Chapter from the book *Wavelet Transforms and Their Recent Applications in Biology and Geoscience*

Brain Computer Interface with Wavelets and Genetic Algorithms

Abdolreza Asadi Ghanbari¹, Mir Mohsen Pedram², Ali Ahmadi³, Hamidreza Navidi⁴, Ali Broumandnia⁵ and Seyyed Reza Aleaghil⁵

¹Young Researchers Club, Boroujerd Branch, Islamic Azad University, Boroujerd, Iran
²Electrical Engineering Department, Tarbiat Moallem University, Tehran, Iran
³Electrical and Computer Department, Khajeh Nasir Toosi University of Technology, Tehran, Iran
⁴Applied Mathematics and Computer Sciences Department, Shahed University, Tehran, Iran
⁵Islamic Azad University-South Tehran Branch, Tehran, Iran

1. Introduction

For many years people have speculated that electroencephalographic activity or other electrophysiological measures of brain function might provide a new non-muscular channel for sending messages and commands to the external world – a brain–computer interface (BCI). BCI is a rapidly growing field of research combining neurophysiological insights, statistical signal analysis, and machine learning (Blankertz et al., 2007). The goal of BCI research is to build a communication channel from the brain to computers bypassing peripheral nerves and muscle activity. Among different techniques for the noninvasive measurement of the human brain, the electroencephalography (EEG) is commercially affordable and has excellent temporal resolution which enables real-time interaction through BCI.

Most popular and many scientific speculations about BCIs start from the ‘mind-reading’ or ‘wire-tapping’ analogy, the assumption that the goal is simply to listen in on brain activity as reflected in electrophysiological signals and thereby determine a person’s wishes. This analogy ignores the essential and central fact of BCI development and operation. A BCI changes electrophysiological signals from mere reflections of central nervous system (CNS) activity into the intended products of that activity: messages and commands that act on the world. It changes a signal such as an EEG rhythm or a neuronal firing rate from a reflection of brain function into the end product of that function: an output that, like output in conventional neuromuscular channels, accomplishes the person’s intent. A BCI replaces nerves and muscles and the movements they produce with electrophysiological signals and the hardware and software that translate those signals into actions.

As a replacement for the brain’s normal neuromuscular output channels, a BCI also depends on feedback and on adaptation of brain activity based on that feedback. Thus, a BCI system must provide feedback and must interact in a productive fashion with the adaptations the brain makes in response to that feedback. This means that BCI operation depends on the interaction of two adaptive controllers: the user’s brain, which produces the signals measured by the BCI; and the BCI itself, which translates these signals into specific commands.
Successful BCI operation requires that the user develop and maintain a new skill, a skill that consists not of proper muscle control but rather of proper control of specific electrophysiological signals; and it also requires that the BCI translate that control into output that accomplishes the user’s intent. This requirement can be expected to remain even when the skill does not require initial training.

Present-day BCIs determine the intent of the user from a variety of different electrophysiological signals. These signals include slow cortical potentials, P300 potentials, and mu or beta rhythms recorded from the scalp, and cortical neuronal activity recorded by implanted electrodes. They are translated in real-time into commands that operate a computer display or other device. Successful operation requires that the user encode commands in these signals and that the BCI derive the commands from the signals. Thus, the user and the BCI system need to adapt to each other both initially and continually so as to ensure stable performance. Current BCIs have maximum information transfer rates up to 10–25 bits/min. This limited capacity can be valuable for people whose severe disabilities prevent them from using conventional augmentative communication methods.

At the same time, many possible applications of BCI technology, such as neuroprosthesis control, may require higher information transfer rates. Future progress will depend on: recognition that BCI research and development is an interdisciplinary problem, involving neurobiology, psychology, engineering, mathematics, and computer science; identification of those signals, whether evoked potentials, spontaneous rhythms, or neuronal firing rates, that users are best able to control independent of activity in conventional motor output pathways; development of training methods for helping users to gain and maintain that control; delineation of the best algorithms for translating these signals into device commands; attention to the identification and elimination of artifacts such as electromyographic and electro-oculographic activity; adoption of precise and objective procedures for evaluating BCI performance; recognition of the need for long-term as well as short-term assessment of BCI performance; identification of appropriate BCI applications and appropriate matching of applications and users; and attention to factors that affect user acceptance of augmentative technology, including ease of use, and provision of those communication and control capacities that are most important to the user.

Three major problems in this novel technology are identifies the brain signal features best suited for communication and artifacts that can occur during the signal acquisition.

Artifacts are undesired signals that can introduce significant changes in brain signals and ultimately affect the neurological phenomenon. Artifacts are attributed either to non-physiological sources (such as 50/60 Hz power-line noise, changes in electrode impedances, etc.) or physiological sources, such as potentials introduced by eye or body movements. Different methods for artifact removal are proposed in the literature. One of the most successful methods is Independent component analysis (ICA) (N. Xu et al., 2004). This method based on a common successful assumption in EEG research is that signals are generated by a linear mixing of independent sources in the brain and other external components and used for artifact removing of EEG signals (Hyvärinen et al., 2000). In this chapter, we use two artifacts removal methods: ICA and liner filtering.

The performance of a BCI, like that of other communication systems, depends on its signal-to-noise ratio. The goal is to recognize and execute the user’s intent, and the signals are those aspects of the recorded electrophysiological activity that correlate with and thereby
reveal that intent. The user’s task is to maximize this correlation; and the system’s first task is to measure the signal features accurately, i.e. to maximize the signal-to-noise ratio.

Feature extraction methods can greatly affect signal-to-noise ratio. Good methods enhance the signal and reduce central nervous system and non-CNS noise. This is most important and difficult when the noise is similar to the signal. For features extraction from the raw EEG data many methods such as time domain, frequency domain, and time–frequency domain are used. Since the EEG is non-stationary in general, it is most appropriate to use time–frequency domain methods like wavelet transform (WT) as a mean for feature extraction (Asadi Ghanbari et al., 2009). In this chapter, we used WT for feature extraction. The WT provides a more flexible way of time–frequency representation of a signal by allowing the use of variable sized windows.

Feature selection is one of the major tasks in classification problems. The main purpose of feature selection is to select a number of features used in the classification and at the same time to maintain acceptable classification accuracy. Various algorithms have been used for feature selection in the past decades. One of the best methods that can be used for features selection is GA (Te-Sheng et al., 2006). The GA plays the role of selector to select a subset of features that can best describe the classification. In this chapter, we employed this idea and used neural network classifier to compare the feature selection classification performance. The GA is a powerful feature selection tool, especially when the dimensions of the original feature set are large (Te-Sheng et al., 2006). Reducing the dimensions of the feature space not only reduces the computational complexity, but also increases estimated performance of the classifiers.

In this chapter, we show how we can convert EEG activity into cursor movement by a BCI using an appropriate feature extraction scheme. The proposed automated method for the classification of EEG activity is based on signal preprocessing, feature extraction and classification. The power spectrum, variance and mean of the Daubechies mother wavelet transform and Hilbert transform used for feature extraction. Finally, we implemented a feed-forward multi-layer perceptron (MLP) with a single hidden layer with five neurons, a probabilistic neural network (PNN), and support vector machine (SVM) classifier with Gaussian RBF kernel.

1.1 Definition and features of a BCI

1.1.1 Dependent and independent BCIs

A BCI is a communication system in which messages or commands that an individual sends to the external world do not pass through the brain’s normal output pathways of peripheral nerves and muscles. For example, in an EEG-based BCI the messages are encoded in EEG activity. A BCI provides its user with an alternative method for acting on the world. BCIs fall into two classes: dependent and independent.

A dependent BCI does not use the brain’s normal output pathways to carry the message, but activity in these pathways is needed to generate the brain activity (e.g. EEG) that does carry it. For example, one dependent BCI presents the user with a matrix of letters that flash one at a time, and the user selects a specific letter by looking directly at it so that the visual evoked potential (VEP) recorded from the scalp over visual cortex when that letter flashes is much larger than the VEPs produced when other letters flash (Sutter, 1992). In this case, the brain’s output channel is EEG, but the generation of the EEG signal depends on gaze...
direction, and therefore on extraocular muscles and the cranial nerves that activate them. A dependent BCI is essentially an alternative method for detecting messages carried in the brain’s normal output pathways: in the present example, gaze direction is detected by monitoring EEG rather than by monitoring eye position directly. While a dependent BCI does not give the brain a new communication channel that is independent of conventional channels, it can still be useful (Sutter, 1992).

In contrast, an independent BCI does not depend in any way on the brain’s normal output pathways. The message is not carried by peripheral nerves and muscles, and, furthermore, activity in these pathways is not needed to generate the brain activity (e.g. EEG) that does carry the message. For example, one independent BCI presents the user with a matrix of letters that flash one at a time, and the user selects a specific letter by producing a P300 evoked potential when that letter flashes (Donchin et al., 2000). In this case, the brain’s output channel is EEG, and the generation of the EEG signal depends mainly on the user’s intent, not on the precise orientation of the eyes (Polich, 2000). The normal output pathways of peripheral nerves and muscles do not have an essential role in the operation of an independent BCI. Because independent BCIs provide the brain with wholly new output pathways, they are of greater theoretical interest than dependent BCIs. Furthermore, for people with the most severe neuromuscular disabilities, who may lack all normal output channels (including extraocular muscle control), independent BCIs are likely to be more useful.

1.1.2 BCI use is a skill

Most popular and many scientific speculations about BCIs start from the ‘mind-reading’ or ‘wire-tapping’ analogy, the assumption that the goal is simply to listen in on brain activity as reflected in electrophysiological signals and thereby determine a person’s wishes. This analogy ignores the essential and central fact of BCI development and operation. A BCI changes electrophysiological signals from mere reflections of central nervous system (CNS) activity into the intended products of that activity: messages and commands that act on the world. It changes a signal such as an EEG rhythm or a neuronal firing rate from a reflection of brain function into the end product of that function: an output that, like output in conventional neuromuscular channels, accomplishes the person’s intent. A BCI replaces nerves and muscles and the movements they produce with electrophysiological signals and the hardware and software that translate those signals into actions.

The brain’s normal neuromuscular output channels depend for their successful operation on feedback. Both standard outputs such as speaking or walking and more specialized outputs such as singing or dancing require for their initial acquisition and subsequent maintenance continual adjustments based on oversight of intermediate and final outcomes (Ghez et al., 2000). When feedback is absent from the start, motor skills do not develop properly; and when feedback is lost later on, skills deteriorate.

As a replacement for the brain’s normal neuromuscular output channels, a BCI also depends on feedback and on adaptation of brain activity based on that feedback. Thus, a BCI system must provide feedback and must interact in a productive fashion with the adaptations the brain makes in response to that feedback. This means that BCI operation depends on the interaction of two adaptive controllers: the user’s brain, which produces the signals measured by the BCI; and the BCI itself, which translates these signals into specific commands.
Successful BCI operation requires that the user develop and maintain a new skill, a skill that consists not of proper muscle control but rather of proper control of specific electrophysiological signals; and it also requires that the BCI translate that control into output that accomplishes the user’s intent. This requirement can be expected to remain even when the skill does not require initial training. In the independent BCI described above, the P300 generated in response to the desired letter occurs without training. Nevertheless, once this P300 is engaged as a communication channel, it is likely to undergo adaptive modification (Coles et al., 1995), and the recognition and productive engagement of this adaptation will be important for continued successful BCI operation.

That the brain’s adaptive capacities extend to control of various electrophysiological signal features was initially suggested by studies exploring therapeutic applications of the EEG. They reported conditioning of the visual alpha rhythm, slow potentials, the mu rhythm, and other EEG features (Neidermeyer., 1999). These studies usually sought to produce an increase in the amplitude of a specific EEG feature. Because they had therapeutic goals, such as reduction in seizure frequency, they did not try to demonstrate rapid bidirectional control, that is, the ability to increase and decrease a specific feature quickly and accurately, which is important for communication. Nevertheless, they suggested that bidirectional control is possible, and thus justified and encouraged efforts to develop EEG-based communication. In addition, studies in monkeys showed that the firing rates of individual cortical neurons could be operantly conditioned, and thus suggested that cortical neuronal activity provides another option for non-muscular communication and control (Schmidt., 1980).

At the same time, these studies did not indicate to what extent the control that people or animals develop over these electrophysiological phenomena depends on activity in conventional neuromuscular output channels (Dewan., 1967). While studies indicated that conditioning of hippocampal activity did not require mediation by motor responses (Black., 1971), the issue was not resolved for other EEG features or for cortical neuronal activity.

1.1.3 The parts of a BCI

Like any communication or control system, a BCI has input (e.g. electrophysiological activity from the user), output (i.e. device commands), components that translate input into output, and a protocol that determines the onset, offset, and timing of operation. Fig. 1 shows these elements and their principal interactions.

1.1.3.1 Signal acquisition

In the BCIs discussed here, the input is EEG recorded from the scalp or the surface of the brain or neuronal activity recorded within the brain. Thus, in addition to the fundamental distinction between dependent and independent BCIs electrophysiological BCIs can be categorized by whether they use non-invasive (e.g. EEG) or invasive (e.g. intracortical) methodology. They can also be categorized by whether they use evoked or spontaneous inputs. Evoked inputs (e.g. EEG produced by flashing letters) result from stereotyped sensory stimulation provided by the BCI. Spontaneous inputs (e.g. EEG rhythms over sensorimotor cortex) do not depend for their generation on such stimulation. There is, presumably, no reason why a BCI could not combine non-invasive and invasive methods or evoked and spontaneous inputs. In the signal-acquisition part of BCI operation, the chosen input is acquired by the recording electrodes, amplified, and digitized.
1.1.3.2 Feature extraction

The digitized signals are then subjected to one or more of a variety of feature extraction procedures, such as spatial filtering, voltage amplitude measurements, spectral analyses, or single-neuron separation. This analysis extracts the signal features that (hopefully) encode the user’s messages or commands. BCIs can use signal features that are in the time domain (e.g. evoked potential amplitudes or neuronal firing rates) or the frequency domain (e.g. mu or beta-rhythm amplitudes) (McFarland et al., 2000). A BCI could conceivably use both timedomain and frequency-domain signal features, and might thereby improve performance (Schalk et al., 2000).

In general, the signal features used in present-day BCIs reflect identifiable brain events like the firing of a specific cortical neuron or the synchronized and rhythmic synaptic activation in sensorimotor cortex that produces a mu rhythm. Knowledge of these events can help guide BCI development. The location, size, and function of the cortical area generating a rhythm or an evoked potential can indicate how it should be recorded, how users might best learn to control its amplitude, and how to recognize and eliminate the effects of non-CNS artifacts. It is also possible for a BCI to use signal features, like sets of autoregressive parameters, that correlate with the user’s intent but do not necessarily reflect specific brain events. In such cases, it is particularly important (and may be more difficult) to ensure that the chosen features are not contaminated by EMG, electrooculography (EOG), or other non-CNS artifacts.

1.1.3.3 The translation algorithm

The first part of signal processing simply extracts specific signal features. The next stage, the translation algorithm, translates these signal features into device commands orders that carry out the user’s intent. This algorithm might use linear methods (e.g. classical statistical analyses (Jain et al., 2000) or nonlinear methods (e.g. neural networks). Whatever its nature, each algorithm changes independent variables (i.e. signal features) into dependent variables (i.e. device control commands).

Effective algorithms adapt to each user on 3 levels. First, when a new user first accesses the BCI the algorithm adapts to that user’s signal features. If the signal feature is mu-rhythm amplitude, the algorithm adjusts to the user’s range of mu-rhythm amplitudes; if the feature is P300 amplitude, it adjusts to the user’s characteristic P300 amplitude; and if the feature is the firing rate of a single cortical neuron, it adjusts to the neuron’s characteristic range of firing rates. A BCI that possesses only this first level of adaptation, i.e. that adjusts to the user initially and never again, will continue to be effective only if the user’s performance is very stable. However, EEG and other electrophysiological signals typically display short- and long-term variations linked to time of day, hormonal levels, immediate environment, recent events, fatigue, illness, and other factors. Thus, effective BCIs need a second level of adaptation: periodic online adjustments to reduce the impact of such spontaneous variations. A good translation algorithm will adjust to these variations so as to match as closely as possible the user’s current range of signal feature values to the available range of device command values.

While they are clearly important, neither of these first two levels of adaptation addresses the central fact of effective BCI operation: its dependence on the effective interaction of two adaptive controllers, the BCI and the user’s brain. The third level of adaptation accommodates and engages the adaptive capacities of the brain. As discussed in Previous Sections, when an
electrophysiological signal feature that is normally merely a reflection of brain function becomes the end product of that function, that is, when it becomes an output that carries the user’s intent to the outside world, it engages the adaptive capacities of the brain. Like activity in the brain’s conventional neuromuscular communication and control channels, BCI signal features will be affected by the device commands they are translated into; the results of BCI operation will affect future BCI input. In the most desirable (and hopefully typical) case, the brain will modify signal features so as to improve BCI operation. If, for example, the feature is mu-rhythm amplitude, the correlation between that amplitude and the user’s intent will hopefully increase over time. An algorithm that incorporates the third level of adaptation could respond to this increase by rewarding the user with faster communication. It would thereby recognize and encourage the user’s development of greater skill in this new form of communication. On the other hand, excessive or inappropriate adaptation could impair performance or discourage further skill development. Proper design of this third level of adaptation is likely to prove crucial for BCI development. Because this level involves the interaction of two adaptive controllers, the user’s brain and the BCI system, its design is among the most difficult problems confronting BCI research.

1.1.3.4 The output device

For most current BCIs, the output device is a computer screen and the output is the selection of targets, letters, or icons presented on it (Wolpaw et al, 1991). Selection is indicated in various ways (e.g. the letter flashes). Some BCIs also provide additional, interim output, such as cursor movement toward the item prior to its selection (Pfurtscheller et al, 2000). In addition to being the intended product of BCI operation, this output is the feedback that the brain uses to maintain and improve the accuracy and speed of communication. Initial studies are also exploring BCI control of a neuroprosthesis or orthosis that provides hand closure to people with cervical spinal cord injuries (Lauer et al, 2000). In this prospective BCI application, the output device is the user’s own hand.

1.1.3.5 The operating protocol

Each BCI has a protocol that guides its operation. This protocol defines how the system is turned on and off, whether communication is continuous or discontinuous, whether message transmission is triggered by the system (e.g. by the stimulus that evokes a P300) or by the user, the sequence and speed of interactions between user and system, and what feedback is provided to the user.

Most protocols used in BCI research are not completely suitable for BCI applications that serve the needs of people with disabilities. Most laboratory BCIs do not give the user on/off control: the investigator turns the system on and off. Because they need to measure communication speed and accuracy, laboratory BCIs usually tell their users what messages or commands to send. In real life the user picks the message. Such differences in protocol can complicate the transition from research to application.

2. Our BCI system: Brain computer interface with wavelets genetic algorithms

2.1 Materials and methods

In this research, EEG signal used as the basic data for classification. The EEG data is from an open EEG database of University of Tuebingen. Two types of the EEG database are employed as (BCI Competition, 2003).
2.1.1 Dataset I

The datasets were taken from a healthy subject. The subject was asked to move a cursor up and down on a computer screen, while his cortical potentials were taken. During the recording, the subject received visual feedback of his slow cortical potentials (Cz-Mastoids). Each trial lasted 6s. During every trial, the task was visually presented by a highlighted goal at either the top or bottom of the screen to indicate negativity or positivity from second 0.5 until the end of the trial. The visual feedback was presented from second 2 to second 5.5. Only this 3.5 second interval of every trial is provided for training and testing. The sampling rate of 256 Hz and the recording length of 3.5s results in 896 samples per channel for every trial. This dataset contain 266 trials that 70% of this dataset is considered as train dataset and the rest are considered as test.

2.1.2 Dataset II

The datasets were taken from an artificially respirated ALS patient. The subject was asked to move a cursor up and down on a computer screen, while his cortical potentials were taken. During the recording, the subject received auditory and visual feedback of his slow cortical potentials (Cz-Mastoids). Each trial lasted 8s. During every trial, the task was visually and auditorily presented by a highlighted goal at the top or bottom of the screen from second 0.5 until second 7.5 of every trial. In addition, the task ("up" or "down") was vocalised at second 0.5. The visual feedback was presented from second 2 to second 6.5. Only this 4.5 second interval of every trial is provided for training and testing. The sampling rate of 256 Hz and the recording length of 4.5s results in 1152 samples per channel for every trial. This dataset contain 200 trials that 70% of this dataset is considered as train dataset and the rest are considered as test.

2.1.3 Proposed methods

The block diagram of the proposed method for EEG signal classification is depicted in Fig.1. The method is divided into five steps: (1) EEG acquisition and sampling, (2) EEG preprocessing, (3) calculation of feature vector, (4) feature selection, (5) classification (Goncharova et al, 2003; Hyvärinen & Oja, 2000; McFarland et al, 2005).

Fig. 1. Block diagram of the proposed method for EEG signal classification.
3. Artifacts in BCI systems

Artifacts are undesirable potentials that contaminate brain signals, and are mostly of non-cerebral origin. Unfortunately, they can modify the shape of a neurological phenomenon used to drive a BCI system. Thus, even cerebral potentials may sometimes be considered as artifacts. For example, in an MRP-based BCI system, a visual evoked potential (VEP) is considered as an artifact. Visual alpha rhythms can also appear as artifacts in a Mu-based BCI system (Goncharova et al, 2003). One problem with such artifacts is that they could mistakenly result in controlling the device (McFarland et al, 2005). Therefore, there is a need to avoid, reject or remove artifacts from recordings of brain signals.

Artifacts originate from non-physiological as well as physiological sources. Non-physiological artifacts originate from outside the human body (such as 50/60 Hz power-line noise or changes in electrode impedances), and are usually avoided by proper filtering, shielding, etc.

Physiological artifacts arise from a variety of bodily activities. Electrocardiography (ECG) artifacts are caused by heart beats and may introduce a rhythmic activity into the EEG signal. Respiration can also cause artifacts by introducing a rhythmic activity that is synchronized with the body’s respiratory movements. Skin responses such as sweating may alter the impedance of electrodes and cause artifacts in the EEG signals.

Physiological artifacts such as ocular (EOG) and muscle (EMG) artifacts are much more challenging to handle than non-physiological ones. Moreover, controlling them during signal acquisition is not easy. There are different ways of handling these types of artifacts in BCI systems. In next Section, we examine the methods for handling Physiological artifacts in BCI systems.

3.1 Methods of handling artifacts

In this section, we briefly address methods of handling artifacts. Our focus throughout this section will be on Artifact removal in BCI systems.

3.1.1 Artifact avoidance

The first step in handling artifacts is to avoid their occurrence by issuing proper instructions to users. For example, users are instructed to avoid blinking or moving their body during the experiments. Instructing users to avoid generating artifacts during data collection has the advantage of being the least computationally demanding among the artifact handling methods, since it is assumed that no artifact is present in the signal (or that the presence of artifacts is minimal). However, it has several drawbacks. First, since many physiological signals, such as the heart beats, are involuntary, artifacts will always be present in brain signals. Even in the case of EOG and EMG activities, it is not easy to control eye and body movements during data recording. Second, the occurrence of ocular and muscle activity during an online operation of any BCI system is unavoidable. Third, the collection of a sufficient amount of data without artifacts may be difficult, especially in cases where a subject has a neurological disability. Finally, avoiding artifacts may introduce an additional cognitive task for the subject. For example, it has been shown that refraining from eye blinking results in changes in the amplitude of some evoked.
3.1.2 Artifact rejection

Artifact rejection refers to the process of rejecting the trials affected by artifacts. It is perhaps the simplest way of dealing with brain signals contaminated with artifacts. It has some important advantages over the artifact avoidance approach. For example, it would be easier for users to participate in the experiments and perform the required tasks, especially those individuals with motor disabilities. Also, the “secondary” cognitive task, resulting from a subject trying to avoid generating a particular artifact, will not be present in the EEG signal. Artifact rejection is usually done by visually inspecting the EEG or the artifact signals, or by using an automatic detection method.

3.1.3 Artifact removal

Artifact removal is the process of identifying and removing artifacts from brain signals. An artifact-removal method should be able to remove the artifacts as well as keeping the related neurological phenomenon intact. In this paper we use artifact removal methods that contain ICA and linear filtering.

3.1.3.1 Linear filtering

Linear filtering is useful for removing artifacts located in certain frequency bands that do not overlap with those of the neurological phenomena of interest (Barlow, 1984). For example, low-pass filtering can be used to remove EMG artifacts and high-pass filtering can be used to remove EOG artifacts. Linear filtering was commonly used in early clinical studies to remove artifacts in EEG signals (Zhou, 2005).

The advantage of using filtering is its simplicity. Also the information from the EOG signal is not needed to remove the artifacts. This method, however, fails when the neurological phenomenon of interest and the EMG or EOG artifacts overlap or lie in the same frequency band (de Beer et al., 1995). A look at the frequency range of neurological phenomena used in BCI systems unfortunately shows that this is usually the case. As a result, a simple filtering approach cannot remove EMG or EOG artifacts without removing a portion of the neurological phenomenon. More specifically, since EOG artifacts generally consist of low frequency components, using a high-pass filter will remove most of the artifacts. Such methods are successful to some extent in BCI systems that use features extracted from high-frequency components of the EEG (e.g., Mu and Beta rhythms). However, for BCI systems that depend on low frequency neurological phenomena (such as MRPs), this methods are not as desirable, since these neurological phenomena may lie in the same frequency range as that of the EOG artifacts.

In the case of removing EMG artifacts from EEG signals, filtering specific frequency bands of the EEG can be used to reduce the EMG activity. Since artifacts generated by EMG activity generally consist of high-frequency components, using a low-pass filter may remove most of these artifacts. Again, such methods may be successful to some extent for BCI systems that rely on low-frequency components (e.g., MRPs), but they cannot be effective for BCI systems that use a neurological phenomenon with high-frequency content (such as Beta rhythms).

3.1.3.2 Blind source separation (BSS)

BSS techniques separate the EEG signals into components that “build” them. They identify the components that are attributed to artifacts and reconstruct the EEG signal without these...
Brain Computer Interface with Wavelets and Genetic Algorithms

components (Choi et al, 2005). Among the BSS methods, Independent Component Analysis (ICA) is more widely used. ICA is a method that blindly separates mixtures of independent source signals, forcing the components to be independent. It has been widely applied to remove ocular artifacts from EEG signals (Jung et al, 2001). Preliminary studies have shown that ICA increases the strength of motor-related signal components in the Mu rhythms, and is thus useful for removing artifacts in BCI systems (Makeig et al, 2000).

One advantage of using BSS methods such as ICA is that they do not rely on the availability of reference artifacts for separating the artifacts from the EOG signals (Zhou, 2005). One disadvantage of ICA, along with other BSS techniques, is that they usually need prior visual inspection to identify artifact components (Jung et al, 2001). However, some automatic methods have been proposed (Joyce et al, 2004).

3.1.3.3 Independent component analysis

ICA was originally developed for blind source separation whose goal is to recover mutually independent but unknown source signals from their linear mixtures without knowing the mixing coefficients.

ICA is a computational technique for revealing hidden factors that underlie sets of measurements or signals. ICA assumes a statistical model whereby the observed multivariate data, typically given as a large database of samples, are assumed to be linear or nonlinear mixtures of some unknown latent variables. The mixing coefficients are also unknown. The latent variables are nongaussian and mutually independent and they are called the independent components of the observed data. By ICA, these independent components, also called sources or factors, can be found. Thus ICA can be seen as an extension to Principal Component Analysis and Factor Analysis. ICA is a much richer technique, however, capable of finding the sources when these classical methods fail completely.

In this chapter, we use a basic form of the FastICA algorithm as follows (Hyvärinen & Oja, 2000):

\[1. \text{Choose an initial (e.g random) weight vector } w.\]
\[2. \text{let } w^* = E\{xg(w^T x)\} - \{g'(w^T x)\} w\]
\[3. w = w^*/\|w^*\| \quad (1)\]
\[4. \text{If not converged, go back to 2.}\]

Where \(g = u \exp(-u^2/2)\), x observed data and w is a weight matrix that does ICA. Note that convergence means that the old and new values of w point in the same direction, i.e. their dot-product are (almost) equal to 1.

4. Feature extraction

For features extraction from the raw EEG data many methods such as time domain, frequency domain, and time–frequency domain are used. In this article we used Hilbert and Wavelet Transform for feature extraction.
4.1 Wavelet transforms

4.1.1 Wavelets

Wavelets are a recently developed signal processing tool enabling the analysis on several timescales of the local properties of complex signals that can present non-stationary zones. They lead to a huge number of applications in various fields, such as, for example, geophysics, astrophysics, telecommunications, imagery and video coding. They are the foundation for new techniques of signal analysis and synthesis and find beautiful applications to general problems such as compression and denoising.

The propagation of wavelets in the scientific community, academic as well as industrial, is surprising. First of all, it is linked to their capacity to constitute a tool adapted to a very broad spectrum of theoretical as well as practical questions. Let us try to make an analogy: the emergence of wavelets could become as important as that of Fourier analysis. A second element has to be noted: wavelets have benefited from an undoubtedly unprecedented trend in the history of applied mathematics. Indeed, very soon after the grounds of the mathematical theory had been laid in the middle of the 1980s (Meyer, 1990), the fast algorithm and the connection with signal processing (Mallat, 1989) appeared at the same time as Daubechies orthogonal wavelets (Daubechies, 1988). This body of knowledge, diffused through the Internet and relayed by the dynamism of the research community enabled a fast development in numerous applied mathematics domains, but also in vast fields of application.

Thus, in less than 20 years, wavelets have essentially been imposed as a fruitful mathematical theory and a tool for signal and image processing. They now therefore form part of the curriculum of many pure and applied mathematics courses, in universities as well as in engineering schools.

4.1.2 Wavelet transforms

For features extraction from the raw EEG data many methods such as time domain, frequency domain, and time–frequency domain are used. Since the EEG is non-stationary in general, it is most appropriate to use time–frequency domain methods like wavelet transform (WT) as a mean for feature extraction (Nazari et al, 2009). The WT provides a more flexible way of time–frequency representation of a signal by allowing the use of variable sized windows. In WT long time windows are used to get a finer low-frequency resolution and short time windows are used to get high-frequency information. Thus, WT gives precise frequency information at low frequencies and precise time information at high frequencies. This makes the WT suitable for the analysis of irregular data patterns, such as impulses occurring at various time instances. The EEG recordings were decomposed into various frequency bands through fourth-level wavelet packet decomposition (WPD). The decomposition filters are usually constructed from the Daubechies or other sharp mother wavelets, when the data has discontinuities. In this research, based on the analysis of the data, Daubechies mother wavelet was used in the decomposition. The power spectrum, variance and mean of the signal (each channel) are extracted as features. So the feature set for each subject in each trial consisted of $3^*\text{number of channels}$. As a result, the feature matrix was 266*18 and 200*21 for subject A and B respectively. Finally the feature matrix is normalized.
4.2 Hilbert transforms

Hilbert transforms are essential in understanding many modern modulation methods. These transforms effectively phase shift a function by 90 degrees independent of frequency. Of course practical implementations have limitations (Huang et al, 1998). For example, the phase shifting of a low frequency implies a long delay, which in turn implies a computational process that maintains a long history of the signal. Hilbert transforms are useful in creating signals with one sided Fourier transforms. Also the concepts of analytic functions and analytic signals will be shown to be related through Hilbert transforms.

The Hilbert transform is a convenient tool to use in dealing with band-pass signals (Huang, 2005). The Hilbert transform of a signal \( x(t) \) will be denoted by \( \hat{x}(t) \) and is obtained by passing \( x(t) \) through a filter with the transfer function:

\[
H(\omega) = -j \text{sign} \omega = \begin{cases} 
  -j & \text{for } \omega > 0 \\
  0 & \text{for } \omega = 0 \\
  j & \text{for } \omega < 0 
\end{cases}
\]

This is illustrated in “Fig. 2”. The Hilbert transform filter is an ideal 90° phase shifter. In the frequency domain:

\[
\hat{X}(\omega) = H(\omega)X(\omega) = -j \text{sign}(\omega)X(\omega)
\]

It can be shown that the inverse Fourier transform of \( H(\omega) \) is \( h(t) = \frac{1}{\pi t} \). Therefore, a signal and its Hilbert transform are related by the convolution integral:

\[
\hat{X} = x(t) * \frac{1}{\pi t} \approx \frac{1}{\pi} \int_{-\infty}^{\infty} x(\tau) \frac{1}{t-\tau} d\tau
\]

Where * represents convolution.

Fig. 2. System for Forming Hilbert Transforms.

5. Feature selection

Feature selection is one of the major tasks in classification problems. The main purpose of feature selection is to select a number of features used in the classification and at the same time to maintain acceptable classification accuracy. Besides deciding which types of features to use, the weighting of features also plays an important role in classification. Emphasizing
features that have better discriminative power will usually boost classification. Feature selection can be seen as a special case of feature weighting, in which features that are eliminated are assigned zero weight. Feature selection reduces the dimensionality of the feature space, which leads to a reduction in computational complexity. Furthermore, in some cases, classification can be more accurate in the reduced space. Various algorithms have been used for feature selection in the past decades. One of the best methods that can be used for features selection is Genetic Algorithms (Sheng, 2006).

5.1 Genetic algorithms

Genetic Algorithms are adaptive heuristic search algorithm premised on the evolutionary ideas of natural selection and genetic (Andries, 2007). The basic concept of Gas is designed to simulate processes in natural system necessary for evolution. The main operator of GA to search in pool of possible solutions is Crossover, Mutation and selection.

The genetic search process is iterative: evaluating, selection and recombining string in the population during each one of iterations (generation) until reaching some termination condition. Evaluation of each string is based on a fitness function that is problem-dependent. It determines which of the candidate solutions are better. This corresponds to the environmental determination of survivability in national selection. Selection of a string, which represents a point in the search space, depends on the string’s fitness relative to those of other strings in the population, those points that have relatively low fitness.

Mutation, as in natural systems, is a very low probability operator and just flips bit. The aim of mutation is to introduce new genetic material into an existing individual; that is, to add diversity to the genetic characteristics of the population. Mutation is used in support of crossover to ensure that the full range of allele is accessible for each gene.

Crossover in contrast is applied with high probability. It is a randomized yet structured operator that allows information exchange between points. Its goal is to preserve the fittest individual without introducing any new value.

The proposed approach to the use of Gas for Feature selection involves encoding a set of d, Feature s as a binary string of d elements, in which a 0 in the string indicates that the corresponding Feature is to be omitted, and a 1 that it is to be included. This coding scheme represents the presence or absence of a particular Feature from the Feature space (see Fig. 3). The length of chromosome equal to Feature space dimensions.

![Fig. 3. Schema of the proposed GA-based feature selection approach](www.intechopen.com)
6. Translation algorithm

6.1 Neural Networks

An artificial neural network (ANN) is an interconnected group of artificial neurons simulating the thinking process of human brain. One can consider an ANN as a “magical” black box trained to achieve expected intelligent process, against the input and output information stream. ANN are useful in application areas such as pattern recognition, classification etc.

6.1.1 Multilayered Perceptron Neural Networks

The decision making process of the ANN is holistic, based on the features of input patterns, and is suitable for classification of biomedical data. Typically, multilayer feed forward neural networks can be trained as non-linear classifiers using the generalized back-propagation (BP) algorithm.

Our network has one hidden layer with five neurons and output layer with one neuron. Generalized BP algorithm with momentum used as training procedure. Momentum is a standard training technique which is used to speed up convergence and maintain generalization performance (Hagan, 1995). For hidden and output layers, we used bipolar and unipolar sigmoid functions respectively as decision function on the other hand we normalized weights and inputs. With these methods we achieved a NN classifier that is the most suitable classifier for the task at hand. We determined the most effective set as well as the optimum vector length for high accuracy classification. This NN classifier was trained and tested by using the feature sets described above.

By means of minimizing error optimized the number of neurons in hidden layer to five with tansig functions and sigmoid function for output layer.

6.1.2 Probabilistic Neural Network

The probabilistic approach to neural networks has been developed in the framework of statistical pattern recognition. Probabilistic neural network (PNN) is derived from radial basis function (RBF) network which is an ANN using RBF. RBF is a bell shape function that scales the variable nonlinearly. PNN is adopted for it has many advantages (Kim & Chang, 2007). Its training speed is many times faster than a BP network. PNN can approach a Bayes optimal result under certain easily met conditions. Additionally, it is robust to noise examples. We choose it also for its simple structure and training manner. The most important advantage of PNN is that training is easy and instantaneous. Weights are not “trained” but assigned. Existing weights will never be alternated but only new vectors are inserted into weight matrices when training. So it can be used in real-time. Since the training and running procedure can be implemented by matrix manipulation, the speed of PNN is very fast.

6.2 Support Vector Machine

The SVM is a relatively new classification technique developed by Vapnik (Avidan, 2004) which has shown to perform strongly in a number of real-world problems, including BCI.
The invention of SVM was driven by underlying statistical learning theory, i.e., following the principle of structural risk minimization that is rooted in VC dimension theory, which makes its derivation even more profound. The SVMs have been a topic of extensive research with wide applications in machine learning and engineering. The output of a binary SVM classifier can be computed by the following expression:

\[ y = \text{sgn} \left( \sum_{i=1}^{N} \alpha_i y_i k(x_i, x) + b \right) \]  

(5)

where \( \{x_i, y_i\}_{i=1}^{N} \) are training samples with input vectors \( x_i \in \mathbb{R}^d \), and class labels \( y_i = \{-1, 1\}, \alpha_i \geq 0 \), are Lagrangian multipliers obtained by solving a quadratic optimization problem, \( b \) is the bias, and \( k(x_i, x_j) \) is called kernel function in SVM. The most commonly used kernel function is the Gaussian RBF as:

\[ k(x_i, x_j) = \exp \left( -\frac{||x_i - x_j||^2}{2\sigma^2} \right) \]  

(6)

The SVM for the linearly separable case find optimal separating hyper plane, as shown in “Fig. 4” (Chandaka et al, 2008).

**Fig. 4.** SVM classification with a hyperplane that maximizes the separating margin between the two classes (indicated by data points marked by “X”s and “O”s). Support vectors are elements of the training set that lie on the boundary hyperplanes of the two classes.

### 7. Simulation results

In this paper, we proposed a scheme to combine liner filtering, Genetic Algorithm and neural network classifiers for EEG signal classification. Liner filtering and independent component analysis is used to artifact removal from EEG signals. The GA select essential EEG features then selected features serve as input feature vector for the following classifiers. Two neural networks, including probabilistic neural network (PNN), Multilayered...
Perceptron (MLP) and support vector machine (SVM) were employed in the study and their effects were compared. In neural network structure, the output layer unit has sigmoid function, which makes network capable of nonlinearly mapping and capturing dynamics of signals. In SVM classifier different values for $\sigma$ which is a very essential parameter in designing a SVM classifier with Gaussian RBF kernel examined and the best one selected.

To classify cursor movements two types of the EEG database are used, 70% of each dataset used for training and the rest for test classifiers. Both neural network classifiers and SVM demonstrated high classification accuracies with relatively small number of features. Between the three classifiers, SVM shows slightly better performance than MLP and PNN in terms of classification accuracy and robustness to different number of features. The results prove that the proposed scheme a promising model for the discrimination of clinical EEG signals. The performance of a classifier is not just measured as the accuracy achieved by the network, but aspects such as computational complexity and convergence characteristics are just as important. To reduce complexity, the GA used to select essential EEG features. This approach to BCI helps to reduce the computational complexity of the Classification process, and helps to improve transfer rate in real-time BCI systems.

Generally, the classification accuracy over files, which were included in training, is higher than the accuracy for the testing set. Tables I and II indicate the results of classification accuracy during training and test stages for both datasets. In comparison with the neural network classifier, SVM has a better training and test accuracy rate of neural network classifier, because of the nature of SVM classifier, this classifier is more general than neural network and this specification is very important in the use of classifiers. The most important advantage of PNN is that training is easy and instantaneous in comparison with SVM and MLP classifiers.

<table>
<thead>
<tr>
<th>Features</th>
<th>Wavelet transform</th>
<th>Hilbert transform</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training</td>
<td>Test</td>
</tr>
<tr>
<td><strong>MLP</strong></td>
<td>99.56%</td>
<td>86.75%</td>
</tr>
<tr>
<td><strong>PNN</strong></td>
<td>99.98%</td>
<td>88.75%</td>
</tr>
<tr>
<td><strong>SVM</strong></td>
<td>99.95%</td>
<td>90.25%</td>
</tr>
</tbody>
</table>

Table 1. Results of the dataset type I

<table>
<thead>
<tr>
<th>Features</th>
<th>Wavelet transform</th>
<th>Hilbert transform</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training</td>
<td>Test</td>
</tr>
<tr>
<td><strong>MLP</strong></td>
<td>99.46%</td>
<td>87.75%</td>
</tr>
<tr>
<td><strong>PNN</strong></td>
<td>99.95%</td>
<td>88.55%</td>
</tr>
<tr>
<td><strong>SVM</strong></td>
<td>99.92%</td>
<td>90.05%</td>
</tr>
</tbody>
</table>

Table 2. Results of the Dataset type II

8. Conclusion

Wavelets are a recently developed signal processing tool enabling the analysis on several timescales of the local properties of complex signals that can present non-stationary zones. They lead to a huge number of applications in various fields, such as, geophysics, astrophysics, telecommunications, imagery and video coding. They are the foundation for
new techniques of signal analysis and synthesis and find beautiful applications to general problems such as compression and denoising. In order to extract the most suitable features from the raw EEG data different methods in time or frequency domain can be used. Since the EEG is non-stationary in general, it is most appropriate to use time–frequency domain methods like wavelet transform (WT) as a mean for feature extraction. The simulation results confirm this fact. The Genetic algorithm is applied in order to choose the best features from the feature space. GA is an evolutionary algorithm which its optimality has been proved in other fields, the computation complexity is low and it is an appropriate method in real time problems.

9. Acknowledgment

This study was supported by Islamic Azad University, Borujerd Branch, Iran. The authors would like to acknowledge staff of university.

10. References


Goncharova II, McFarland DJ, Vaughan TM, Wolpaw JR. "EMG contamination of EEG: 

Hyvärinen, A. and Oja, E, “Independent Component Analysis: Algorithms and 

Jain AK, Duin PW, Mao J. Statistical pattern recognition: a review. IEEE Trans Pattern Anal 

Joyce CA, Gorodnitsky IF, Kutas M. Automatic removal of eye movement and blink artifacts

Jung TP, Makeig S, Westerfield M, Townsend J, Courchesne E, Sejnowski TJ. Analysis and 

Lauer RT, Peckham PH, Kilgore KL, Heetderks WJ. Applications of cortical signals to 

M. Hagan, H. Demuth, and M. Beale, “Neural Network Design,” PWS Publishing Company,
1995.

M. R. Nazari Kousarrizi, A. Asadi Ghanbari, M. Teshnehab, M. Aliyari, A. Gharaviri, 
“Feature Extraction and Classification of EEG Signals using Wavelet Transform, 
SVM and Artificial Neural Networks for Brain Computer Interfaces,” IEEE International Joint Conferences on Bioinformatics, Systems Biology and Intelligent Computing, 2009.

Makeig S, Enghoff S, Jung TP, Sejnowski TJ. A natural basis for efficient brain-actuated 

Mallat S., “A theory for multiresolution signal decomposition: the wavelet representation”,

operation: signal and noise during early training sessions,” Clin Neurophysiol; vol. 

McFarland DJ, Sarnacki WA, Vaughan TM, Wolpaw JR. EEG-based brain–computer 
interface communication effect of target number and trial length on information 


“The empirical mode decomposition and Hilbert spectrum for nonlinear and 
1998.

N. Huang, “The Hilbert–Huang Transform in Engineering,”, CRC Press, Boca Raton, FL,
2005.

N. Xu, X. Gao, B. Hong, X. Miao, S. Gao, and F.Yang. BCI Competition 2003—Data set Iib: 
Enhancing P300 wave detection using ICA-based subspace projections for BCI 

Neidermeyer E. The normal EEG of the waking adult. In: Niedermeyer E, Lopes da Silva 
FH, editors. Electroencephalography: basic principles, clinical applications and 


This book reports on recent applications in biology and geoscience. Among them we mention the application of wavelet transforms in the treatment of EEG signals, the dimensionality reduction of the gait recognition framework, the biometric identification and verification. The book also contains applications of the wavelet transforms in the analysis of data collected from sport and breast cancer. The denoting procedure is analyzed within wavelet transform and applied on data coming from real world applications. The book ends with two important applications of the wavelet transforms in geoscience.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following: