded.

The time frequency approach was used to display the significant power increase or decrease in a predefined frequency band, thereby representing clear and easy visualisation of the movement-related behaviour of the induced activity averaged over several trials. In a pedagogical sense, increases or decreases of the power of the EEG related to a particular event is represented by the increase or decrease in the synchrony of the neuronal populations. Averaging over trials has been employed to deal with evoked potentials in order to improve the signal to noise ratio. This deals with any on-going base EEG activity, which may be considered to be noise as opposed to the actual potential instigated by the event or the task. The ratio of power calculated after averaging over all the trials to the power within a reference interval expressed in terms of percentage is the ERD/S strength.

ERD/S is not phase locked but time locked and specific to a frequency range. It is therefore conceivable that nearly identical ERD/S activity can be observed at different spatial locations. This will be indicative from some of the figures 1(a) and (b).

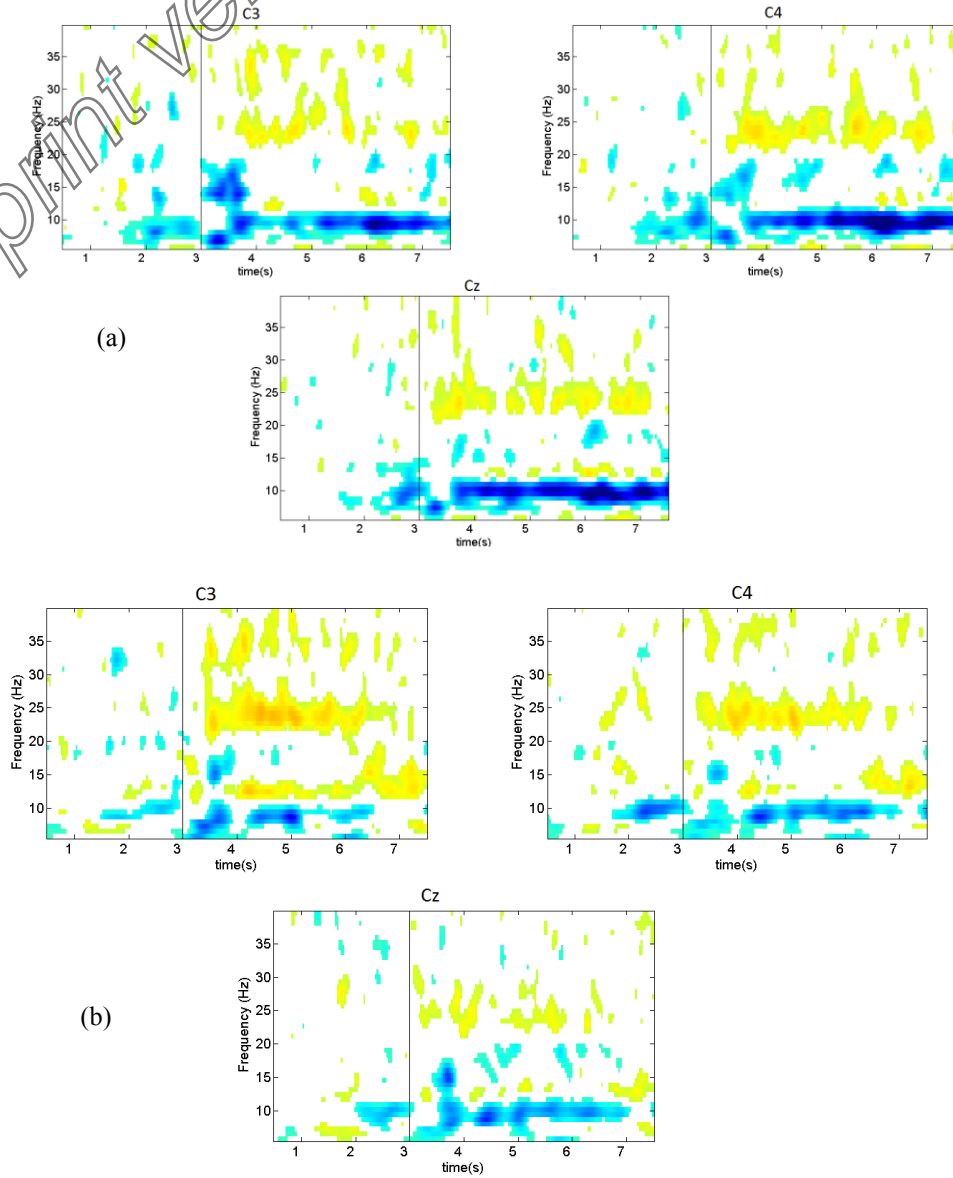


Fig 1: ERD/S maps for (a) Left hand motor imagery and (b) Right hand motor imagery. Three maps are presented for three channels C3, C4, and Cz. The vertical black line indicates the start of the motor imagery period at $t=3s$ and the maps were computed with a baseline period of 0.5-1.5s.

The timeline for each trial is shown in figure 1. The vertical black line shows the presentation of the cue at time = 3s. It is clearly observed that the activity in all the channels before the presentation of the cue is statistically significantly (bootstrap test) (Graimann et al., 2002) less when compared to the activity after the cue ($p < 0.05$). So the activity before the cue acts as a kind of reference to ascertain the fact that there is a clear increase or decrease in the 'indicative power levels' for a particular frequency band after the presentation of the cue. This is also seen in 1(b) for right hand motor imagery. Visual inspection of figures 1 (a) and (b) provides clear demarcation in the level of power activity in particular frequency bands: 8-13 Hz (alpha) and 20-30 Hz (upper beta), for the two respective tasks. This information is used for the manual selection of the frequency bands to be used for feature extraction. These bands can be employed for computing band power features.

2.2 Functional Magnetic Resonance Imaging (fMRI) and Magnetoencephalography (MEG)

The relative suitability of each of the brain sources for BCI purposes may depend on the specific application(s) being considered i.e. the relative suitability of each source is relative to the anticipated outcome of the interfacing process. The metabolic consequences of neural activity is observed as the changes in blood flow and metabolism. Imaging techniques like functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) help us to visualise these changes in the form of images and may be used in BCI applications, for example (Sitaram et al., 2007).

Associated magnetic fields produced by the neuronal activity can be detected as the magnetoencephalogram (MEG). MEG is more susceptible to noise and is not portable compared to EEG but has a far better spatial resolution ($< 1\text{cm}$) and depth sensitivity ($\sim 4\text{cm}$), while also having a similar time resolution. MEG has been used for BCI control. For example, in (Lal et al., 2005; Mellinger et al., 2007) a MEG based BCI is proposed and discussed.

2.3 Intracranial recordings ECoG (electrocorticogram)

The Electrocorticogram (ECoG) is recorded by placing invasive electrodes under the skull on the surface of the cortex. A surgical procedure is required to place these intracranial electrodes. Experience suggests that the quality of the signal is better for ECoG than the scalp EEG and is less contaminated with artifacts, but that the ease of recording of the scalp EEG is superior to ECoG for BCI purposes (Schalk and Leuthardt, 2011). It has also been observed in some studies that the performance of the classifier is dependent on the user population (Hill et al., 2006).

2.4 Functional near infra-red spectroscopy (fNIRS)

Diffuse Optical imaging (DOI) is another method to measure distributions and concentrations of Oxy(HbO) and deoxy (HbR) haemoglobin in the brain. This technique is based on near-infrared (NIR) light and provides continuous measures of changes in oxygenated haemoglobin (HbO), deoxygenated haemoglobin (HbR) and total haemoglobin (HbT) concentrations (Gibson and Dehghani, 2009; Izzetoglu et al., 2005; Wolf et al., 2007).

A recent review of current state of the art in the Brain computer interface technology at the fourth international meeting of the BCI community has stressed the need to develop robust and wearable brain acquisition methods for domesticating BCIs (Vaughan and Wolpaw 2011; Nicolas-Alonso and Gomez-Gil 2012). Near infrared spectroscopy (NIRS) has been identified to provide a better, easier asynchronous control of BCI applications for people unable to control EEG based BCIs (Naito, Michioka et al. 2007; Jackson, McClendon et al. 2010; Power, Falk et al. 2010; Sorger, Reithler et al. 2012). But there are limitations of transfer rates for a BCI developed using only NIRS (Coyle, Ward et al. 2007; Sitaram, Caria et al. 2009). Interestingly, a BCI developed using simultaneous NIRS and EEG acquisitions has been reported to enhance the BCI performance by at least 5% (Fazli, Mehnert et al. 2012). It is envisaged that, if combined, EEG and NIRS can provide robust BCI control by providing an accessible, portable, wearable solution not only on the bedside but also for home-use. There are several advantages of combining a NIRS-EEG for BCI development. Specifically, NIRS is

relatively robust to movement, has better spatial sensitivity, and allows non-invasive measurement of localized cognitive activity, thereby, empowering the BCI with minimal training while EEG provides optimized, precise sensor placement and quick transfer rates.

3. BCI DEVELOPMENT PROCESS

The building blocks of a BCI are depicted in figure 2. The signals from the brain are acquired by scalp electrodes, or intracranial electrodes and are processed to extract features such as the amplitudes of evoked potentials, sensorimotor cortex rhythms, or firing rates of cortical neurons that reflect the user's intent. These features are then translated into commands that operate devices such as a simple word processing program (Krusienski et al., 2008), a wheelchair (Leeb et al., 2007), or a neuroprosthesis (Schalk et al., 2004).

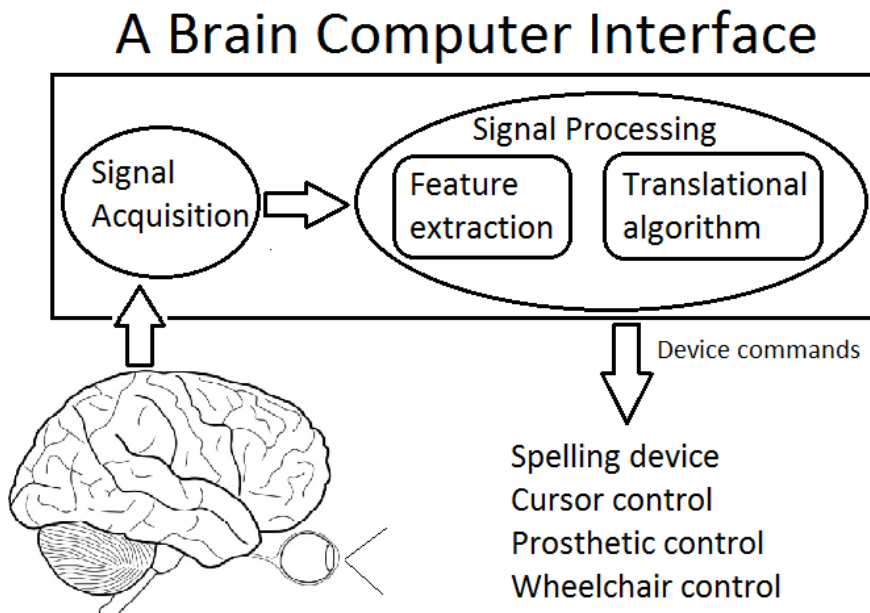


Fig 2: The basic framework of a BCI (adopted from (Schalk et al., 2004).

A more standardized model of a BCI was presented by Mason & Birch, (2003), who presented a taxonomy of the terms more often used in development of the BCI (Mason and Birch, 2003). This model lists the main components which would define the basic BCI arrangement, namely the user (the person/the subject), electrodes (sensors to convert the user's brain state to electrical signals), amplifier, feature extractor (transforms the amplified electrical signals that correspond to related neurological states into feature values), feature translator (converts the feature values into logical and 'feed-able' control signals), the control interface (maps the control signals to the specific device to be controlled), the device controller (translates semantic control signals from the control interface into physical signals that can be used within a device) and the device itself. This was regarded as a standard model and streamlined the BCI terminology but still required some clarifications. For example, the feature translator may be renamed the feature classifier or simply the classifier for discrete control signals.

This model, shown in figure 3, was designed using EEG signals as the 'source of information'; a similar analogy can be inferred for imaging sources such as fMRI and positron emission tomography (PET). To explain the analogy, the subject would remain the same, while the counterpart of the electrodes, would be replaced by the imaging equipment. The other terminology for the model in terms of its applicability to the imaging techniques explicitly remain identical but have significant implicit changes for feature extraction and translational algorithms. It is pertinent to state here that for BCI development, the imaging techniques works in tandem with signals as sources to extract relevant information from the brain and its underlying processes while the task is being performed by the subject. The imaging techniques, so far, have not been justified as sources of information which could sustain BCI development on their own.

Pre-print version

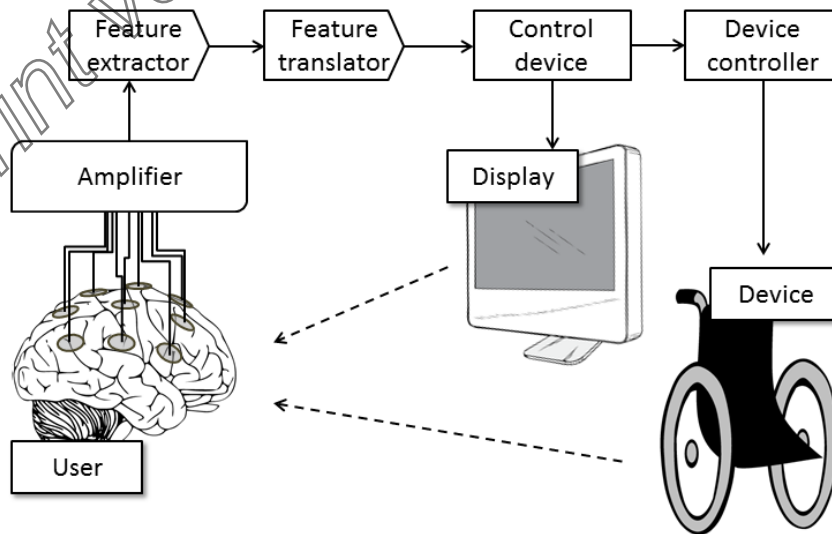


Fig 3: The BCI block diagram [adopted from (Mason and Birch, 2003).

4. TYPES OF BCI

BCIs are categorised on the basis of their method of working and their functionality. Some of the key classifications and terminology associated with present day BCI are considered as follows.

4.1 Invasive and non-invasive BCIs

Non-invasive BCIs are based on EEG measured with the scalp electrodes. In invasive BCIs, the electrical activity of the brain is recorded from inside the head (e.g., from the cerebral cortex) via intracranial electrodes or microelectrodes. These BCIs can be based on, for example, ECoG recordings (Schalk and Leuthardt, 2011; Scherer et al., 2003). Microelectrodes can also be used to record activity of a single neuron which gives rise to ‘spikes’ as sources of brain information from which BCIs can be developed (Kennedy et al., 2004).

4.2 Synchronous and asynchronous BCIs

Many BCIs work in a synchronous mode, i.e. in an externally paced mode. This requires the user to produce specific mental states in a predefined time window. For synchronous BCI, the control is system-initiated. In an asynchronous mode, the brain activity is analyzed continuously and the user can freely initiate the specific mental task(s) used as the control signal(s); the control is not system-initiated but user-initiated. This requires the BCI to detect when the EEG correlates of intended control occur. For example, Mason and Birch have tried to implement an asynchronous BCI with a switch (Mason and Birch, 2003).

5. COMPUTATIONAL TECHNIQUES

The heart of the BCI development process concerns the techniques employed for feature extraction and classification of the data. Apart from the application, the construction and detailed implementation aspects of the central translational algorithm i.e. the methods used for feature extraction and classification also determine the selection of the different brain signals, the recording technique and the equipment. All these are interrelated and the selection of these regulates the aspects of the efficiency and reliability of the BCI module developed. In this section we try to establish benchmark criteria for selection of the various techniques being used by BCI researchers. The international meetings on brain computer interface technology (Wolpaw et al., 2000), (Vaughan et al., 2003), (Vaughan and Wolpaw, 2006) (Vaughan and Wolpaw, 2011), and (Huggins et al., 2014) have been instrumental in providing a forum for research and clinical experts. The discussions at such meetings have been documented and have helped in identifying the desirable characteristics of these techniques. Some of the key characteristics are precision, responsiveness (speed), interpretability and ease of setting up.

To get more information from the brain, medical instrumentation companies have come up with computer based monitoring systems which can record a large number of channels from the brain giving rise to multichannel data (for example, 128 channel recordings). The challenge lies in the processing of this data and for example cleaning it of artefacts and then modifying it in order to extract relevant information/knowledge about the

underlying neurological processes going on in the brain at the time it is recorded. In the case of online BCIs, where the data is processed in real time, it is necessary to use appropriate techniques for processing. These might be different from the techniques used for offline processing of the BCI data or to investigate novel paradigms.

5.1 Dimensionality reduction

BCIs base their control upon the characterisation and classification of biophysiological datasets (EEG) (Niedermeyer and Silva, 2005), (ECoG) (Menon et al., 1996). However, biophysiological datasets may often be multi-dimensional. For example, EEG data may be recorded from a large number of electrodes positioned across the scalp. The data is often broad frequency and spread across a range of time periods, resulting in potentially very high dimensional data for use in BCI control.

However, for many cognitive tasks, changes in neurological activity may only be observed in a subset of the data, for example within specific time periods, particular frequency bands, and/or over certain channels. Additionally, effects such as volume conduction (the spread of electrophysiological activity across the scalp resulting in similar activity being recorded at several EEG electrodes) and high levels of redundancy in the data may mean similar information is available across multiple dimensions in the data.

Dimensionality reduction techniques aim to identify an optimal sub-set of dimensions from a highly dimensional dataset. For example, they may be used to extract a subset of EEG channels that contain the most relevant information pertaining to a particular cognitive process (Singh et al., 2007, 2012; Yang et al., 2012).

A number of computational techniques are available for dimensionality reduction. These include principal component analysis (PCA), single value decomposition (SVD), and canonical correlation analysis (CCA) etc. These three examples of dimensionality reduction methods are each discussed below.

5.1.1 Principal component analysis

Principal component analysis (PCA) attempts to identify an orthogonal transformation that translates a set of potentially correlated variables into a set of linearly uncorrelated variables. These new variables are referred to as the principal components (PCs) (Smith, 2002).

PCA operates by first subtracting the mean from the data to centre it. Thus, for a feature set X a new, zero mean, feature set \bar{X} is derived by subtracting the mean from X . The covariance matrix of the zero mean feature set \bar{X} is then used to measure the strength of the relationships between all rows of \bar{X} . This is defined as the matrix C where each element $C_{i,j}$ denotes the covariance between rows i and j in the feature set \bar{X} .

$$C_{i,j} = \frac{\sum_{k=1}^n (X_{i,k} - \bar{X}_{i,:}) (X_{j,k} - \bar{X}_{j,:})}{(n - 1)}$$

(5.1)

where $X_{i,k}$ and $X_{j,k}$ denote the k 'th features from different examples of the data i and j , of the data and $\bar{X}_{i,:}$ denotes the mean over a feature vector for an individual example of the data i .

The covariance matrix is then decomposed into a matrix of eigenvectors and a vector of eigenvalues. This may be defined as

$$Cu = \lambda u,$$

where u denotes an eigenvector of the covariance matrix C , and λ denotes the corresponding eigenvalue.

The eigenvalues identified for C may be ranked in decreasing value. The corresponding eigenvectors then contain projections of the feature set onto the principal components and are ordered by decreasing variance. Thus, the eigenvector corresponding to the largest eigenvalue contains a projection of the feature set X which has the greatest variance.

PCA is used in a large number of BCIs where dimensionality reduction is required. For example, in (Kottaimalai et al., 2013). PCA is used to identify a subset of electrodes for classifying 5 different mental tasks (resting, letter association, math, visual counting, and geometric figure rotation) via an artificial neural network (ANN). Accuracies of up to 100% are achieved when PCA is used, which is reported to be significantly higher than accuracies achieved without PCA. PCA is also very helpful in initial data exploration and to estimate the interrelationships between BCI classes/tasks in the data (Singh et al., 2012). It can also indicate the minimum number of channels that can be used for classification of novel BCI tasks (Azar et al., 2014; Singh, 2009).

5.1.2 Singular value decomposition

Singular value decomposition (SVD) attempts to identify a factorization of a matrix of dimensions $m \times n$. SVD is closely related to eigen decomposition, but may be applied to matrices of any values of m and n , whereas eigen decomposition may only be applied to square matrices.

For a matrix A of dimensions $m \times n$ the SVD is given by

$$A = USV^T,$$

where U and V denote orthogonal matrices of dimensions $m \times m$ and $n \times n$ respectively, and S denotes an $m \times n$ rectangular diagonal matrix. The columns of U are denoted u_i and contain the left singular vectors, while the columns of V are denoted v_i and contain the right singular vectors. The diagonal elements of S are denoted σ_i and contain the singular values.

Singular values are sorted in descending order and used to identify the first n left or right singular vectors. These singular vectors may then be taken to represent a reduced subset of the data.

Singular value decomposition may be used directly with EEG data to reduce dimensionality. For example, in (Daly, Nicolaou, et al., 2013) three approaches to artifact removal are compared, including an approach, originally proposed in (Teixeira et al., 2005), for using singular value decomposition as a dimensionality reduction step in an artefact removal method.

The SVD-based artefact removal method is observed to remove significant proportions of ocular artefacts and is comparable in performance to a number of other state-of-the-art artefact removal methods.

5.1.3 Canonical correlation analysis

Canonical correlation analysis (CCA) is a method for identifying pairs of vectors from two matrices with maximum correlations. For example, for two matrices X and Y of dimensions $x \times n$ and $y \times n$ respectively CCA will attempt to find a subset of elements of X and Y which are maximally correlated.

CCA may be defined as a maximisation problem, which can be solved by maximising the following term

$$\max_{w_x, w_y} \rho(u, v) = \frac{w_x^T C_{xy} w_y}{\sqrt{(w_x^T C_{xx} w_x)(w_y^T C_{yy} w_y)}}$$

where C_{xx} and C_{yy} denote the autocovariance matrices of X and Y respectively and C_{xy} denotes the cross covariance matrix of X and Y .

CCA may be used to identify a subset of dimensions of an EEG dataset which maximally correlate with some interesting properties of the data. For example, in (De Clercq et al.,

2006) CCA is used to identify muscle artefacts in the EEG via their relatively low auto-correlation with one another when compared to the EEG.

CCA is also commonly used as the basis for classifying steady state visual evoked potentials (SSVEPs). For example, in (Daly, Billinger, et al., 2013) CCA is used in an online brain-computer interface (BCI) to identify frequency bands in the EEG which maximally correlate with SSVEP stimulation frequencies, and hence identify which stimuli a user is attending to and, therefore, their intended control action.

5.2 Feature extraction

Biophysiological data may be described via a number of different feature types. When considering neurological data measured by either the EEG or ECoG there are three broad groups of features that may be extracted. These are features based upon the amplitude of the data, features based upon the frequency content of the data, and features based upon the phase content of the data (Lotte, et.al., 2007; Hwang, et.al., 2013). It's also possible to consider feature types in which two or more types of feature are combined. For example, time-frequency features may be used to describe changes in amplitude of the data across different frequency bands.

Features based upon the amplitude of the data include measures of the peaks in the data. For example, the event-related potential (ERP) is a change in amplitude of the EEG in response to certain stimuli or cognitive processes and may be identified via the size and/or latency of the peak amplitude. Amplitude based features may also include measures of distributions of the data, measures of relationships between different amplitudes (e.g. correlation), and statistical measures of amplitude differences in the data.

Frequency based features may be used to describe how the frequency content of the signals change over time or in relation to certain events or stimuli. For example, BCI control may be based upon SSVEPs, an increase in frequency content in a narrow frequency band, in response to entrainment by an external stimuli, of a particular neural

oscillators in the visual cortex (Müller-Putz et al., 2005). Relationships between different measures of frequency content (e.g. coherence) may also be used to measure how frequency content changes across particular regions of the brain.

Phase based features are traditionally used much less in BCI control than other feature types (Hwang, et.al., 2013). However, they have been shown, in some cases, to exhibit significant improvements in performance compared to some traditional features such as the event-related (de)synchronisation (a combined amplitude-frequency feature) (Daly et al., 2012).

Features based upon a combination of different feature types are becoming increasingly popular in BCI research (Hwang et al., 2013). It's likely that the recent development and exploration of hybrid BCIs (BCIs that combine two or more control mechanisms or physiological measures (Pfurtscheller et al., 2010)) are driving an increased interest in combined feature types (Hwang et al., 2013).

5.3 Feature selection

Feature selection refers to techniques which search a set of possible features and identify a subset of those features which are optimal for some purpose. For example, consider EEG recorded during an ERP oddball task from 19 channels positioned over the scalp. The oddball task is designed to produce a P300 ERP by presenting a selection of stimuli with a small probability of an out of sequence stimuli (i.e. different in some property from the majority of the other stimuli). The resulting P300 ERP is usually most apparent over the occipital, parietal cortices. Hence, EEG channels positioned over this region are more likely to see increased EEG amplitude. A feature selection approach may be used to identify which channels exhibit this increase in amplitude in response to the unexpected (target) stimuli.

Feature selection may be described as manual, automated, or semi-automated. Manual feature selection refers to the process of selecting feature sets based upon prior

knowledge of the dataset and expected features of interest. Automated feature selection refers to computer driven selection of features and may often be performed via a machine learning (ML) technique. Semi-automated techniques attempt to combine these two methods to take advantage of the benefits of both. For example, a region of interest may be manually selected and an automated method then used to select the best feature set from within this region.

Automated feature selection may be performed by either supervised or unsupervised methods. Supervised methods include meta data about the training data in the selection processes, while unsupervised methods do not. Typically, unsupervised methods amount to dimensionality reduction methods as discussed in section 5.4.

Supervised methods may be described as filter based methods; wrapper based methods, or embedded methods.

Filter based methods attempt to select features independently of the classification step (Tangermann, 2007). Relationships between different features may be used to identify the most relevant features i.e. those which give the most information about the cognitive phenomena of interest. As such filter methods may be either supervised or unsupervised.

The main advantage of filter methods often lies in their computational efficiency. There is no need to optimize an objective function by repeated re-evaluations with a filter method as one would when applying a wrapper (Kohavi, 1997). As such they can be considered independent of the classification method chosen. This makes filter methods very scalable and potentially able to offer similar levels of performance on both large and small data-sets.

Typical examples of filter techniques include clustering techniques for dimensionality reduction (Gan, 2007) and measures of feature similarity such as correlation (Hall, 1999). When applying a similarity measure features that are very similar to another can be

thought to be redundant and hence removed. When a supervised filter is applied the amount of information each feature gives about the class labels, for example the correlation between features and the class labels, can give an indication of how suitable a particular feature is for classifying the dataset correctly (Hall, 1999).

A wrapper method attempts to optimize an objective function by repeated re-evaluation with different candidate feature sets (Kohavi, 1997). The objective function is the chosen classifier for the BCI hence the choice of classifier becomes an integral part of the feature selection process. Subsequently the classifier cannot be evaluated without the class labels and all wrapper methods are, therefore, supervised.

Many wrapper methods are based upon the idea of meta-heuristics. A meta-heuristic search attempts to traverse a search space in such a way as to avoid getting stuck in local optima solutions (Guyon, 2003). They do this by either using multiple population members with randomization, by allowing backtracking of some variety, or a combination of the two. They are often based upon biologically inspired search metaphors (Kohavi, 1997).

Examples of meta-heuristic wrapper techniques that are popular in BCI research include Genetic Algorithms (GAs), which are inspired by natural evolution (Kohavi, 1997) and (Yorn-Tov and Inbar, 2001). A population of candidate feature sets is created. Each member of this population is then evaluated against some objective function (the classifier) and the members that give the highest classification accuracies are taken to be the 'fittest'. These members are 'bred' with other fit members to create new 'fitter' child members (Yorn-Tov and Inbar, 2001). Over enough generations the overall fitness of the population increases and in ideal circumstances arrives at an optimal feature set which allows highly accurate classification of the data. Random mutation is applied to each generation to attempt to avoid the solution getting stuck in local maxima in the way a simple gradient ascent algorithm would (Goldberg and Holland, 1988).

An example of a type of genetic algorithm being used in feature selection for BCIs is provided in (Daly et al., 2011). In this example differential evolution (a variant of a genetic algorithm) is used to identify optimal channels and frequency bands for ERD detection during a motor imagery task.

An embedded method attempts to combine feature selection and classification into one step (Guyon, 2003). An example of this is an Artificial Neural Network (ANN). In the training routine for ANNs the weights of the network are adjusted such that optimal classification accuracy may be achieved with the best feature set. Therefore the neural network incorporates both a feature selection step and a classification step (Schalk et al., 2004). Embedded techniques are often quicker than wrapper techniques as there is often less need for multiple re-evaluations of candidate feature sets against an objective function in the training process (Guyon, 2003).

Several BCI systems incorporate embedded methods for feature selection. However, it is not always clear that this is what is happening as the feature sets may be selected based, in part, upon a-priori knowledge of the paradigm. Thus, systems such as (Pregenzer and Pfurtscheller, 1999) select a set of features that are known a-priori to be effective at differentiating classes related to the paradigm. This feature set is then passed to a classifier with a further embedded feature selection step which sub-selects optimal features from this candidate feature set.

5.4 Classification techniques

Classification techniques attempt to identify which class a previously unseen dataset belongs to by applying a classification rule to the data. The classification rule is typically trained on previously seen examples of the data from each of the available classes before being applied to new data.

Formally, for a two class problem a function is estimated as $f: R^N \rightarrow \{-1, +1\}$, where a new, previously unseen, dataset is labeled either as -1 or +1 depending upon which side

of a decision boundary it falls. The process of identifying an optimal function f from training data effectively amounts to learning the class membership / learning the inter-class decision boundary.

Classifier training may often be split into several stages to attempt to minimise the bias-variance trade-off problem. This typically involves using a training set to first estimate the classifier's decision boundary, followed by a validation set, which may be used to test and further refine the boundary. The final data the classifier is to be applied to is referred to as the testing set.

Classifier techniques are used in the majority of BCI applications (Hwang et al., 2013). For example, in (Jin et al., 2013) a Bayesian linear discriminant analysis (BLDA) classifier is used to identify trials which contain P300 event-related potentials (ERPs) for the control of an ERP based online BCI. BLDA operates by attempting to identify a linear combination of features which optimally separate them into the correct classes via a decision boundary. Mean online classification accuracies of $94.5 \pm 5.1\%$ are obtained.

Other types of classification technique that are popular for use in BCIs include support vector machines (SVMs), hidden markov models (HMMs), artificial neural networks (ANNs) etc. (Hwang et al., 2013). To discuss an example of one more of these classifier types hidden markov models (HMMs) are considered.

HMMs attempt to characterise the temporal dynamics of a time series of bio-signals as a markov processes. Specifically, the observed datasets are assumed to be generated by an underlying markov process, with each state in that markov process generating the observed features according to some probability distribution. Estimation of these distributions and state transition probabilities for a markov process may be performed as part of the classifier training process. A HMM is typically trained for each class and then used to classify new data by identifying which HMM has the greatest probability of having generated the new data.

HMMs have been successfully applied to classify EEG data in a BCI context in (Daly et al., 2012). In this example EEG data recorded during a motor imagery task is first characterised via phase based features before HMMs are trained to differentiate left vs. right index finger motor imagery. Highly significant classification accuracies are achieved using this approach.

Another algorithm used for classification purposes is the Linear Discriminant Analysis (LDA) classifier. LDA is a simple technique widely used in BCI applications because it is computationally efficient and robust. This method maximizes the ratio of ‘between-class’ variance to the ‘within-class’ variance in any particular dataset thereby guaranteeing maximal separability (Perez and Cruz, 2007). It can also handle the cases in which the ‘within-class’ frequencies are unequal. Evolving from the same principles as of PCA, it holds the advantage over PCA in that the shape and location of the original datasets changes when transformed to a different space whereas LDA does not change the location but only tries to provide more class separability and identifies a decision boundary between the given classes.

The future of the BCI techniques lies in the standardisation of the computational methods. This will not only help to launch a BCI based product in the market but will also broaden the spectrum of BCI applications.

6. BCI EVALUATION

BCI research has reached a level where there are large varieties of combinations of feature extraction and classification techniques being used by research groups all over the world (Hwang et al., 2013). It is important to have benchmark criteria which not only provide a quick and easy means of identifying the relative success of the technique but also act as a standard to compare the relative performance of all the techniques. Possible criteria are discussed in (Billinger et al., 2012) and include classification accuracy, Cohen’s Kappa coefficient, mutual information of discrete, and continuous output.

Two measures of classification accuracy may be considered, overall accuracy and individual accuracy. Overall accuracy is calculated as the ratio of the sum of the diagonal elements of the classifier generated confusion matrix and the total number of samples. The specific accuracy for each class is the ratio of the diagonal element (hits) for that particular class and the total number of sample points for that class (sum of the row elements of the confusion matrix for that particular class).

The output of the validation procedure is devised in the form of specific parameters using BCI terminology. Transfer rates may be calculated to measure the speed with which users may operate a BCI. Cohen's kappa coefficient, which is reported to be a better representation of the accuracy as it takes into account the occurrence of a chance (Schlogl et al., 2007), may also be used. Kappa coefficient is 1 in case of a perfect classification and is 0 if the predicted class has no correlation with the actual class. Specific accuracy for each of the two classes gives an idea of the performance of the classification algorithm for each of the classes. Correlation and signal to noise ratio may also be used.

7. Conclusion

Brain-computer interfaces (BCIs) are a rapidly growing field (Hwang et al., 2013), with a large and growing number of researchers attempting to tackle the multiple challenges presented by this highly interdisciplinary area of research.

BCIs may be described by multiple component stages. These have been outlined and described in this chapter and include components for data acquisition, pre-processing, dimensionality reduction, feature extraction/selection, classification, and application. Each of these component parts is researched with the aim of developing novel and improved components and applications to allow BCIs to meet the needs of new user groups and improve the performance of BCI systems provided to existing user groups.

One such example is the hybrid BCI, in which two or more different data acquisition methods or BCI paradigms may be combined (Pfurtscheller et al., 2010). For example,

simultaneous electroencephalogram (EEG) and near infrared spectroscopy (NIRS) may be combined with the aim of using the combination of electrophysiological and haemodynamic responses to provide more accurate control to BCI user groups (Fazli et al., 2012).

Another example is the use of BCIs in neuro-rehabilitation for stroke (Ang et al., 2010). This represents a novel application of BCI technology, in which closed loop BCI control of an application is utilised to induce changes in levels of neuroplasticity in the user. The aim is to promote beneficial neuroplastic changes which allow compensatory neural areas to “take over” from lesioned cortical regions (Silvoni et al., 2011).

In each of these new developments in BCI technology the heart of the BCI remains the translational algorithms. Indeed, translational algorithms retain an essential role in all types of BCI, that of converting raw, noisy, non-stationary neural data into stable and realisable control commands for a computer, robotic device, or other application.

We categorise translational algorithms into four different types: dimensionality reduction, feature extraction, feature selection, and classification. Not every BCI type requires every one of these types of translational algorithm, but all BCI types require some combination of some of these translational algorithms.

Dimensionality reduction algorithms are typically used as a pre-processing step in the processing pipeline at the heart of a BCI. They aim to take a high dimensional dataset (for example, one composed of multiple channels, sample points, frequency bands, etc.) and reduce it to just the most relevant dimensions of interest. Typical examples of dimensionality reduction methods were discussed in this chapter and include methods such as principal component analysis and singular value decomposition.

Feature extraction refers to the method(s) by which raw brain signals are translated into features describing some relevant property of the brain state. Raw signals are the signals recorded from the brain at the data acquisition stage and can include EEG amplitudes, electrocortigram (ECoG) amplitudes, blood oxygenation levels (fNIRS, fMRI) etc. The

process of converting these raw signals to features of interest may vary considerably depending on the type of raw data, the cognitive task(s) the user is attempting in order to control the BCI, and the processing limitations of the available computers. Examples can include measures of signal magnitude (e.g. EEG amplitude), measures of frequency content of the data, and more complex measures such as connectivity measures between different regions of interest within the brain.

Feature selection refers to the automated or semi-automated selection of a relevant subset of the features extracted from the raw brain signals. Automated feature selection may be used in cases where there isn't a clear set of well known "good features" for the particular cognitive paradigm. It may also be used to provide some compensation for the problem of inter-subject variability in neural data (Niedermeyer and Silva, 2005). The neurological data recorded from two individuals is not completely alike and automated feature selection may be used to find the best set of features for a particular individual during a particular BCI control task. Feature selection is most commonly performed using methods from machine learning and may include methods such as genetic algorithms, random forest searches etc.

Finally, classification refers to the identification of corresponding discrete class labels for a selected feature set. This may also be performed using tools taken from the machine learning toolbox, for example, support vector machines (SVMs) or linear discriminant analysis (LDA) classifiers, and may be combined with the feature selection method or performed independently.

New algorithms and methods are in a constant state of development by the BCI community and are opening up increasing possibilities for applications, data acquisition types, and target user groups. Many of these new possibilities require advancements in translational algorithms. Thus, research and development of translational algorithms, both by the BCI community and others, may be seen as the engine of BCI research and development and forms a vital part of the field of multidisciplinary BCI research.

Regarding the future directions of BCI research it's possible to make a few statements about the role of translational algorithms. First, every existing method of data acquisition from the central nervous system produces data which is non-stationary and noisy. Thus, there is likely to be a need for a number of the translational algorithms described in this chapter for the foreseeable future in BCI for tasks such as de-noising the data, extracting and selecting features of interest, and identifying the relevant class labels corresponding to the data. Second, while there will be a need for these algorithms they will not all be required in every application. For example, features may be mapped directly to a control action, removing the need for a classifier to identify corresponding class labels, or the need for feature selection may be mitigated if a-priori knowledge of suitable features is available.

Finally, advances in computational technology and machine learning research mean that new and more advanced translational algorithms are constantly been developed. This development is not just confined to the BCI research community and often machine learning algorithms developed for other purposes are co-opted into the community (for example LDA classifiers (Xu et al., 2011)), while algorithms developed within the community may also find other uses elsewhere (for example common spatial patterns (Koles et al., 1990)). This process is likely to continue and accelerate of the coming years and advancements in computing technology allow the deployment of ever increasingly advanced algorithms. An example of this may be seen in the use of machine learning algorithms for feature selection. More advanced machine learning algorithms may be employed as computing technology allows them to be run in sufficiently short times to allow their use during online BCI applications.

Acknowledgements

The authors are thankful to Mathew Dyson, PhD student, Computer Science Department, University of Essex for his support in recording the dataset used in the case study described in this work. The authors are thankful to the BCI research community for their support and guidance through their publications and resources on the web.

References

- Ang KK, Guan C, Chua KSG, et al. (2010) Clinical study of neurorehabilitation in stroke using EEG-based motor imagery brain-computer interface with robotic feedback. In: *Conf Proc IEEE Eng Med Biol Soc*.
- Azar A, Balas V and Olariu T (2014) Classification of EEG-Based Brain-Computer Interfaces. In: Iantovics B and Kountchev R (eds), *Advanced Intelligent Computational Technologies and Decision Support Systems SE - 9*, Studies in Computational Intelligence, pp. 97–106.
- Billinger M, Daly I, Kaiser V, et al. (2012) Is it significant? Guidelines for reporting BCI performance. In: *Toward Practical BCIs: Bridging the Gap from Research to Real-World Applications*.
- Birbaumer N (2003) The Thought-Translation Device (TTD): Neurobehavioral Mechanisms and Clinical Outcome. *Rehabilitation*, 11(2), 120–123.
- Birbaumer N (2006) Breaking the silence: brain-computer interfaces (BCI) for communication and motor control. *Psychophysiology*, 43(6), 517–32.
- Chen X, Bin G, Daly I, et al. (2013) Event-related desynchronization (ERD) in the alpha band during a hand mental rotation task. *Neuroscience Letters*.
- Daly I, Nasuto SJ and Warwick K (2011) Single tap identification for fast BCI control. *Cognitive Neurodynamics*, 5(1), 21–30.
- Daly I, Nasuto SJ and Warwick K (2012) Brain computer interface control via functional connectivity dynamics. *Pattern Recognition*, 45(6), 2123–2136.
- Daly I, Nicolaou N, Nasuto SJ, et al. (2013) Automated artifact removal from the electroencephalogram: a comparative study. *Clinical EEG and neuroscience*, 44(4), 291–306.
- Daly I, Billinger M, Laparra-Hernández J, et al. (2013) On the control of brain-computer interfaces by users with cerebral palsy. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, 124(9), 1787–97.
- De Clercq W, Vergult A, Vanrumste B, et al. (2006) Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram. *IEEE transactions on bio-medical engineering*, 53(12 Pt 1), 2583–7.
- Donchin E, Heffley E, Hillyard SA, et al. (1984) Cognition and Event-Related Potentials II. The Orienting Reflex and P300. *Annals of the New York Academy of Sciences*, 425(1 Brain and Inf), 39–57.

- Ebersole JS and Pedley TA (2003) Current practice of clinical electroencephalography, 3rd edn. *European Journal of Neurology*, 10(5), 604–605.
- Farwell LA and Donchin E (1988) Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and clinical neurophysiology*, 70(6), 510–23.
- Fazli S, Mehnert J, Steinbrink J, et al. (2012) Enhanced performance by a hybrid NIRS-EEG brain computer interface. *NeuroImage*, 59(1), 519–29.
- Friedrich EVC, Scherer R and Neuper C (2012) The effect of distinct mental strategies on classification performance for brain–computer interfaces. *Int J Psychophysiol*, 84, 86–94.
- Gan G (2007) *Data Clustering: Theory, Algorithms and Applications*.
- Gibson A and Deghani H (2009) Diffuse optical imaging. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 367(1900), 3055–72.
- Goldberg DE and Holland JH (1988) Genetic Algorithms and Machine Learning. *Machine Learning*, 3(2), 95 – 99.
- Grimann B, Huggins J., Levine S., et al. (2002) Visualization of significant ERD/ERS patterns in multichannel EEG and ECoG data. *Clinical Neurophysiology*, 113(1), 43–47.
- Guger C, Harkam W, Hertenæs C, et al. (1999) Prosthetic Control by an EEG-based Brain-Computer Interface (BCI). In: *Proceedings of the AAATE Conference*.
- Guyon I (2003) An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research*, 3(7-8), 1157–1182.
- Hall MA (1999) Correlation-based Feature Selection for Machine Learning. *PhD Thesis*.
- Handy TC (2005) *Event-related Potentials: A Methods Handbook*.
- Hill NJ, Lal TN, Schröder M, et al. (2006) Classifying EEG and ECoG signals without subject training for fast BCI implementation: comparison of nonparalyzed and completely paralyzed subjects. *IEEE transactions on neural systems and rehabilitation engineering*, 14(2), 183–6.
- Huggins JE, Guger C, Allison B, et al. (2014) Workshops of the Fifth International Brain-Computer Interface Meeting: Defining the Future. *Brain-Computer Interfaces*, 1(1), 27–49.
- Hwang H-J, Kim S, Choi S, et al. (2013) EEG-based Brain-Computer Interfaces (BCIs): A Thorough Literature Survey. *International Journal of Human-Computer Interaction*, 29(12), 130429122442009.

- Izzetoglu M, Izzetoglu K, Bunce S, et al. (2005) Functional near-infrared neuroimaging. *IEEE transactions on neural systems and rehabilitation engineering: a publication of the IEEE Engineering in Medicine and Biology Society*, 13(2), 153–9.
- Jin J, Sellers EW, Zhang Y, et al. (2013) Whether generic model works for rapid ERP-based BCI calibration. *Journal of neuroscience methods*, 212(1), 94–9.
- Kennedy P, Andreasen D, Ehirim P, et al. (2004) Using human extra-cortical local field potentials to control a switch. *Journal of neural engineering*, 1(2), 72–7.
- Kohavi R (1997) Wrappers for feature subset selection. *Artificial Intelligence*, 97(1-2), 273–324.
- Koles ZJ, Lazar MS and Zhou SZ (1990) Spatial patterns underlying population differences in the background EEG. *Brain Topography*, 2(4), 275–284.
- Kottaimalai R, Rajasekaran MP, Selvam V, et al. (2013) EEG signal classification using Principal Component Analysis with Neural Network in Brain Computer Interface applications. In: *2013 IEEE International Conference ON Emerging Trends in Computing, Communication and Nanotechnology (ICECCN)*, pp. 227–231.
- Krusienski DJ, Sellers EW, McFarland DJ, et al. (2008) Toward enhanced P300 speller performance. *Journal of neuroscience methods*, 167(1), 15–21.
- Lal TN, Schr M, Hill NJ, et al. (2005) A Brain Computer Interface with Online Feedback based on Magnetoencephalography. In: *Proceedings of 22nd International Conference on Machine Learning*.
- Leeb R, Friedman D, Müller-Putz GR, et al. (2007) Self-paced (asynchronous) BCI control of a wheelchair in virtual environments: a case study with a tetraplegic. *Computational Intelligence and Neuroscience*.
- Luck SJ (2005) Ten simple rules for designing ERP Experiments. In: Handy TC (ed.), *Event-Related Potentials - A methods handbook*, pp. 17–32.
- Mason SG and Birch GE (2003) A general framework for brain-computer interface design. *IEEE transactions on neural systems and rehabilitation engineering: a publication of the IEEE Engineering in Medicine and Biology Society*, 11(1), 70–85.
- Mellinger J, Schalk G, Braun C, et al. (2007) An MEG-based brain-computer interface (BCI). *NeuroImage*, 36(3), 581–93.
- Menon V, Freeman WJ, Cuttillo BA, et al. (1996) Spatio-temporal correlations in human gamma band electrocorticograms. *Electroencephalography and Clinical Neurophysiology*, 98(2), 89–102.

- Miranda ER, Magee WL, Wilson JJ, et al. (2011) Brain-Computer Music Interfacing (BCMI): From Basic Research to the Real World of Special Needs. *Music and Medicine*, 3(3), 134–140.
- Müller-Putz GR, Scherer R, Brauneis C, et al. (2005) Steady-state visual evoked potential (SSVEP)-based communication: impact of harmonic frequency components. *Journal of neural engineering*, 2(4), 123–130.
- Nam CS, Li Y and Johnson S (2010) Evaluation of P300-Based Brain-Computer Interface in Real-World Contexts. *International Journal of Human-Computer Interaction*, 26(6), 621–637.
- Neuper C, Scherer R, Reiner M, et al. (2005) Imagery of motor actions: differential effects of kinesthetic and visual-motor mode of imagery in single-trial EEG. *Brain research. Cognitive brain research*, 25(3), 668–77.
- Niedermeyer E and Silva FHL Da (2005) *Electroencephalography: basic principles, clinical applications, and related fields*.
- Nijholt A and Desney T (2009) *BCI for games: A state of the Art survey*. Lecture notes in computer science.
- O'Brien JH (1982) P300 wave elicited by a stimulus-change paradigm in acutely prepared rats. *Physiology & behavior*, 28(4), 711–3.
- Perelmouter J and Birbaumer N (2000) A binary spelling interface with random errors. *IEEE Transactions on Rehabilitation Engineering*, 8(2), 227–232.
- Perez JLM and Cruz AB (2007) Linear Discriminant Analysis on Brain Computer Interface. In: *2007 IEEE International Symposium on Intelligent Signal Processing*, pp. 1–6.
- Pfurtscheller G and Lopes da Silva FH (1999) Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical Neurophysiology*, 110(11), 1842–1857.
- Pfurtscheller G, Müller-Putz GR, Schlögl A, et al. (2006) 15 years of BCI research at Graz University of Technology: current projects. *IEEE transactions on neural systems and rehabilitation engineering*, 14(2), 205–210.
- Pfurtscheller G, Allison BZ, Brunner C, et al. (2010) The hybrid BCI. *Frontiers in Neuroprosthetics*, 4(30).
- Pregenzer M and Pfurtscheller G (1999) Frequency component selection for an EEG-based brain to computer interface. *IEEE Transactions on Rehabilitation Engineering*, 7(4), 413–419.
- Schalk G and Leuthardt EC (2011) Brain-computer interfaces using electrocorticographic signals. *IEEE reviews in biomedical engineering*, 4, 140–54.

- Schalk G, McFarland DJ, Hinterberger T, et al. (2004) BCI2000: a general-purpose brain-computer interface (BCI) system. *IEEE transactions on bio-medical engineering*, 51(6), 1034–43.
- Scherer R, Graimann B, Huggins JE, et al. (2003) Frequency component selection for an ECoG-based brain-computer interface. *Biomedizinische Technik. Biomedical engineering*, 48(1-2), 31–6.
- Schlogl A, Kronegg J, Huggins JE, et al. (2007) Evaluation criteria in BCI research,.
- Schomer L and Lopes de Silva F (eds) (2011) *Niedermeyer's electroencephalography: Basic principles, clinical applications, and related fields*. 6th ed.
- Sellers EW and Donchin E (2006) A P300-based brain-computer interface: initial tests by ALS patients. *Clinical neurophysiology: official journal of the International Federation of Clinical Neurophysiology*, 117(3), 538–48.
- Silvoni S, Ramos-Murguialday A, Cavinato M, et al. (2011) Brain-computer interface in stroke: a review of progress. *Clinical EEG and neuroscience*, 42(4), 245–52.
- Singh H (2009) Development of EEG based BCI approaches for detection of awareness in human disorders of consciousness.
- Singh H, Qin X, Hines E, et al. (2007) Classification and feature extraction strategies for multi channel multi trial BCI data. *International Journal of Bioelectromagnetism*, 9(4), 233–236.
- Singh H, Yang J, Singh S, et al. (2012) Channel selection for multi channel multi trial invasive BCI data. In: Shama K, Nayak KP, and Bhat S (eds), *Electronic design and signal processing*, pp. 92–96.
- Sitaram R, Caria A, Veit R, et al. (2007) FMRI brain-computer interface: a tool for neuroscientific research and treatment. *Computational intelligence and neuroscience*, 25487.
- Smith LI (2002) *A tutorial on principal components analysis*.
- Tangermann MW (2007) Feature Selection for Brain-Computer Interfaces. *Naturwissenschaften*.
- Teixeira AR, Tome AM, Lang EW, et al. (2005) On the Use of Clustering and local Singular Spectrum Analysis to Remove Ocular Artifacts from Electroencephalograms. In: *Proceedings of International Joint Conference on Neural Networks*, pp. 2514–2519.
- Tudor M, Tudor L and Tudor KI (2005) [Hans Berger (1873-1941)--the history of electroencephalography]. *Acta medica Croatica : časopis Hrvatske akademije medicinskih znanosti*, 59(4), 307–13.
- Vaughan TM and Wolpaw JR (2006) The Third International Meeting on Brain-Computer Interface Technology: making a difference. *IEEE transactions on neural systems and rehabilitation*

engineering: a publication of the IEEE Engineering in Medicine and Biology Society, 14(2), 126–7.

Vaughan TM and Wolpaw JR (2011) Special issue containing contributions from the Fourth International Brain-Computer Interface Meeting. *Journal of neural engineering*, 8(2), 020201.

Vaughan TM, Heetderks WJ, Trejo LJ, et al. (2003) Brain-computer interface technology: a review of the Second International Meeting. *IEEE transactions on neural systems and rehabilitation engineering: a publication of the IEEE Engineering in Medicine and Biology Society*, 11(2), 94–109.

Wolf M, Ferrari M and Quaresima V (2007) Progress of near-infrared spectroscopy and topography for brain and muscle clinical applications. *Journal of biomedical optics*, 12(6), 062104.

Wolpaw JR, Birbaumer N, Heetderks WJ, et al. (2000) Brain-computer interface technology: a review of the first international meeting. *IEEE transactions on rehabilitation engineering: a publication of the IEEE Engineering in Medicine and Biology Society*, 8(2), 164–173.

Wolpaw JR, McFarland DJ, Vaughan TM, et al. (2003) The Wadsworth Center brain-computer interface (BCI) research and development program. *IEEE transactions on neural systems and rehabilitation engineering: a publication of the IEEE Engineering in Medicine and Biology Society*, 11(2), 204–7.

Xu P, Yang P, Lei X, et al. (2011) An enhanced probabilistic LDA for multi-class brain computer interface. *PloS one*, 6(1), e14634.

Yang J, Singh H, Hines EL, et al. (2012) Channel selection and classification of electroencephalogram signals: an artificial neural network and genetic algorithm-based approach. *Artificial intelligence in medicine*, 55(2), 117–26.

Yorn-Tov E and Inbar GF (2001) Selection of relevant features for classification of movements from single movement-related potentials using a genetic algorithm. *2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 1364–1366.