University of Essex

Stages for Developing Control Systems using EMG and EEG Signals: A survey

TECHNICAL REPORT: CES-513
ISSN 1744-8050

Ericka Janet Rechy-Ramirez and Huosheng Hu

School of Computer Science and Electronic Engineering
University of Essex, United Kingdom

Email: ejrech@essex.ac.uk, hhu@essex.ac.uk

June 9, 2011

Abstract

Bio-signals such as EMG (Electromyography), EEG (Electroencephalography), EOG (Electrooculogram), ECG (Electrocardiogram) have been deployed recently to develop control systems for improving the quality of life of disabled and elderly people. This technical report aims to review the current deployment of these state of the art control systems and explain some challenge issues. In particular, the stages for developing EMG and EEG based control systems are categorized, namely data acquisition, data segmentation, feature extraction, classification, and controller. Some related Bio-control applications are outlined. Finally a brief conclusion is summarized.

Contents

1 Introduction

Bio-signals are physical quantities that vary with time[88], which can be used to control machines and systems. There are different types of bio-signals, including EMG (Electromyography signal: electrical activity generated during the contraction of a skeletal muscle), EEG (Electroencephalography signal: electrical activity of the brain recorded from the scalp), EOG (Electrooculogram signal: electrical activity of the movements of the eyeball "eye gaze position" recorded around the eyes) and ECG(Electrocardiogram signal: electrical activity of the heart).

Human Machine Interaction (HMI) is a discipline with the goal of designing a control system in which the human can communicate with the computer or other devices more naturally. It is necessary for the elderly and disabled people to use different ways of communication and interaction with machines. On the other hand an interface using EEG signals as a way of communication is called BCI (Brain Computer Interface). This document overviews two kinds of bio-signals, namely EMG and EEG.

According to [91, 93, 3], there are 4 stages in EMG or EEG based control systems: 1) data acquisition and data segmentation, 2) feature extraction, 3) classification and 4) controller, as shown in Figure 1. The stage 1 is also called signal conditioning and preprocessing, in which the signal is acquired from the human body and is filtered to reduce the noise produced by other electrical activities of the body or inappropriate contact of the sensors, namely artifact. At this stage the outcome is raw signal.

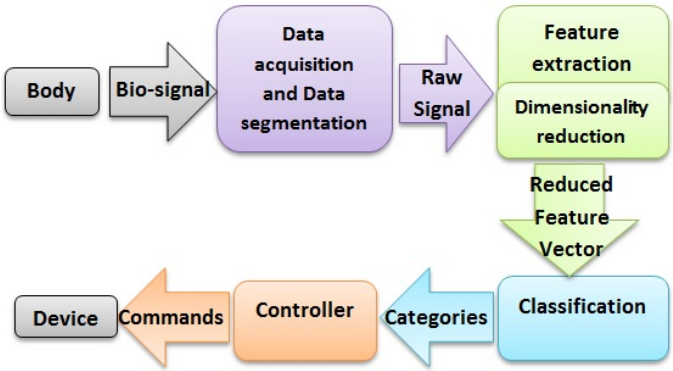

Figure 1: Stages for developing control systems using EMG or EEG signals

The feature extraction stage converts the raw signals obtained from the stage 1 into a feature vector. The feature vector represents relevant structure in the raw data. Dimensionality reduction eliminates redundant information in the feature vector, generating a reduced feature vector [93]. The Stage 3, classification, is also called as translation algorithm, involving pattern recognition, since a classification algorithm is applied to the reduced feature vector in order to obtain categories. Finally, the stage 4, controller, is to translate the categories obtained from Stage 3 to control commands for execution.

The rest of the survey is organized as follows. Section 2 presents the anatomy of the EMG and EEG signals. The stage of data acquisition and data segmentation are described in Section 3. Then Sections 4 and 5 explain the feature extraction and dimensionality reduction stage, and the control system classification stage respectively. Finally Section 6 outlines some applications of EMG and EEG signals based control systems.

2 Anatomical background in EMG and EEG signals

This section briefly introduces a general idea of how skeletal muscles and the brain work, which are the source of EMG and EEG signals respectively.

2.1 EMG: anatomy of the muscles

Reaz *et al.* [73] explain that muscles are composed of bundles of specialized cells capable of contraction and relaxation; they are in charge of producing motion, moving substance within the body, providing stabilization and generating heat. There are three types of muscle tissue depending on the basis of structure, contractile properties and control mechanism: a) skeletal muscle, b) smooth muscle and c) cardiac muscle [73, 80]. According to Sörnmo and Laguna [80], skeletal muscle is attached to the skeleton and facilitates movement and position of the body; smooth muscle is found within the intestines and position of the body, while the cardiac muscle is responsible for creating a heartbeat. EMG signal is acquired from the skeletal muscle.

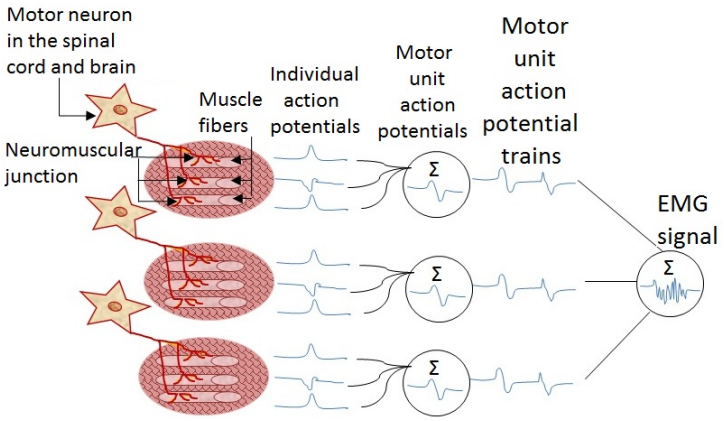

Figure 2: Participants in the skeletal muscle contraction producing EMG signal

In skeletal muscle, contraction is controlled by electrical impulses propagating between the central and peripheral nervous systems and the muscles. The contraction of a muscle fiber is initiated when neuronal action potentials reach the neuromuscular junction [1] and fire action potentials which spread along the excitable membranes of the muscle fiber. A Motor Unit Action Potential (MUAP) results from spatial and temporal summation of individual action potentials (electrical activity) as they spread through the different muscle fibers of a single motor unit [2] [57, 7].

According to Bida [8], a MUAP is the response of the motor unit to a single motor neuron excitation, while a MUAPT (motor unit action potential train) is a repetitive sequence of stimulations to the motor units; as a result of the motor unit responses to the impulse train is independent from the sequence and the total series response is random, the EMG signal is the superposition of the MUAPTs and can be treated as a stochastic process (see Figure 2).

2.2 EEG: anatomy of the brain

The brain consists of 10^{10}-10^{11} neurons that are very closely interconnected via axons and dendrites [61]. The neurons are in charge of communicating information to and from the brain. According to Malmivuo and Plonsey [55], the brain has 5 main parts: a) the interbrain (diencephalon), b) the midbrain, c) the pons Varolii and cerebellum, d) the medulla oblongata and e) the cerebrum, including the two cerebral hemispheres, as shown in Figure 3.

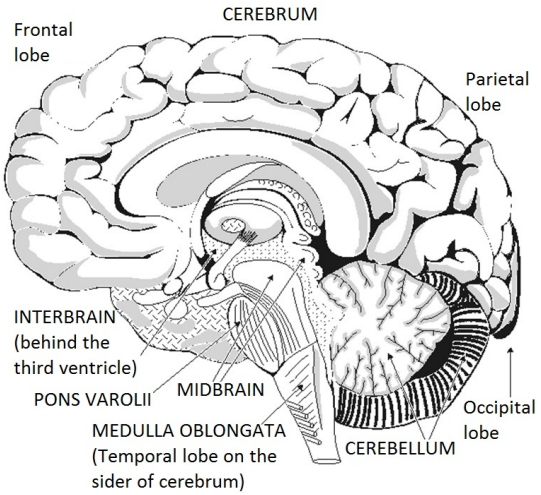

Figure 3: Anatomy of the brain [55]

Malmivuo and Plonsey [55] explain that the **interbrain** or diencephalon includes the thalamus and the hypothalamus. The Thalamus is a bridge connecting the sensory paths, while the hypothalamus is important for the regulation of the autonomic (involuntary) functions; together with the hypophysis, it regulates hormonal secretions. The **midbrain**

[1] Neuromuscular junction. Specialized synapse connecting each motor neuron to muscle fiber and allowing the action potentials to stimulate contraction

[2] Motor unit. Basic functional unit of a muscular contraction, it is the collection of the α-motor neuron in the spinal cord and the fibers, which it innervates[70]

is a small part of the brain. The **pons Varolii** is an interconnection of neural tracts; while the **cerebellum** controls fine movement. On the other hand, the **medulla oblongata** resembles the spinal cord to which it is immediately connected. Many reflex centers, such as the vasomotor center and the breathing center, are located in the medulla oblongata.

Finally, the **cerebrum**, which is the largest area of the brain, includes the higher cerebral functions, accurate sensations and the voluntary motor control of muscles. It consists of two hemispheres, the right and left cerebral hemispheres. The right cerebral hemisphere controls the left side of the body and the left cerebral hemisphere the right. The outer layer of the cerebrum, called the cerebral cortex, is made up of grey matter. The inner portion of the cerebrum is white matter. Grey matter is composed of nerve cells, these cells control brain activity; while white matter is composed of nerve cell axons that carry information between nerve cells in the brain and spinal cord. Deep indentations called fissures divide each hemisphere of the cerebrum into four lobes: a)frontal lobe, b) parietal lobe, c) temporal lobe and d)occipital lobe, Novák *et al.* [60].

Penfield and Rasmussen [69] state that in the cerebral cortex are located many different areas of specialized brain function; the higher brain functions occur in the frontal lobe, the visual center is located in the occipital lobe, and the sensory area and motor area are located on both sides of the central fissure. There are specific areas in the sensory and motor cortex whose elements correspond to certain parts of the body. The size of each such area is proportional to the required accuracy of sensory or motor control, see figure 4.

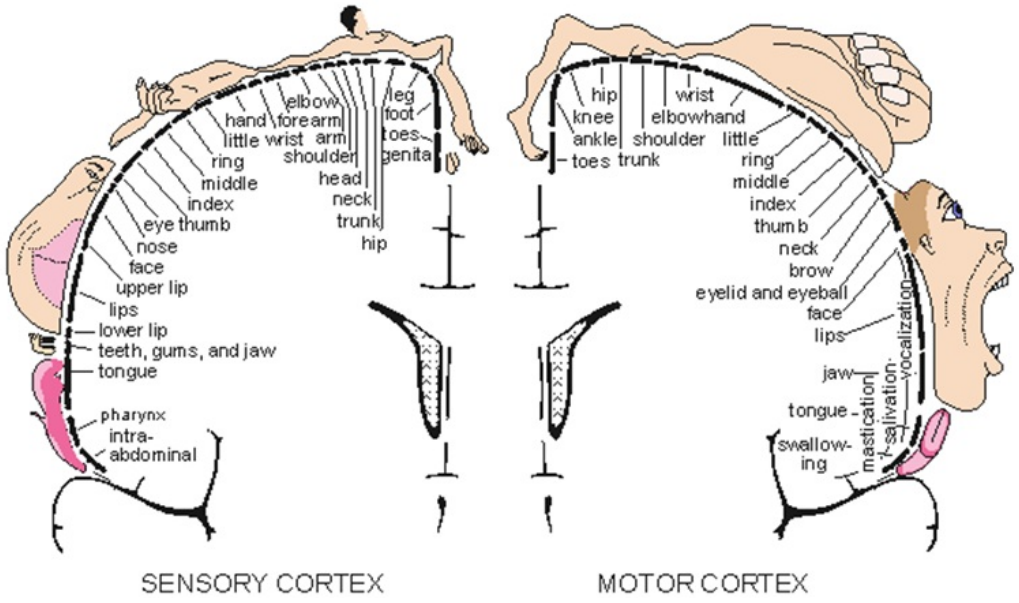

Figure 4: The division of sensory (left) and motor (right) functions in the cerebral cortex [69]

Based on Malmivuo and Plonsey [55], Sörnmo and Laguna [80], and Novák *et al.* [60], there are several waves that can be differentiated from the EEG signal as follows:

- **alpha** (α) **waves**, have the frequency spectrum of 8- 13 Hz; they can be measured from the occipital region in an awake person when the eyes are closed; alpha activity disappears normally with attention (e.g., mental arithmetic, stress, opening eyes);
- **beta** (β) **waves**, have the frequency band of 13 - 30 Hz, these are detectable over the parietal and frontal lobes, beta activity is generally regarded as a normal rhythm and is the dominant rhythm in patients who are alert or anxious or who have their eyes open;
- **delta** (δ) **waves**, have the frequency range of 0.5-4 Hz; they are detectable in infants and sleeping adults, and can be an indicative of cerebral damage or brain disease;
- **theta** (θ) **waves**, have the frequency range of 4-8 Hz; they are obtained from children and sleeping adults; the theta activity occurs during drowsiness and in certain stages of sleep, being abnormal in awake adults but is perfectly normal in children up to 13 years and in sleep; and
- **gamma** (γ) **waves**, have the frequency spectrum of upper 30 Hz, it is related to a state of active information processing of the cortex.

3 Stage 1: data acquisition and data segmentation

As described above, the first stage of a control system is data acquisition and data segmentation. In this stage, it is necessary to collect the data from the muscle for EMG signals or from the scalp for EEG signals. Once we have the data collected, the second part is to divide the signal into representative segments in order to extract features from each one.

3.1 EMG: data acquisition and data segmentation

3.1.1 EMG data acquisition

Sörnmo and Laguna [80] explain that myoelectric activities can be acquired:

- **invasively** by inserting a needle electrode through the skin directly into the muscle. This technique is a standard clinical tool used mainly for diagnostic purposes, since it provides a high-resolution, localized description of the muscle's electrical activity, albeit relatively painful for the patient; or
- **non-invasively** by placing a surface electrode on the skin overlying the muscle. The applications using this technique are discussed in section 6.1. The spatial resolution of this technique is more limited than that of the needle EMG, and the high frequency content of a MUAP is smoothed.

3.2 EMG data segmentation

According to Oskoei and Hu [66], there are two methods of EMG segmentation: disjoint and overlapped segmentation. In disjoint segmentation, separate segments with a predefined length are used for feature extraction, while in overlapped segmentation the new segment slides over the current segment, with an increment [3] time less than the segment length and more than the processing time. Therefore, disjoint segmentation is associated with segment length, while overlapped segmentation is associated with length and increment, as shown in Figure 5.

Christodoulou and Pattichis [17] employed a window with a constant length and a segmentation algorithm that calculated a threshold depending of the maximum value and the mean absolute value of the whole EMG signal. Peaks over the calculated threshold were considered as candidate segment. Gut and Moschytz [33] used a sliding time window to determine the beginning and the end of a segment. If the mean slope within this window exceeded a certain threshold, the beginning of an segment was detected; while the end of a segment was reached when the total variation of the EMG within the window fell below another threshold.

Oskoei and Hu [66] evaluated disjoint and overlapped segmentation by comparing classification performance over disjoint segments with a length of 50ms and overlapped segments, with those with a length of 200ms and an increment of 50ms. Their results showed that a disjoint segmentation with a length of 200ms provided high performance during EMG classification and a reasonable response time to allow real-time application, whereas overlapped segmentation with a length of 200 ms and an increment of 50ms shortened the response time without a noticeable degradation in accuracy.

Kaur *et al.* [48] analyzed three EMG segmentation techniques: 1) by identifying the peaks of the MUAPs, 2) by finding the beginning extraction point (BEP) and ending extraction point (EEP) of MUAPs, and 3) by using discrete wavelet transform (DWT). In the first technique, the EMG signal was segmented using an algorithm that detected areas of low activity and candidate MUAPs; the second technique identified the BEPs and EEPs of the possible MUAPs by sliding a window throughout the signal; and in the third technique, EMG signal was decomposed with the help of daubechies4 (db4) wavelet to detect MUAPs. In general, the first technique had the best performance with a total success rate of 95.90%, in comparison with the total success rates of 75.39% and 66.64% for the second and third techniques respectively.

3.3 EEG: data acquisition and data segmentation

3.3.1 EEG data acquisition

The most used recording technique for clinical EEG and for the study of event related potentials in non-clinical settings is the International 10/20 system; which is a standardized system for electrode placement proposed by Jasper [43]. This system employs 21 electrodes attached to the surface of the scalp at locations defined by certain anatomical reference points; the numbers 10 and 20 are percentage signifying relative distances between different electrode locations on the skull perimeter. The sampling rate for EEG signal acquisition is usually selected to be at least 200Hz [80].

[3] Increment. It is the time interval between two consecutive segments [66].

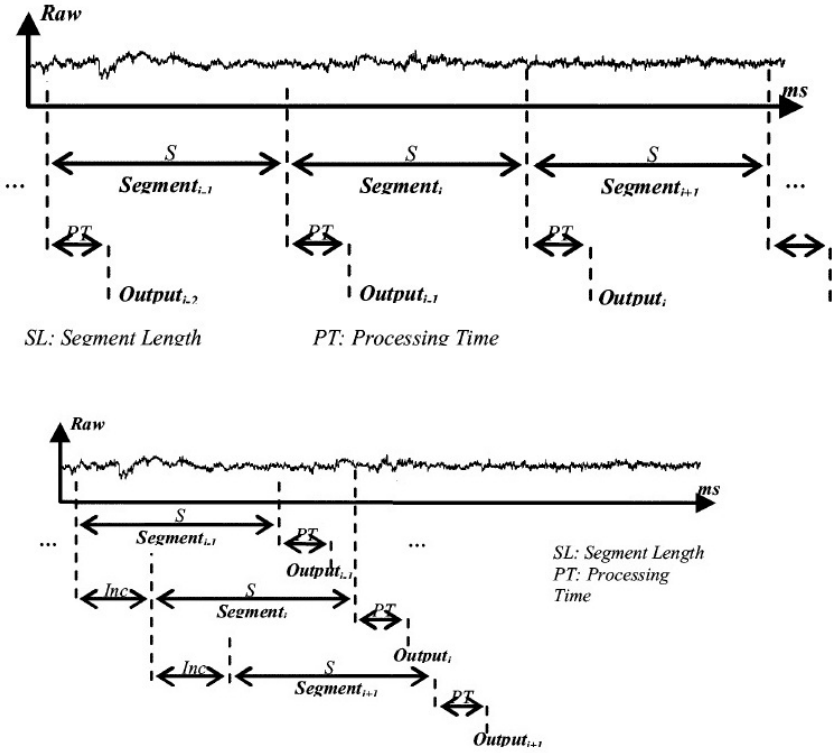

Figure 5: Disjoint (left) and overlapped (right) segmentation [66].

3.3.2 EEG data segmentation

Based on Biscay *et al.* [9], there are three methods to segment EEG recording into consecutive, stationary intervals: 1) adaptive segmentation, which is based on the detection of changes of an auto-regressive model that describes the data within a moving time window of fixed length; 2) a priori piece-wise segmentation followed by clustering; and 3) the syntactic approach, which incorporates grammatical rules to take into account the temporal contextual information for segmentation.

According to Kaplan *et al.* [45], there are two main approaches to divide the EEG signal into segments, a) fixed-interval segmentation, which separates the EEG recording into segments of the same length, and b) adaptive segmentation, which splits the EEG recording into quasy-stationarity segments of variable length. The adaptive segmentation methods can be classified as parametric methods and non-parametric methods.

Fixed-length segmentation

Kaplan *et al.* [45] explain that the fixed-length segmentation consists of four stages: 1) the EEG recording is divided preliminary into equal minimal ('elementary') segment lengths, 2) each segment is characterized by a certain set of features (e.g., spectral estimations or auto-regressive coefficients), 3) the main EEG segments are assigned to one of a number of classes accordingly to their characteristics, using one of the multivariate statistical procedures, and 4) the boundaries between the segments belonging to a same class are erased. Each of these stationary segments is characterized by its specific duration and typological features, but some EEG fragments contain transition processes and, are not strictly stationary, since this segmentation approach do not take into account the properties of the EEG recording.

Adaptive segmentation

Kaplan and Shishkin [46] state that the procedure of adaptive segmentation could be based on the estimation of the extent of similarity of an initial fixed interval of EEG with an EEG interval of the same duration viewed through the time window running along the EEG recording. The similarity index will drop sharply when the window runs over a segment boundary, giving a formal indication of the transition to the following segment.

Parametric methods

Parametric methods allow to describe adequately the piecewise stationary structure of the EEG signal. However, a drawback is that all these methods designed for the analysis of nonstationary processes are based on a procedure which may be applied only to stationary processes [45]. These methods make a good many assumptions, and rather stringent ones, about the nature of the population (in this case the EEG recording) from which the observations were drawn [77].

Parametric methods are effective if the phenomenological model of the process under study is known [37]. Nevertheless in the case of EEG, no phenomenological model is generally accepted and different mathematical models can be fitted to the same signal, resulting in different estimates of change-points [4]. The correctness of application of parametric methods for the EEG segmentation is therefore questionable [24]. Dvořák and Holden [23] established auto-regressive model (AR), autoregressive moving average (ARMA) and Kalman filter, as the most used parametric methods of EEG signal analysis.

Aufrichtigl *et al.* [4] examined AR model for segmenting EEG signals in four ways: 1) an AR-model estimated for the reference window and the signal in the moving window filtered with the corresponding inverse filter; 2) AR model estimated for the moving window, followed by an inverse filtering and calculation of test statistic for the reference window; 3) an asymptotic Gaussian distribution of the AR-parameters used to achieve a test statistic for the difference between the AR-parameters of the reference and moving windows; and 4) a calculated sum of two statistical tests, one corresponding to the difference between the AR-parameters of the reference and moving windows and the other one was the same difference, but inverted the order of the windows, both differences using an asymptotic Gaussian distribution of the AR-parameters.

Non-parametric methods

Non-parametric methods do not make numerous or stringent assumptions about the population (EEG recording), they do not need a priori information about probability distributions of random sequences [77, 45]. Brodsky *et al.* [24] proposed a non-parametric method for the segmentation of the EEG, the algorithm of change-point detection consisted of five steps: 1)construction of the diagnostic sequence [5] from an initial signal, 2) checking the homogeneity hypothesis, 3) preliminary estimation of change-points, 4) rejecting doubtful change-points, and 5) final estimation of change-points.

Other methods

Biscay *et al.* [9] proposed an EEG segmentation method that could incorporate the relevant information on the signal through parametric and non-parametric methods. The segmentation was performed according to an optimality criterion derived from the maximum a posteriori criterion and was not constrained by any predefined time scale of analysis, locating arbitrarily change points on the time axis.

4 Stage 2: feature extraction

As described above, the feature extraction stage involves the transformation of the raw signal to relevant data structure, called feature vector, by deleting noise and highlighting important data. Also, it could imply "dimensionality reduction", which eliminates redundant data from the feature vector, with the aim to facilitate the classification process.

4.1 EMG feature extraction

According to Zecca *et al.* [93], there are three types of features in EMG control systems: a) time domain, b) frequency domain and c) time-frequency domain.

4.1.1 Time domain

The time domain features are the most popular in EMG pattern recognition, because they are easy and quick to calculate, since they do not require a transformation. The time domain features are computed based on signal amplitude and resultant values give a measure of waveform amplitude, frequency and duration within some limited parameters [65].

1. **Integrated EMG (IEMG)**. This parameter is found by calculating the summation of the absolute values of EMG signals. It can be treated as a signal power estimator. It is defined as [39]:

$$IEMG_k = \sum_{i=1}^{N} |x_i|$$

where x_i is the value of each part of the segment k, and N is the length of the segment.

[4] A change point indicates a division in the EEG recording to produce a segment.
[5] Diagnostic sequence. A random sequence of detection of changes [24].

2. **Mean Absolute Value (MAV).** Estimates the mean absolute value of a signal MAV_k, by adding the absolute value of all of the values x_i in a segment k and dividing it by the length of the segment N [26].

$$MAV_k = \frac{1}{N}\sum_{i=1}^{N}|x_i|$$

3. **Modified Mean Absolute Value 1 (MMAV1).** An extension of MAV using weighting window function w_i [71].

$$MMAV1_k = \frac{1}{N}\sum_{i=1}^{N} w_i|x_i|$$

$$w(i) = \begin{cases} 1, & 0.25N \leq i \leq 0.75N \\ 0.5, & otherwise \end{cases}$$

4. **Modified Mean Absolute Value 2 (MMAV2).** Another extension of MAV, but here the weighting window function w_i is improved, because it is a continuous function [71].

$$MMAV2_k = \frac{1}{N}\sum_{i=1}^{N} w_i|x_i|$$

$$w(i) = \begin{cases} 1, & 0.25N \leq i \leq 0.75N \\ 4i/N, & 0.25N > i \\ 4(i-N)/N, & 0.75N < i \end{cases}$$

5. **Mean Absolute Value Slope (MAVS).** Estimates the difference between mean absolute values of the adjacent segments $k+1$ and k [71].

$$MAVS_k = MAV_{k+1} - MAVk$$

6. **Root Mean Square (RMS).** It is modeled as amplitude modulated Gaussian random process whose RMS is related to the constant force and non-fatiguing contraction [71].

$$RMS_k = \sqrt{\frac{1}{N}\sum_{i=1}^{N} {x_i}^2}$$

7. **Variance (VAR).** It is given by [65]:

$$VAR_k = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

where $\bar{x}$ is the mean value of the segment k.

8. **Waveform Length (WL).** It is the cumulative length of the waveform over the segment. The resultant values indicate a measure of waveform amplitude, frequency and duration all within a single parameter [26].

$$WL_k = \sum_{i=1}^{N-1}|x_{i+1} - x_i|$$

9. **Zero Crossings (ZC).** It is the number of times the waveform crosses zero, it is the number of times when the waveform changes its sign. A threshold ϵ must be included in the zero crossing calculation to reduce the noise induced zero crossings. Given two consecutive samples x_i and x_{i+1}, increment the zero crossing count [26], if

$$\{x_i > 0 \text{ and } x_{i+1} < 0\} \text{ or } \{x_i < 0 \text{ and } x_{i+1} > 0\} \text{ and } |x_i - x_{i+1}| \geq \epsilon$$

10. **Slope Sign Changes (SSC).** A feature that may provide another measure of frequency content is the number of times the slope changes sign. Again, a suitable threshold must be chosen to reduce noise induced slope sign changes. Given three consecutive samples, x_{i-1}, x_i and x_{i+1}, [26] the slope sign change is incremented if

$$\{x_i > x_{i-1} \text{ and } x_i > x_{i+1}\} \text{ or } \{x_i < x_{i-1} \text{ and } x_i < x_{i+1}\}$$
$$\text{and}$$
$$|x_i - x_{i+1}| \geq \epsilon \text{ or } |x_i - x_{i-1}| \geq \epsilon$$

11. **Willison Amplitude (WAMP).** Calculates the number of times that the absolute value of the difference between EMG signal amplitude of two consecutive samples (x_i and x_{i+1}) exceeds a predetermined threshold, ϵ [71].

$$WAMP_k = \sum_{i=1}^{N-1} f(|x_i - x_{i+1}|)$$

$$f(x) = \begin{cases} 1 & x > \epsilon \\ 0 & otherwise \end{cases}$$

12. **Simple Square Integral (SSI).** Uses the energy of the EMG signal as a feature [71].

$$SSI_k = \sum_{i=1}^{N} (|{x_i}^2|)$$

13. **Histogram of EMG (HEMG).** Divides the elements in EMG signal into b equally spaced segments and returns the number of elements in each segment [71].

4.1.2 Frequency domain

The frequency domain features are based on signal's estimated power spectrum density (PSD) and are computed by periodogram or parametric methods, but these features in comparison with time domain features require more computation and time to be calculated [65].

1. **Auto-Regressive coefficients (AR).** Describes each sample of EMG signal as a linear combination of previous samples plus a white noise error term. AR coefficients are used as features in EMG pattern recognition [71].

$$x_k = -\sum_{i=1}^{N} a_i x_{k-i} + e_k$$

where a_i is AR coefficients, e_k is white noise or error sequence, and N is the order of AR model.

2. **Frequency Median (FMD).** The frequency median splits the power spectrum density into two equal parts. It is given by [65]:

$$F_{MD} = \frac{1}{2} \sum_{i=1}^{M} PSD_i$$

where M is the length of the power spectrum density, and PSD_i is the i^{th} line of the power spectrum density.

3. **Frequency Mean (FMN).** The frequency mean is given by [65]:

$$F_{MN} = \frac{\sum_{i=1}^{M} f_i PSD_i}{\sum_{i=1}^{M} PSD_i}$$

where M is the length of the power spectrum density, $f_i = (i * sampling_{rate})/(2 * M)$, and PSD_i is the i^{th} line of the power spectrum density.

4. **Modified Frequency Median (MFMD).** This feature was proposed by Phinyomark *et al.* [71], it is the frequency at which the spectrum is divided into two regions with equal amplitude. It is given by:

$$MFMD = \frac{1}{2} \sum_{j=1}^{M} A_j$$

where A_j is the EMG amplitude spectrum at frequency bin j.

5. **Modified Frequency Mean (MFMN).** Also, this feature was proposed by Phinyomark *et al.* [71], it is the average of the frequency. MFMN is calculated as the sum of the product of the amplitude spectrum A_j and the frequency f_j, divided by the total sum of the spectral intensity, it is given by:

$$MFMN = \frac{\sum_{j=1}^{M} f_j A_j}{\sum_{j=1}^{M} A_j}$$

where f_j is the frequency of the spectrum at frequency bin j.

It is important to remark, as Phinyomark *et al.* [71] said, that traditional frequency median (FMD) and frequency mean (FMN) are calculated based on power spectrum density; the outline of amplitude spectrum and power spectrum are similar, but the amplitude value of amplitude spectrum is larger than the amplitude value of power spectrum.

6. **Frequency Ratio (FR).** It was proposed by Han *et al.* [35] in order to distinguish the difference between contraction and relaxation of a muscle in frequency domain, by applying fast Fourier transform to EMG signals in time domain.

$$FR_j = \frac{|F(\cdot)|_{jlowfreq}}{|F(\cdot)|_{jhighfreq}}$$

where $|F(\cdot)|_j$ is the fast Fourier transform of EMG signal in channel j, $lowfreq$ is the low frequency band, and $highfreq$ is the high frequency band. The threshold for dividing a low frequency band and a high frequency band is decided through experiments.

4.1.3 Time-Frequency domain

In his thesis [25], Englehart explains that time-frequency representation can localize the energy of the signal both in time and in frequency, allowing a more accurate description of the physical phenomenon; but these features generally requires a transformation that could be computationally heavy.

1. **Short Time Fourier Transform (STFT).** Gabor [30] extended the applicability of Fourier transform method by dividing the input signal into segments, by doing this the signal in each window can be assumed to be stationary. The STFT for a given signal $x(t)$ can be expressed as:

$$STFT_x(t, w) = \int W^*(\tau - t)x(\tau)e^{-jw\tau}d\tau$$

where $W(t)$ is the window function [6], $*$ is the complex conjugate, τ represents time, and w stands for frequency.

2. **Wavelet Transform (WT).** It is a transform where a signal is integrated with a shifted and scaled mother wavelet function. The continuous wavelet transform is represented by [88]:

$$W_x(a, b) = \int x(t)\left(\frac{1}{\sqrt{a}}\right)\Psi^*\left(\frac{t-b}{a}\right)dt$$

where $x(t)$ is the function representing the input signal, Ψ^* is the complex conjugate of the mother wavelet function, and $\Psi((t-b)/a)$ is the shifted and scaled version of the wavelet at time b and scale a.

3. **Wavelet Packet Transform (WPT).** WPT is a generalized version of the continuous wavelet transform and the discrete wavelet transform [93]. The basis for the WPT is chosen using an entropy-based cost function [19].

The main difference between STFT, WT and WPT is the way each one divides the time-frequency plane. The STFT has a fixed tiling, each cell has an identical aspect ratio; while the WT has a variable tiling and the aspect ratio of the cell varies in a form that the frequency resolution is proportional to the centre frequency. Finally, the WPT has an adaptive tiling, providing several alternatives of tilings [27], as shown in Figure 6.

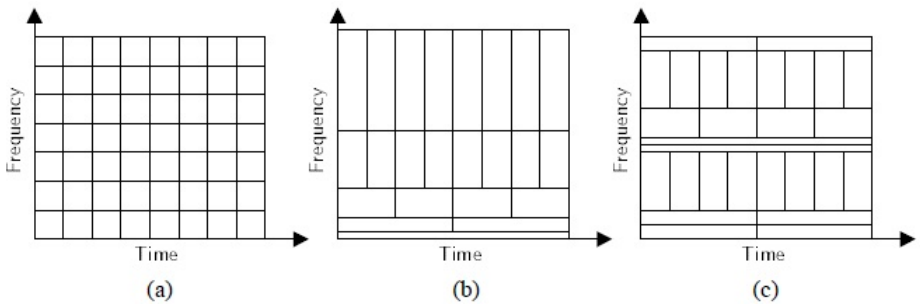

Figure 6: The time-frequency tiling of a) the STFT, b) the WT and c) the WPT [27].

[6] When in a STFT is used a Gaussian window is often called Gabor transform [30].

Huang and Chen [39] did a comparison of the performance of EMG time domain features (integrated EMG (IEMG), variance (VAR), bias zero crossings (BZC) [7], slope sign changes (SSC), waveform length (WL) and Willison amplitude (WAMP)) and EMG frequency domain features (auto-regressive coefficients (AR) of order four), in order to distinguish hand movements. Their work consisted in two stages. In the first stage, they applied the Davies-Bouldin index [8] with the aim to evaluate each feature, resulting that variance (VAR), wavelength (WL) and IEMG had better cluster separability than others. In the second stage, in order to get the best performance, the best features (VAR, WL and IEMG) determined in the previous stage were combined with other features (WAMP, BZC and AR of second order) to reinforce the whole clustering performance. From the results, two combinations of features were used in the neural network classification engine to test the results. The first combination consisted of IEMG, VAR, WL and WAMP; and the second combination composed of IEMG, VAR, WL, WAMP, BZC and second order AR model parameters. The combination of IEMG, VAR, WAMP, WL, BZC and second order AR parameters had the best performance for the pattern recognition of the hand movements.

Oskoei and Hu [65] applied advanced subset search algorithms rather than comparing index to evaluate EMG features of upper limb. These algorithms consisted of a genetic algorithm [9] adopted as the search strategy, Davies Bouldin index and Fishers linear discriminant index [10] employed as the filter objective functions [11] and linear discriminant analysis [12] used as the wrapper objective functions [13]. An artificial neural network was implemented as the classifier in the upper limb EMG system. Oskoei, and Hu evaluated EMG time domain features such as mean absolute value (MAV), mean absolute value slope (MAVS), root mean square (RMS), variance (VAR), waveform length (WL), zero crossings (ZC), slope sign changes (SSC) and Willison amplitude (WAMP), and auto-regressive coefficients (AR), frequency mean (FMN), frequency median (FMD) and frequency ratio(FR), as EMG frequency domain features. As a result, they found that WL demonstrated high capability in providing discriminating information for EMG classification; after that MAV and RMS were placed in second grade and AR in third grade.

Phinyomark *et al.* [71] compared eighteen time domain features (integrated EMG (IEMG), mean absolute value (MAV), modified mean absolute value 1 (MMAV1), modified mean absolute value 2 (MMAV2), mean absolute value slope (MAVS), simple square integral (SSI), variance (VAR), root mean square (RMS), waveform length (WL), zero crossings (ZC), slope sign changes (SSC), Willison amplitude (WAMP) and histogram of EMG (HEMG)) and frequency domain features (auto-regressive coefficients, frequency mean, frequency median, modified frequency mean and modified frequency median) in a noisy environment, with the aim to determine which one has a better tolerance of white Gaussian noise [14]. Their results showed that from the point of view of white Gaussian noise, modified frequency mean (MFMN) was the best feature comparing with others on the quality of the robustness of EMG features. MFMN had an average error of 6% on strong EMG signals and 10% on weak EMG signal at signal-to-noise ratio [15] value of 0 dB; also MFMN had an average error of 0.4% in both strong and weak EMG signals at signal-to-noise ratio value of 20 dB. In addition, MFMN and other robust features (WAMP and HEMG) were used as an input to the EMG pattern recognition. The experiment showed that these features were excellent candidates for a multi-source feature vector.

4.2 EEG feature extraction

Although there are time domain, frequency domain and time-frequency domain features in EEG feature extraction, there are different features used in EEG signal in comparison with EMG signal; this is because of the characteristics of the EEG signal, besides there are different waves in this signal, such as α, β, δ, γ and θ that have an important role in some applications (e.g. detecting sleeping disorders and brain diseases).

4.2.1 Time domain

These features are derived directly from the signal and include the (averaged) time-course.

[7] In this paper, it is called bias zero crossings because, they added a bias to the calculation of the zero crossings value in order to prevent influence from background disturbance.

[8] Davies-Bouldin index (DBI). It is a measure that reflects the degree of overlapping of the clusters with respect to their nearest neighbors. A lower value of the DBI implies a higher degree of cluster separability [39].

[9] Genetic algorithm. It is an optimization technique that mimics the evolutionary process of survival of the fittest [65].

[10] Fishers linear discriminant index. This index represents clusters' dispersion comparing to their scatter [65].

[11] Filter objective functions. Evaluate candidate subsets according to their information content(interclass distance, statistical dependence, or information-theoretic measures) and return a measure of their "goodness" as a feedback to select the new candidates [65].

[12] Linear discriminant analysis. Using training sample set, discriminant function $g(x)$ divides the feature space to distinct and labeled subspaces by the hyperplane decision surfaces [65].

[13] Wrapper objective function. It is a classifier evaluating candidate subsets according to the accuracy of its classification and returning a measure of their goodness as a feedback to select the new candidates [65].

[14] As a result, noise removal algorithm is not needed [71].

[15] Signal-to-noise ratio. It is a function of complicated interactions between the electrolytes in the skin and the metal of the detection surfaces of the electrode [21].

1. **Mean Value.** Since the sum of (positive and negative) EEG potential is usually on the order of a few microvolts when the analysis time is not too short, the mean is essentially a constant, although of a small value. Any shifts in values of the mean, therefore, are indicative of changes in potential that are of technical origin [11]. The mean value is given by:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

where x_i is time series for $i = 1, 2, ..., N$, and N is the number of data points.

2. **Standard deviation.** It is given by:

$$x_{std} = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}}$$

where $\bar{x}$ is the mean value.

3. **Maximum peak value.** It is the maximum absolute value of the segment k and is given by [31]:

$$x_k = max|x_i|$$

4. **Skewness.** Bronzino [11] explains that the skewness measures the degree of deviation from the symmetry of a normal or Gaussian distribution. This measure has the value of zero when the distribution is completely symmetrical and assumes some nonzero value when the EEG waveforms are asymmetrical with respect to the baseline(as is the case in some characteristic sleep patterns, mu rhythms, morphine spindles, etc.)

$$S_{k_{mc}} = \frac{\sum_{i=1}^{N}\frac{(x_i-\bar{x})^3}{N}}{\left[\sum_{i=1}^{N}\frac{(x_i-\bar{x})^2}{N}\right]^{3/2}}$$

where $S_{k_{mc}}$ is the moment coefficient of skewness.

5. **Kurtosis.** According to Bronzino [11], this measure reveals the peakedness or flatness of a distribution. In clinical electroencephalography, when EEG with little frequency and amplitude modulation is analyzed, negative values of kurtosis are observed; on the other hand high positives values of kurtosis are present when the EEG contains transient spikes, isolated high-voltage wave group, etc. The moment coefficient of Kurtosis K_{mc} is given by:

$$K_{mc} = \frac{\sum_{i=1}^{N}\frac{(x_i-\bar{x})^4}{N}}{\sum_{i=1}^{N}\left[\frac{(x_i-\bar{x})^2}{N}\right]^2} - 3$$

6. **Cross correlation.** A cross correlation sequence between two energy signals measures the extent of similarity between these two signals [72]. Chandaka *et al.* [14] explain that if a signal is correlated with itself, the resulting sequence is called the auto correlation sequence. The cross correlation of $x(n)$ and $y(n)$ is given by:

$$\hat{R}_{xy}(m) = \begin{cases} \sum_{n=0}^{N-m-1} x_{n+m}y_n & m \geq 0 \\ \hat{R}_{yx}(-m) & m < 0 \end{cases}$$

where $x(n)$ and $y(n)$ are two signal sequences, each of which with a finite energy, $m = ... - 2, -1, 0, 1, 2, ...$, represents the time shift parameter, and subscript xy stands for sequences being correlated. The order of the subscripts, with x preceding y, indicates the direction in which one sequence is shifted, relative to other.

4.2.2 Frequency domain

These features characterize the power of the brain signal in several frequency bands.

1. **Auto-Regressive coefficients (AR).** As we can see in works of [38], [75] and [92] this feature is used in EEG signal as well as in EMG signal.

2. **Power spectrum density (PSD).** It shows at what point frequency variations are strong and weak. The power spectrum density (PSD) is calculated by [40]:

$$PSD = \left| \sum_{i=0}^{N-1} x_i e^{\frac{-j2\pi ki}{N}} \right|^2$$

where $k = 0, 1, ..., N-1$, N is the length of the EEG data, and x_i represents the discrete samples of EEG data.

3. **Band power.** Based on Sabeti *et al.* [75] and Iscan *et al.* [40], EEG contains different specific frequency components, which carry the discriminative information. This type of feature reflects the energy in several bands (α, β, δ, γ and θ [16]). Once that the bands are filtered, the power spectrum density can be applied to each one in order to obtain important features.

4. **Asymmetry ratio PSD.** According to Palaniappan [67], this value is derived from the power spectrum density and can be used as a feature, also this feature is useful for EEG analysis when the mental activities that are studied exhibit inter-hemispheric difference. The asymmetry ratio power spectrum density is given by:

$$AS_{PSD} = \left[\frac{PSD_1 - PSD_2}{PSD_1 + PSD_2} \right]$$

where PSD_1 is the power spectrum density in one channel, and PSD_2 is the power spectrum density in another channel, but in the opposite hemisphere.

4.2.3 Time-Frequency domain

These features describe how spectrum power varies over time. As well as in EMG signal, the short time Fourier transform and the Wavelet transform are the most used.

4.2.4 Spatial filtering

This type of filtering uses signals from multiple electrodes to focus on activity at a particular location in the brain.

- **Bipolar montage.** Bipolar channels are computed subtracting the signals from two neighboring electrodes [87].
- **Common average reference.** This technique subtract the average value of the entire electrode montage (the common average) from that of the channel of interest [56].
- **Laplacian method.** It calculates for each electrode location the second derivative of the instantaneous spatial voltage distribution. The value of the Laplacian at each electrode location is calculated by combining the value at that location with the values of a set of surrounding electrodes. The distances to the set of surrounding electrodes determine the spatial filtering characteristics of the Laplacian [56].
- **Common spatial patterns.** It is a technique to analyze multi-channel data based on recordings from two classes (tasks). It is given by:

$$x_{CSP}(t) = x(t) \cdot W$$

where $x(t)$ is the signal, and W is a matrix that projects the signal in the original sensor space to a surrogate sensor space $x_{CSP}(t)$. Each column vector of a W is a spatial filter. CSP filters maximize the variance of the spatially filtered signal under one task and minimize it for the other task [87].

4.2.5 Kalman filter

Welch and Bishop [90] defined the Kalman filter as a set of mathematical equations that provides an efficient computational (recursive) means to estimate the state of a process, in a way that minimizes the mean of the squared error. Omidvarnia *et al.* [64] proposed the Kalman filter as an EEG feature vector.

[16] The frequencies and characteristics of each type of wave are explained in subsection 2.2

4.2.6 Fractal dimension

Sabeti *et al.* [75] explain that fractal dimension has a relation with entropy, and entropy has a direct relationship with the amount of information inside a signal; fractal dimension represents the degree of meandering (or roughness or irregularity) of a signal. According to Sabeti, there are three methods for calculating fractal dimension: Higuchi method, Katz method and Petrosian method.

Omidvarnia *et al.* [64] compared the features of AR, power of signal in different EEG bands (α, β, δ, γ and θ), wavelet coefficients and Kalman filter. Bayesian with a Gaussian kernel, Parzen estimation, K-nearest neighbor and back-propagation neural network were employed as the classifiers of the features. Kalman filter had the best performance over the other features when Parzen estimation, K-nearest neighbor and back-propagation neural network were used; while AR had better performance than Kalman filter when Bayes with a Gaussian kernel was used.

Sabeti *et al.* [75] tested the performance of the features of AR, band power, fractal dimension (calculated by Katz's, Higuchi and Petrosian methods) and wavelet energy, by using as a classifier linear discriminant analysis and support vector machines with the aim to select the most relevant features for EEG signal classification of schizophrenic patients. As a result the most consistent feature for discrimination of the schizophrenic patients and control participants was AR coefficients.

4.3 Dimensionality reduction

Once that we have our feature vector, in order to increase the performance of the classifier it is advisable to reduce its dimensionality by keeping the representative information in the feature vector and eliminating the redundant one from it. The resulting vector is called reduced feature vector. Based on Englehart [25], there are two main strategies of dimensionality reduction:

- **Feature projection.** This strategy tries to determine the best combination of the original features to form a new feature set, generally smaller than the original one [93]. Principal component analysis (PCA) is used in this type of strategy, it is an orthogonal linear transformation that can be used for dimensionality reduction in a signal dataset by retaining those characteristics of the dataset that contribute most to its variance, in order to do this, lower-order principal components are kept and higher-order ones are dismissed [6]. This could be applied to EMG or EEG signals.

- **Feature selection.** This strategy chooses the best subset of the original feature vector according to some criteria for judging whether one subset is better than another. The ideal criterion for classification should be the minimization of the probability of misclassification, but generally simpler criteria based on class separability are chosen [93].

Englehart *et al.* [27] compared principal component analysis (PCA) as a feature projection technique and Euclidean distance class separability (CS) as a feature selection technique. The features used were the Hudgins' time domain features (MAV, MAVS, ZC, SSC and WL), the short time Fourier transform, the wavelet transform and the wavelet packet transform in the classification of EMG signal. The result was that PCA provided more effective means of dimensionality reduction than feature selection by CS, when time-frequency feature sets were employed.

5 Stage 3: classification

Once the features have been extracted from the raw signal ("feature extraction") and the features with redundant information have been reduced ("dimensionality reduction") in the previous stage; it is necessary to distinguish different categories among the features by applying a classifier. These obtained categories are going to be used for the controller in the next stage. There are several techniques to classify data, for instance neural networks (NN), Bayesian classifier (BC), fuzzy logic (FL), linear discriminant analysis (LDA), support vector machines (SVM), hidden Markov models (HMM) and K-nearest neighbor (KNN).

5.1 Neural networks (NN)

A neuron is a cell in the brain whose principal function is the collection, processing and dissemination of electrical signals [74]; while an artificial neuron is an information-processing unit that is fundamental to the operation of a neural network [36]. This artificial neuron receives inputs x and weights w; inside the neuron is performing an arithmetic summing [17] followed by an activation function [50]. The arithmetic summing net consists in:

[17] The arithmetic summing constitutes a linear combiner [36].

$$net = \sum_{i=0}^{n} w_i x_i$$

where x_i is the input to the neuron, and w_i is the weight associated to x_i.

An activation function f_{act} is applied to net, giving as a result the output of the neuron O. This activation function limits the amplitude of the output of a neuron [36]; it needs to meet two conditions: 1) "activate" the neuron (near +1) when the "right" inputs are given and "inactive" it (near 0) when the "wrong" inputs are given; and 2) the activation needs to be nonlinear [74]. According to Kil, and Shin [50], there are several activation functions, such as, threshold linear, step, arbitrary step, exponential and sigmoid [18], see figure 7.

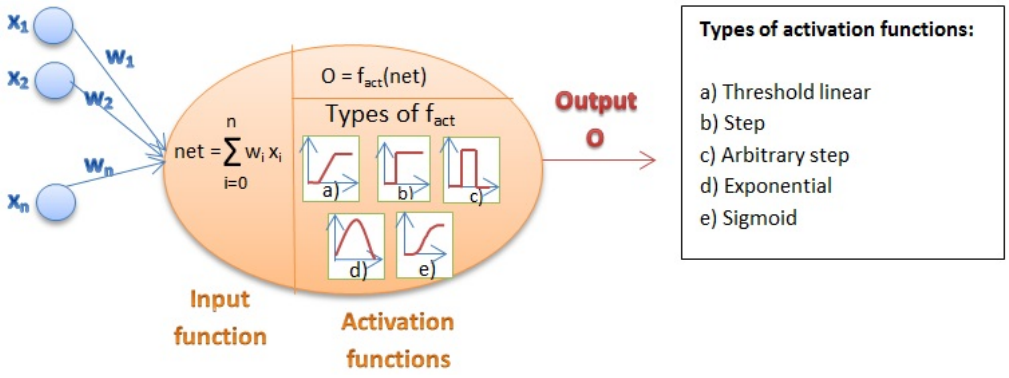

Figure 7: An artificial neuron

Feedforward networks are usually arranged in layers, such that each unit receives input only from units in the immediately preceding layer (feedforward type) [74]. In a multilayer feedforward network, there are one input layer, one or more hidden layers [19] and one output layer.

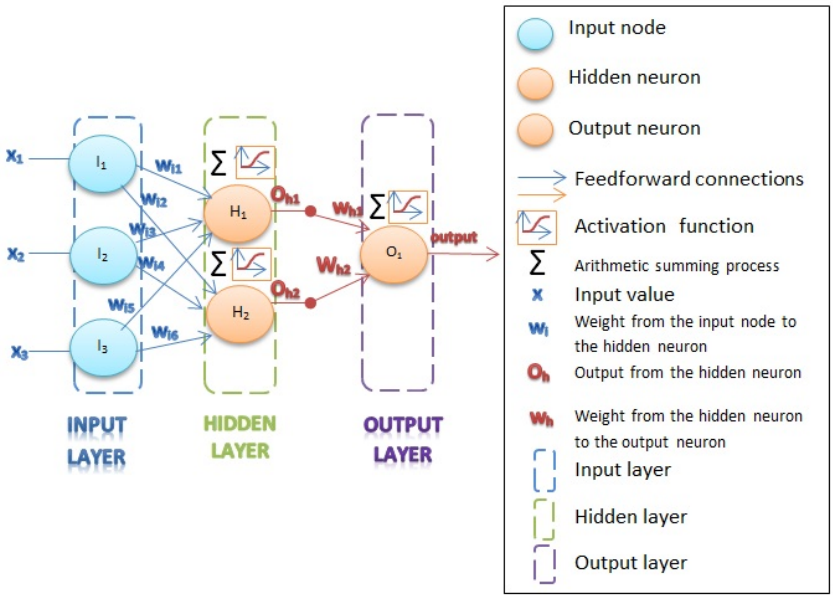

Figure 8: A fully connected feedforward network with one input layer, one hidden layer and one output layer

Haykin [36] explains that the input nodes in the input layer of the network supply respective elements of the activation pattern (input vector), which constitute the input signals applied to the neurons (computation nodes) in the second layer (i.e., the first hidden layer). The output signals of the second layer are used as inputs to the third layer (i.e., the second hidden layer or the output layer) and so on for the rest of the network. The set of output signals of the neurons in the output layer of the network constitutes the overall response of the network to the activation pattern supplied by the input nodes in the input layer, see figure 8. In a fully connected network, every node in each layer is

[18] The sigmoid function is the most common form of activation function used in the construction of neural networks [36].

[19] By adding one or more hidden layers, the network is enabled to extract higher-order statistics from its input [36].

connected to every other node in the adjacent forward layer; if some of the communication links are missing from the network, then the network is partially connected.

5.1.1 Back-propagation neural network (BPNN)

Kil and Shin [50] define a back-propagation neural network (BPNN) as a feedforward network with at least one hidden layer. Each neuron performs arithmetic summing followed by the sigmoid activation. The back-propagation algorithm is an iterative gradient descent algorithm to minimize the mean-squared error between the desired output and the actual network output.

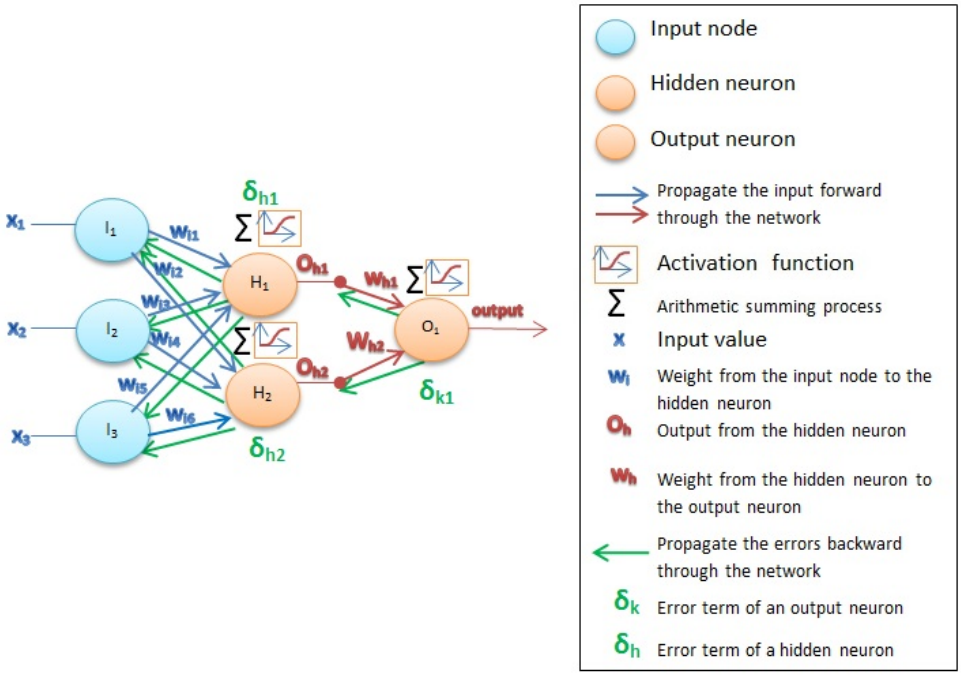

Figure 9: A back-propagation neural network

The steps for applying the backpropagation algorithm are the following [58], see figure 9:

1. Create a feedforward network with n_{in} input nodes, n_{hidden} hidden neurons and n_{out} output neurons.
2. Initialize all network weights to small random numbers.
3. Until the termination condition [20] is met, do:

   For each $< X, T >$ in training examples, do: **Propagate the input forward through the network:**

   (a) Input the instance X to the network and compute the output o_u of every unit u in the network.

   Propagate the errors backward through the network:

   (b) For each output neuron k, calculate its error term δ_k

   $$\delta_k \leftarrow o_k(1 - o_k)(t_k - o_k)$$

   (c) For each hidden neuron h, calculate its error term δ_h

   $$\delta_h \leftarrow o_h(1 - o_h) \sum_{k \in outputs} w_{kh}\delta_k$$

   (d) Update each network weight w_{ji}

   $$w_{ji} \leftarrow w_{ji} + (n\delta_j x_{ji})$$

[20] A termination condition to halt the procedure could be: a) after a fixed number of iterations through the loop; b) once the error on the training examples falls below some threshold; c) or once the error on a separate validation set of examples meets some criterion [58].

where n_{in} is the number of input nodes; n_{hidden} is the number of hidden neurons; n_{out} is the number of output neurons; $< X, T >$ is a training example; X is the vector of network input values; T is the vector of target network output values; t_k is the target value associated with the kth output neuron; $output$ is the set of output neurons in the network; w_{kh} is the weight from hidden neuron h to output neuron k; n is the learning rate; x_{ji} is the input from node i to unit j; and w_{ji} is the weight from node i to unit j.

5.1.2 Bayesian neural network (BNN)

It is a directed acyclic graph where the nodes are probability variables representing certain events. According to Bu *et al.* [12], a BNN can be defined as $G = (V, A, P)$; where $V = \{v_1, v_2, ..., v_n\}$ is a set of nodes (variables), A is an assembly of directed arcs between the nodes, and P is a set of conditional probability tables that are associated with each node. A directed arc from v_i to v_j, $(v_i, v_j) \in A$, represents the conditional dependency between the variables; this dependency is indicated with $P(v_j = a|v_i = b)$, which is the conditional probability for $v_j = a$ given that $v_i = b$.

5.1.3 Log-linearized Gaussian mixture network (LLGMN)

LLGMN is a three-layer feedforward NN. The structure of LLGMN is based on the Gaussian mixture model (GMM) and a log-linear model. This NN is able to estimate probability density functions (pdfs) of input patterns [12]. According to Oskoei and Hu [3], a Gaussian mixture model (GMM) has the ability to form a smooth approximation for general probability density functions. The probability density of GMM, which is called mixture density (MD), is a linear combination of multiple standard Gaussian probability densities (named components). The complete Gaussian mixture model is parameterized by the mean vectors, covariance matrices and mixture weights from all component densities.

Tsuji *et al.* [85] present the structure of a LLGMN, see figure 10. In the LLGMM, the pdf of class $c(c = 1, ..., C)$ is approximated with a GMM. The input vector x is converted into the modified vector X, by applying a log-linearization process to the a priori probability for each component $\{c, m\}$ and the probability of x to be generated from the component m in class c. The first layer consists of H units corresponding to the dimension X and the identity function is used for activation of each unit. ${}^{(1)}O_h$ denotes the output of the hth unit in the first layer. In the second layer, each M_c unit receives the output of the first layer weighted by the coefficient $w_h^{(c,m)}$ and outputs the a posteriori probability of each component. The third layer consists of C units corresponding to the number of classes and outputs the a posteriori probability for class $c(c = 1, ..., C)$; each C unit integrates the outputs of the M_c units of the second layer.

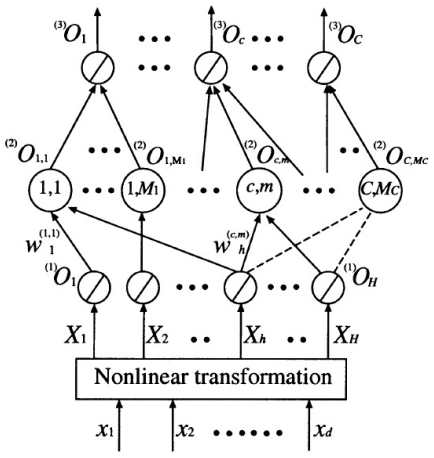

Figure 10: A Log-linearized Gaussian mixture network [85]

5.1.4 Recurrent network

In a recurrent network, additional to the feedforward connections, units have self-connections or connections to units in the previous layers. This recurrency acts as a short-term memory and lets the network remember what happened in the past [2], see figure 11.

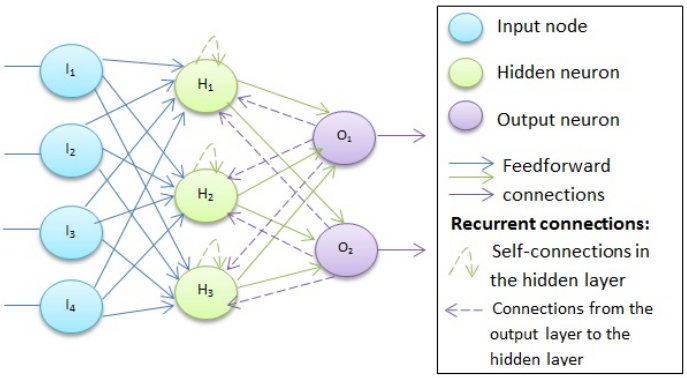

Figure 11: A recurrent network

5.1.5 Wavelet neural network (WNN)

Wavelet neural network (WNN) is a neural network based on wavelet transform, in which discrete wavelet function is used as the node activation function, see figure 12. Because the wavelet space is used as characteristic space of pattern recognition, the characteristic extraction of signal is realized by weighted sum of inner product of wavelet base and signal vector [82].

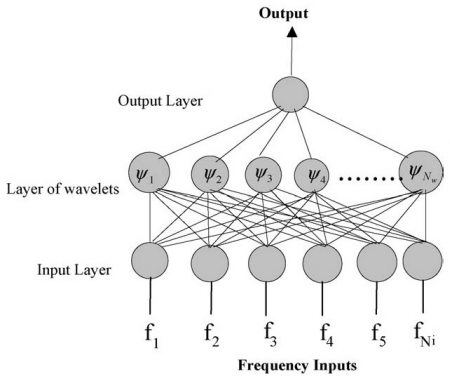

Figure 12: A wavelet neural network model [82]

5.1.6 Applications in EMG and EEG

Huang and Chen [39] developed a myoelectric discrimination system for a multi-degree prosthetic hand. They classified eight types of hand movements, such as three-jaw chuck, lateral hand, hook grasp, power grasp, cylindrical grasp, centralized grip, flattened hand and wrist flexion. They employed a back-propagation neural network (BPNN) for discriminating among the feature sets. The BPNN had one hidden layer and one output layer. The transfer functions for hidden layer neurons and output layer neurons were all nonlinear sigmoid functions. The discrimination system achieved success rates of 85% for off-line test and of 71% for on-line test.

Also, Karlik [47] classified EMG signals for controlling multifunction prosthetic devices by using a three-layered back-propagation neural network (BPNN). The inputs of the BPNN were auto-regressive (AR) parameters of a1, a2, a3, a4 and signal power obtained from different arm muscle motions, see figure 13. The result was an accuracy rate of 97.6% for categorizing six movements (R: resting, EF: elbow flexion, EE: elbow extension, WS: wrist supination, WP: wrist pronation and G: grasp) in 5000 iterations.

Tsuji *et al.* [85] proposed a neural network, called "recurrent log-linearized Gaussian mixture network (R-LLGMN)" for classification of time series, more specific for EEG signals. The structure of this network was based on a hidden Markov model (HMM). R-LLGMN can as well be interpreted as an extension of a probabilistic neural network using a log-linearized Gaussian mixture model, in which recurrent connections had been incorporated to make temporal information in use. Tsuji et al. classified EEG signals of a photic stimulation consisting in opening and closing eyes. They compared five classifiers: LLGMN (log-linearized Gaussian mixture network), LLGMN with RNF (recurrent neural filter), HMM, CDHMM (continuous density HMM) and R-LLGMN. Their results showed R-LLGM as the best classifier with an average accuracy rate of 94.4%, followed by LLGMN with NF and HMM,

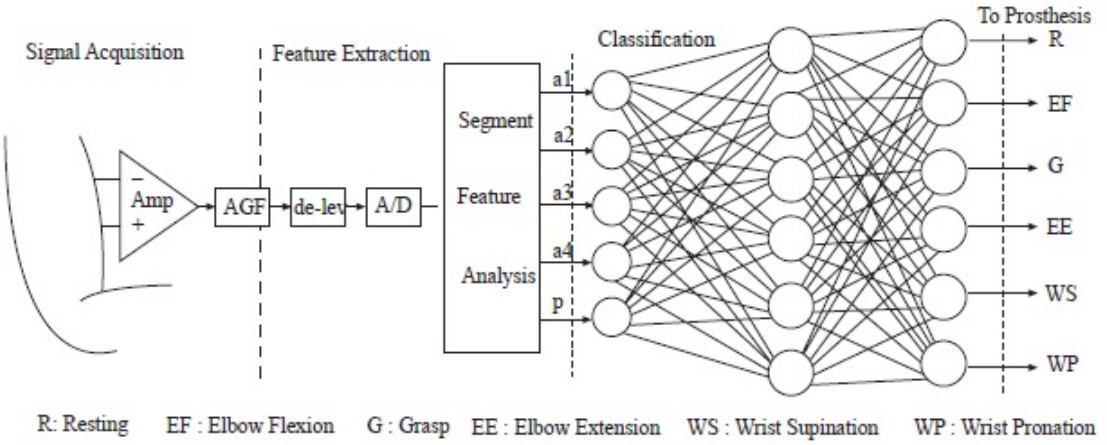

Figure 13: Myoelectric control of a multifunction prosthesis, using a back-propagation neural network as a classifier [47]

both with an accuracy rate of 92.5%. While CDHMM and LLGMN achieved accuracy rates of 92.1% and 88.5%, respectively.

Chu *et al.* [18] proposed a real-time EMG pattern recognition for the control of a multifunction myoelectric hand from four channel EMG signals. To extract a feature vector from the EMG signal, they used a wavelet packet transform. For dimensionality reduction and nonlinear mapping of the features, they proposed a linear-nonlinear feature projection composed of principal components analysis (PCA) and a self-organizing feature map (SOFM). They employed to classify a multilayer perceptron (MLP) with: a) an input layer constructed from the eight outputs of the SOFM for four channels; b) two hidden layers, each hidden layer with nine neurons and c) an output layer with nine neurons for the nine hand motions to be recognized. The average classification success rates of the MLP were 97.024% when PCA+SOFM were used, 97.785% when SOFM was applied and 95.759% when PCA was employed. Their experimental results showed that all processes, including virtual hand control, were completed within 125 ms.

Subasi *et al.* [82] compared back-propagation neural networks (BPNN) and wavelet neural networks (WNN) for classifying neuromuscular disorders of EMG recordings. They used an auto-regressive (AR) model of EMG signals as an input to classification system. The BPNN was designed with AR spectrum of EMG signal in the input layer; the output layer consisted of three nodes representing normal, myopathic or neurogenic disorder. On the other hand, the WNN was designed with mono-hidden-layer forward neural network with its node activation function based on dyadic discrete Morlet wavelet basic function able to unambiguously locate the three classes. A total of 1200 MUPs obtained from 7 normal subjects, 7 subjects suffering from myopathy and 13 subjects suffering from neurogenic disease were analyzed. The success rate for the WNN technique was 90.7% and for the BPNN technique 88%.

5.2 Bayesian classifier (BC)

According to Mitchell [58], Bayes theorem provides a way to calculate the probability of a hypothesis based on its prior probability, the probabilities of observing various data given the hypothesis and the observed data itself:

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)}$$

where $P(h)$, the prior probability, denotes the initial probability that hypothesis h holds, before the training data has been observed [21]; $P(D)$ stands for the prior probability that training data D will be observed (i.e., the probability of D given no knowledge about which hypothesis holds), and $P(D|h)$ is the probability of observing data D given some world in which hypothesis h holds. The posterior probability $P(h|D)$, reflects the confidence that h holds after the training data D has been seen.

Bu *et al.* [12] proposed a task model using a Bayesian network (BN) for motion prediction. Given information of the previous motion, this task model was able to predict occurrence probabilities of the motions concerned in the task. Furthermore, a hybrid motion classification framework was developed based on the BN motion prediction. Besides the motion prediction, EMG signals were simultaneously classified by a log-linearized Gaussian mixture network (LLGMN). Then, the probabilities, which are outputs of LLGMN and the BN task model, were combined to generate motion commands for controlling, see figure 14. Experiments were conducted with four subjects to demonstrate the feasibility of the proposed methods. In these experiments, forearm motions were classified with EMG signals

[21] if there is no prior knowledge, then the same prior probability is assigned to each candidate hypothesis [58]

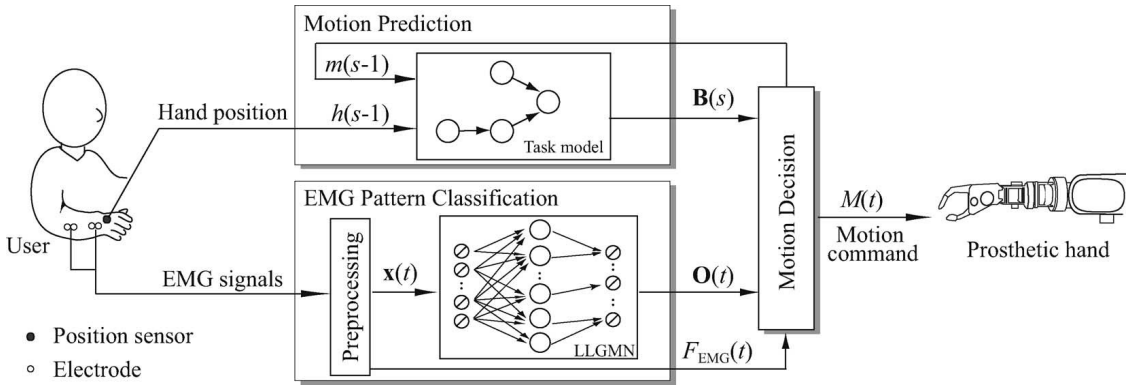

Figure 14: Hybrid structure of an EMG-based human-robot interface, where the Bayesian network task model is incorporated for motion prediction [12].

considering a cooking task. The classification rates of used only the LLGMN and the proposed (BN with LLGMN) method were 85.1% and 92.9%, respectively.

5.3 Fuzzy Logic (FL) Classifier

According to Sivanandam *et al.* [78], the Fuzzy Logic tool was introduced in 1965 by Lotfi Zadeh. It is a mathematical tool for dealing with uncertainty, imprecision and information granularity. The fuzzy theory employs fuzzy sets, membership functions, rules, fuzzification and defuzzification. **Fuzzy sets** model the uncertainty associated with vagueness, imprecision and lack of information regarding a problem; involving linguistic variables such as "low", "medium", "high", "often", "few", etc. In fuzzy logic, a **membership function** has various "degrees of membership", values between on the real continuous interval [0,1], these values represent the "fuzziness". Some basic shapes of membership functions are trapezoidal, triangular and gaussian. The **rules** constitute the basis for the fuzzy logic to obtain the fuzzy output, they convert inputs to outputs. These rules have the general form "IF X is A THEN Y is B", where A and B are fuzzy sets.

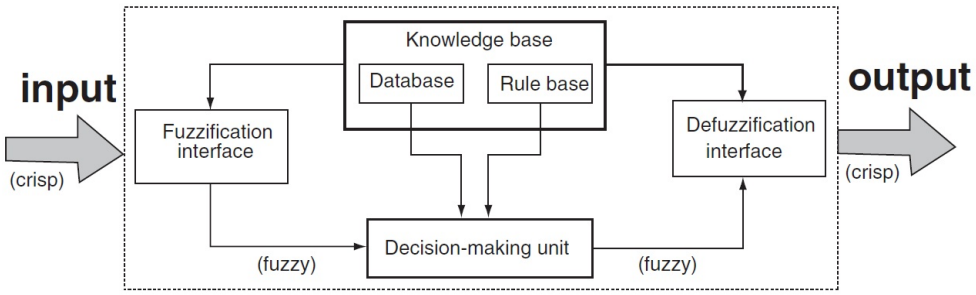

Figure 15: Fuzzy Inference System [78].

Two conversion processes are used in fuzzy logic, fuzzification and defuzzification. **Fuzzification** is the process where the crisp quantities are converted to fuzzy (crisp to fuzzy). By identifying some of the uncertainties present in the crisp values, the fuzzy values are formed. The conversion of fuzzy values is represented by the membership functions. On the other hand, **defuzzification** converts the fuzzy quantities into crisp quantities for further processing, since the fuzzy results generated cannot be used as such to the applications. A Fuzzy inference system, see figure 15, consists of a fuzzification interface, a rule base (containing a number of fuzzy IF THEN rules), a database (which defines the membership functions of the fuzzy sets used in the fuzzy rules), a decision-making unit (which performs the inference operations on the rules) and a defuzzification interface [78].

Si *et al.* [76] designed an expert system for the pediatric intensive care unit with the aim of alerting experts about the level of abnormality of the EEG of the patients. They used fuzzy logic and neural networks to classify the data in the expert system. Four fuzzy sets were used for the amplitude of the EEG: severe, moderate, mild and normal. Results showed an accuracy percentage of 91%.

On the other hand, James *et al.* [42] developed a multi-stage system for automated detection of epileptiform activity in the EEG; using fuzzy logic and an artificial neural network, called organizing feature map (SOFM). SOFM was in charge of assigning a probability value to incoming candidate epileptiform discharges (ED), while fuzzy logic

was employed to incorporate spatial contextual information in the detection process of ED. The spatial-combiner used a number of rules that specify allowable combinations of EDs across channels to detect an epileptiform event (EV). Five fuzzy sets were defined to classify the probabilities of true ED (Inputs to the spatial-combiner): negative big (NB), negative small (NS), zero (ZE), positive small (PS) and positive big (PB). While each sub-system produced a single output that was defuzzified using the fuzzy sets zero (ZE), possible (POS), probable (PRO) and definite (DEF), (outputs of the spatial-combiner), see figure 16. In both cases trapezoidal membership functions were used. The fuzzy-rules were drawn-up based on the authors' pre-defined knowledge of how an EV would manifest itself across a bipolar electrode chain. Results showed that the system has a selectivity of 82%.

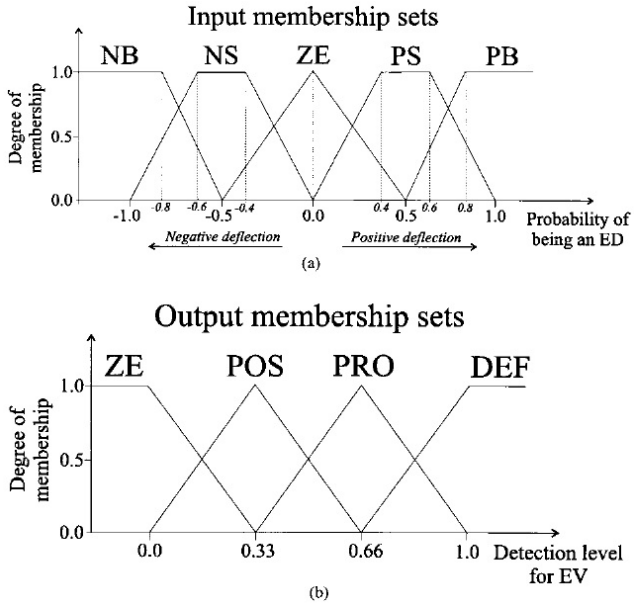

Figure 16: Fuzzy sets used to define (a) the inputs to and (b) the outputs of the spatial-combiner [42]

Kiguchi *et al.* [49] developed a robotic exoskeleton for human upper-limb motion assist. The robotic exoskeleton was controlled using EMG signals. The classifier was a hierarchical neuro-fuzzy controller, consisting of three stages: 1)input signal selection, 2) posture region selection stage and 3) neuro-fuzzy control. Three membership functions were used for the elbow and shoulder regions: flexed angle(FA), intermediate angle (IA) and extended angle (EA), see figure 17. By applying these membership functions, the appropriate controllers were moderately selected in accordance with the arm posture of the robot user. Two kinds of nonlinear functions (and) were applied to express the membership function of the neuro-fuzzy controller. The initial fuzzy $IF-THEN$ control rules were designed based on the analyzed human elbow and shoulder motion patterns, then they were transferred to the neural network form.

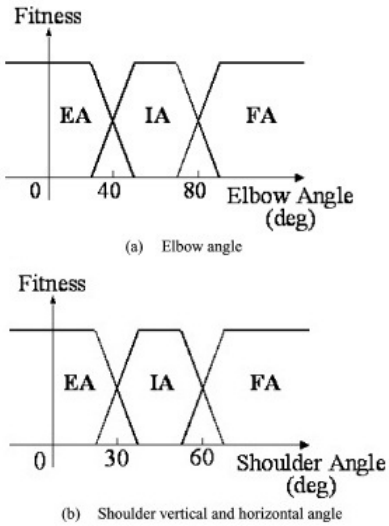

Figure 17: Membership functions [49]

Ajiboye and Weir [1] presented a heuristic fuzzy logic approach to multiple EMG pattern recognition for multi-functional prosthesis control. Basic signal statistics (mean and standard deviation) were used for membership function construction and fuzzy c-means (FCMs) data clustering was used to automate the construction of a simple amplitude-driven inference rule base. The multiinput-single-output fuzzy system consisted of three parts: 1) input membership functions that fuzzify numerical inputs, converting them to linguistic variables; 2) an inference rule base that performs pattern classification by processing the linguistic inputs, returning linguistic outputs and associated degrees of truth and 3) an output membership function that defuzzifies the inference rule base linguistic outputs, converting them to one numerical value. Four fuzzy sets were defined of signal gradation (OFF, LOW, MED, HIGH). Ajiboye, and Weir, used fuzzy -means (FCM) clustering to generate the rules. Overall classification rates ranged from 94% to 99%.

5.4 Linear Discriminant Analysis (LDA) Classifier

Balakrishnama and Ganapathiraju [5] explain that Linear Discriminant Analysis easily handles the case where the within-class frequencies are unequal and their performances have been examined on randomly generated test data. This method maximizes the ratio of between-class variance to the within-class variance in any particular data set thereby guaranteeing maximal separability, with this criterion the axes of the transformed space are defined. Also, they describe two approaches to transform data sets and classify test vectors in the transformed space: a) class-dependent transformation, that maximizes the ratio of between class variance to within class variance and b) class-independent transformation, that maximizes the ratio of overall variance to within class variance and each class is considered as a separate class against all other classes.

Fukunaga [29] points out that LDA (also known as Fishers LDA) uses hyperplanes to separate the data representing the different classes. The separating hyperplane is obtained by seeking the projection that maximizes the distance between the two classes means and minimizes the interclass variance. To solve an $N - class$ problem ($N > 2$) several hyperplanes are used.

Sabeti *et al.* [75] analyzed EEG signals of 20 schizophrenic patients and 20 age-matched control participants using 22 channels, with the aim of determining the most informative channels and finally distinguishing the two groups. Bidirectional search and plus-L minus-R techniques were employed to select the most informative channels; while linear discriminant analysis and support vector machines (SVM) were used as the classifiers. The results were accuracy rates of 84.62% for LDA and 99.38% for SVM when bidirectional search was used, and 88.23% for LDA and 99.54% for SVM when LRS technique was applied.

5.5 Support Vector Machines (SVM) Classifier

Choi and Cichocki [16] explain that the purpose of the technique of support vector machines is to find a hyperplane [22] that corresponds to the largest possible margin between the points of different classes k. This hyperplane then forms the decision boundary for classifying new data points, the points forming the boundary are called support vectors; while the distance from the hyperplane to the instances closest to it on either side is called the margin [2], see figure 18.

The algorithm for linear SVM assigns the value of 1 or -1 to the data points $x_1, \ldots, x_n$, with labels $y_1, \ldots, y_n$, according to:

$$w \cdot x_i + b + \xi_i \geq 1, \text{ if } y_i = 1 \; w \cdot x_i + b + \xi_i \leq -1, \text{ if } y_i = -1$$

where w is the normal to the chosen hyperplane, b is the intercept term, chosen to maximize the margin of the decision boundary, and ξ is the distance of the misclassified points from the hyperplane.

When there are more than two classes, $k > 2$, a solution is to define k two-class problems, each one separating one class from all other classes combined and applying the algorithm for all the classes. If the classes are not linearly separable such that there is no hyperplane to divide them, the problem can be mapped to a new space by doing a nonlinear transformation using suitably chosen basis functions and then use a linear model in this new space. To do this, Kernel functions, such as radial-basis, polynomial and sigmoidal, can be employed [2].

Yom-Tov and Inbar [92] designed a classifier combining a genetic algorithm and support vector machines (SVM) to distinguish between movements of contralateral fingers using movement-related potentials embedded in EEG. Their results showed that, it was possible to select as few as 10 subject-specific features and achieved average accuracy rates of 87% between two limbs and 63% between three limbs. Crawford *et al.* [20] developed a 4-degrees-of-freedom robotic arm. They employed linear SVM as the classifier, achieving accuracy rates of 92-98% in 3 subjects.

Halder *et al.* [34] proposed a combination of blind source separation (BSS) and independent component analysis (ICA) (signal decomposition into artifacts and nonartifacts) with SVM (automatic classification); in order to isolate

[22] Hyperplanes. Surfaces, which could be found from prototypes (or training samples) to separate different classes of patterns in the n-dimensional space and are used to classify unknown patterns [10].

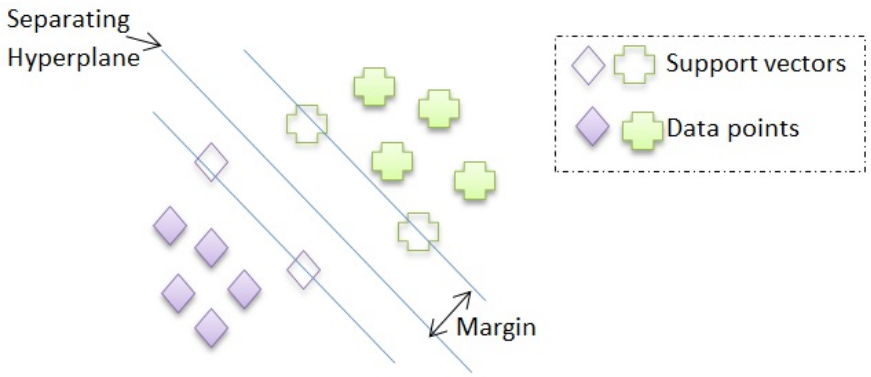

Figure 18: Linear support vector machines

EMG and electrooculographic (EOG) artifacts into individual components. The accuracy percentages of the classification between artifacts and nonartifacts were 99.39% for eye blink, 99.62% for eye movement, 92.26% for jaw muscle and 91.51% for forehead. Choi and Cichocki [16] controlled a motorized wheelchair online, they used the linear SVM for classifying the feature vectors obtained from the EEG data into each class of motor imagery. Three subjects participated in the experiments; each one performed an imaginary movement of the hand and foot, depending of the direction of an arrow, showing on the computer.

Firoozabadi *et al.* [28] developed a hands-free control system for operating a virtual wheelchair, which was based on forehead multi-channels bio-signals as EMG signals. SVM was used for classifying the motion control commands (forward, left, right, backward and stop). Three subjects (one adult and two children) participated in controlling a virtual wheelchair using an interface software on a personal computer. The accuracy percentages of SVM classification were: 100% for the adult and 89.75% and 97.49% for the two children. Lucas *et al.* [54] proposed a multi-channel supervised classification of EMG signals with the aim of controlling myoelectric prostheses. The classification of six hand movements was performed with SVM approach in a multi-channel representation space. The results showed an average misclassification rate of 5%.

Oskoei and Hu [66] evaluated the application of SVM to classify upper limb motions using EMG signals. Four popular kernels were examined: radial-basis, linear, polynomial and sigmoid. The four applied kernels performed similarly. This could be interpreted that the boundaries between classes were almost linear. The average accuracy for all kernels was approximately $95.5 \pm 3.8\%$. Gurmanik *et al.* [32] proposed an integrated binary classifier based on SVM for differentiating neuromuscular disorders, using EMG signal. The objective of SVM was to find optimal hyperplane for separating MUAP clusters. They used threshold technique for segmentating EMG signal and autoregressive coefficients (AR) as features. A total of 12 EMG signals obtained from 3 normal (NOR), 5 myopathic (MYO) and 4 motor neuron diseased (MND) subjects were analyzed. The classification accuracy of binary SVM with AR features was 100%.

Subasy, and Gursoy [81] developed an EEG signal classification method for diagnosing epilepsy. This method was based on discrete wavelet transform; the dimension reduction was performed by principal components analysis (PCA), independent components analysis (ICA) and linear discriminant analysis (LDA), while the classification was carried on by SVM with a radial basis function (RBF) as a kernel. The classification rate with LDA feature extraction was highest (100%), ICA came as second (99.5%); while the PCA had the lowest correct classification percentage (98.75%). SVM without using PCA, ICA or LDA achieved an accuracy rate of 98%.

Wei and Hu [89] designed a human-machine interaction (HMI) for hands-free control of a wheelchair, employing forehead EMG signals and color face image information. They used five recognizable movements: SJC (single jaw clenching), DJC (double jaw clenching) and CJC (continuous jaw clenching) from jaw movements and LEC (left eye close) and REC (right eye close) from eye movements. SJC and DJC patterns were recognized using a threshold based strategy; while CJC, LEC and REC were separated using SVM with a radial-basis function kernel. The result achieved by SVM had an accuracy rate of over 93%.

5.6 Hidden Markov Models (HMM) Classifier

According to Kil and Shin [50], a hidden Markov model consists of: a) the number of states in the model; b) the number of distinct observation symbols per state; c) the observation symbol probability distribution in each state for the discrete HMM, usually modeled as a Gaussian mixture; d) the state transition probability and e) the initial-state occupancy probability, see figure 19. The mainly benefits of using HMM are: a) its effectiveness in capturing time-varying signal characteristics; b) its ability to model unknown signal dynamics statically and c) its computational

tractability due to the inherent statistical property of Markov processes. [23]

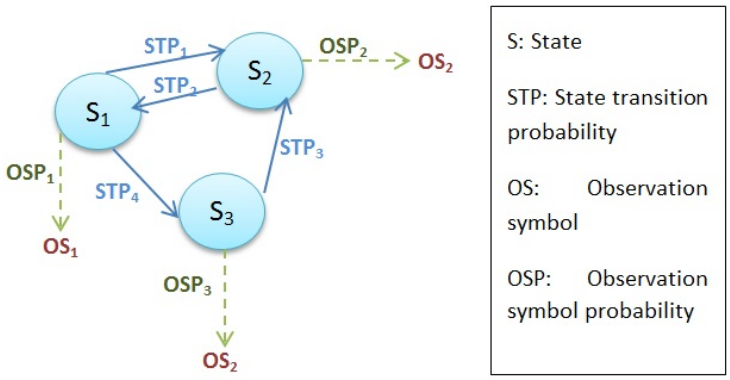

Figure 19: HMM components

Novák *et al.* [59] employed HMM in scoring of human sleep. They used three HMM states, one corresponding to wake state, other representing deep sleep and the other one standing for REM sleep. Obermaier *et al.* [62] developed a letter spelling device operated by spontaneous EEG, whereby the EEG is modulated by mental hand and leg motor imagery. They employed two HMM for representing motor imagery; reporting that the ability of three people in the use of the letter spelling device varied between 0.85 and 0.5 letters/min in error-free writing.

Chan and Englehart [13] used HMM to process four channels of EMG signal, with the task of discriminating six classes of limb movement. In their work, six-state fully connected HMM were applied; each state was associated with an intended limb motion. HMM classification of continuous myoelectric signals resulted in an average accuracy of 94.63%. Solhjoo *et al.* [79] studied the performance of two kinds of HMM, discrete HMM (dHMM) and multi-Gaussian HMM (mHMM), in the classification of EEG based mental task. This task implied the controlling of a feedback bar by means of imagery left or right hand movements according to the cues shown to the subject. The best performance of dHMM was 77.13 % with 2 states and 16 observable symbols/state according to 0.5s segment of data; while for mHMM was 77.5% using first 0.5s segment, with 8 states and 2 Gaussians/state.

5.7 K-nearest neighbor (KNN) Classifier

According to Bow [10], KNN is a process to assign a pattern point to a class to which its nearest neighbor belongs. If membership is decided by a majority vote of the k-nearest neighbors, the procedure will be called a k-nearest neighbor decision rule. The algorithm consists of two stages: a) make pregrouping of data to obtain subclusters by using Euclidean distance and b)merge the subclusters hierarchically by using a similarity measure.

Peleg *et al.* [68] employed KNN as a classifier for EMG signals in finger activation, in order to be used in a robotic prosthesis arm. While Chaovalitwongse *et al.* [15] used KNN to classify normal and abnormal (epileptic) brain activities employing EEG recordings; having as results in the classification a sensitivity of 81.29% and a specificity of 72.86%, on average, across ten patients.

5.8 Combination of Classifiers

Lotte *et al.* [53] present boosting, voting and stacking techniques as classifier combination strategies used in EEG signal analysis. **Boosting** consists in using several classifiers in cascade, each classifier focusing on the errors committed by the previous ones [22]. In **voting**, several classifiers are employed, each of them assigning the input feature vector to a class. The final class will be that of the majority [53]. In **stacking**, the outputs of the individual classifiers are used to train the "stacked" classifier. The final decision is made based on the outputs of the stacked classifier in conjunction with the outputs of individual classifiers [41].

Okamoto *et al.* [63] used a hierarchical pattern classification algorithm based on boosting approach for the estimation of a suitable network structure. In this algorithm, the structure of the classification network was automatically constructed by adding LLGMNs (log-linearized Gaussian mixture network) as classifiers, with the aim of categorizing EMG signals involving six Japanese phonemes.

5.9 Comparison of Classifiers

Huan and Palaniappan [38] used linear discriminant analysis and multilayer perceptron neural network trained by the back-propagation algorithm (MLP-BP) to classify mental tasks using features that were extracted from EEG signals.

[23] A Markov process is a stochastic process whose past has no bearing on the future as long as the present is specified [50].

They employed the following feature methods: AR coefficients computed with Burgs algorithm, AR coefficients computed with a least-squares (LS) algorithm and adaptive auto-regressive (AAR) coefficients computed with a least-mean-square (LMS) algorithm. The results showed that sixth-order AR coefficients with the LS algorithm without segmentation gave the best performances (93.10%) using MLP-BP and (97.00%) using LDA.

Omidvarnia *et al.* [64] compared the performance of several classifiers using as features AR, power of signal in different EEG bands (α, β, δ, γ and θ), wavelet coefficients and Kalman filter. The employed classifiers were Bayesian with a Gaussian kernel, Parzen estimation, K-nearest neighbor and back-propagation neural network. When the AR coefficients were used as a feature, the accuracy rates were: 90.44%, 87.69%, 93.16% and 83.88% for Bayes, Parzen, KNN and MLP, respectively. When the wavelet coefficients were applied, the accuracy percentages were 88.58% for Bayes, 75.33% for Parzen, 81.33% for KNN and 85.19% for MLP. When the power of signal was employed as a feature, the accuracy rates were 83% for Bayes, 77.27% for Parzen, 81.72% for KNN and 82.27% for MLP. Finally, when Kalman filter was used as a feature, the accuracy results were 78.29%, 92.23%, 96.13% and 93.86%, for Bayes, Parzen, KNN and MLP, respectively. In conclusion when the used features were AR and Kalman filter, the best classifier was KNN; and when wavelet coefficients and power of signal were used as features, the best classifier was Bayesian with a Gaussian kernel.

Lotte [52] compared four classifiers in order to categorize motor imagery signals using EEG signal. These classifiers were a fuzzy inference system (FIS), a support vector machines with gaussian kernel (SVM), a multilayer perceptron (MLP) and a perceptron as a linear classifier (LC). The best performance was achievable by SVM with an accuracy percentage of 79.4 %, followed by FIS with an accuracy percentage of 79 %. MLP and LC had accuracy percentages of 78.9 % and 76.2%, respectively.

	Bio-signal	Accuracy percentage of the classifiers							
		Parzen	KNN	NN	Bayes	FL	SVM	LDA	LC
2004, Huan, and Palaniappan	EEG			93.10				97 ☆	
2005, Omidvarnia *et al.* AR / kalman filter	EEG	87.69/ 92.23	93.16/ 96.13 ☆	83.88/ 93.86	90.44/ 78.29				
Wavelet / power of signal		75.33/ 77.27	81.33/ 81.72	85.19/ 82.27	88.58/ 83 ☆				
2006, Lotte	EEG			78.9		79	79.4 ☆		76.2
2008, Oskoei, and Hu	EMG			N/E			95.5 ☆	94.5	
2008, Zhou *et al.* Using own data	EEG			90			91 ☆	88	

KNN: K-nearest neighbour
NN: neural networks
N/E: not specified
FL: fuzzy logic
SVM: support vector machines
☆: best classifier
LDA: linear discriminant analysis
LC: perceptron as a linear classifier

Figure 20: Comparison of classifiers

Oskoei and Hu [66] compared SVM, LDA and multilayer perceptron neural network (MLP) in classifying upper limb motions using myoelectric signals. They used four kernels (radial-basis, linear, polynomial and sigmoid) in SVM, and two multilayer perceptron neural network, one with one hidden layer (MLP1) and the other one with two hidden layers (MLP2). The average accuracy for all kernels in SVM was approximately 95.5 $\pm$ 3.8%. The LDA was placed after SVM with the average performance of 94.5 $\pm$ 4.9%. The MLP2 performed a similar accuracy to the SVM and LDA, while the accuracy of MLP1 dropped approximately 6%.

Zhou *et al.* [94] used a feature set including higher-order statistics based on the bispectrum of EEG signals for classifying EEG signals corresponding to left/right-hand motor imagery. Support vector machines with Gaussian kernel (SVM), linear discriminant analysis (LDA) and neural networks (NN) were used as classifiers and were compared with the winners classifier of BCI-competition 2003, using the same BCI data set and using their own data. In the NN, they employed an input layer with 24 nodes for the features, a hidden layer with 15 nodes and an output layer with two nodes for the classes of hand motor imagery; also they chose back-propagation algorithm to train the NN. The results showed that, using the same BCI data set and their own features, the best classifiers were NN and SVM, both with a minimal misclassification rate of 10%; while using their own data and their own features, the best classifiers were SVM, NN and LDA, with minimal misclassification rates of 9%, 10% and 12 %, respectively. All the above works are summarized in Figure 20.

Multifunction prosthesis			
	Feature extraction	Classifier	Application
1999, Huang, and Chen [39]	IEMG, VAR, bias ZC, SSC, WL, WAMP and AR	BPNN	A myoelectric discrimination system for a multi-degree prosthetic hand
2002, Peleg *et al.* [68]	AR and discrete Fourier transform	KNN	Finger activation for using a robotic prosthetic arm
2006, Chu *et al.* [18]	Wavelet packet transform	Multilayer perceptron	Control of a multifunction myoelectric hand
2008, Oskoei, and Hu [66]	MAV, RMS, WL, VAR,ZC, SSC, WAMP, MAV1, MAV2, power spectrum, AR,FMN and FMD	SVM	Classification of upper limb motions using myoelectric signals
Wheelchairs			
2008, Firoozabadi *et al.* [28]	MAV	SVM	Hands-free control system for operating a virtual wheelchair
2010, Tamura *et al.* [83]	Not indicated	Threshold algorithm	Hands-free control system for operating electric wheelchairs with facial muscles
2010, Wei, and Hu [89]	MAV, RMS, WL, ZC, FMN and FMD	SVM	Hands-free control of electric wheelchairs with forehead EMG signals and color face image information
Other applications			
2004, Jeong *et al.* [44]	IEMG, difference absolute mean value	Fuzzy min-max neural network	Using a computer by clenching teeth
2010, Gurmanik *et al.* [32]	AR	SVM	Differentiating neuromuscular disorders

Table 1: EMG applications.

6 Stage 4: Control Applications

In this section, a number of bio-control applications are outlined, in which output commands produced in the classification stage are fed to a robot or device such as wheelchairs, robotic arms or computers.

6.1 EMG non-invasive applications

Sörnmo and Laguna [80]; Oskoei and Hu [3] have shown, that some EMG non-invasive applications are related to:

1. kinesiology, since EMG can assist on the study of motor control strategies, mechanics of muscle contraction and gait;
2. ergonomics, as EMG provides a valuable, quantitative measure of muscle load, often used to asses physical load during work, therefore it can help to avoid work-related disorders and design better workplaces;
3. prosthesis control, inasmuch as the control signal is derived with surface electrodes placed over muscles or muscle groups under voluntary control within the residual limb ([39], [68], [18], [66]);
4. wheelchair controllers ([28], [83], [89]);
5. virtual keyboards ([44]) and f) diagnoses and clinical applications, such as functional neuromuscular stimulation ([32]) and detection of preterm births based on uterine myoelectric signals.

More details can be seen in Table 1.

6.2 EEG non-invasive applications

Some authors (Wolpaw *et al.* [91]; Sörnmo and Laguna [80]; Van Gerven *et al.* [86]) have outlined as EEG applications the following ones:

	Feature extraction	Classifier	Application
Mental tasks			
2002, Yom-Tov, and Inbar [92]	AR, PSD, Barlow, mean amplitude difference between every pair of recorded electrodes and standard deviation of the amplitude difference between every pair of recorded electrodes	Combination of a genetic algorithm and SVM	Classification of movement-related potentials recorded from the scalp
2004, Huan, and Palaniappan [38]	AR	BPNN and LDA	Two-state BCI from EEG signals extracted during mental tasks
Wheelchairs			
2005, Tanaka *et al.* [84]	Coefficient of correlation	Recursive training using Euclidean distance	Electroencephalogram-based control of an electric wheelchair
2007, Leeb *et al.* [51]	Logarithmic band power	Threshold algorithm	BCI control of a wheelchair in virtual environments
2008, Choi, and Cichocki [16]	Common spatial pattern	SVM	Control of a wheelchair by motor imagery in real time
Mental and neurological disorders			
2007, Sabeti *et al.* [75]	AR, band power, fractal dimension and wavelet energy	LDA and SVM	Selection of relevant features for EEG signal classification of schizophrenic patients
2010, Subasi, and Gursoy [81]	Mean of the absolute values and standard deviation of the coefficients in each sub-band, average power of the wavelet coefficients in each sub-band, ratio of the absolute mean values of adjacent sub-bands	SVM	Diagnostic decision support tool for physicians treating potential epilepsy

Table 2: EEG applications.

1. diagnosing mental disorders including epilepsy and schizophrenia ([75], [81]), also sleep disorders, such as insomnia, hypersomnia, circadian rhythm disorders and parasomnia [24];
2. monitoring mental tasks ([92], [38]) and
3. controlling spelling program, computer cursor for communication with the external world, video games, intelligent wheelchair ([84], [51], [16]), television, robotic arm or a neuroprosthesis that enables the multi-dimensional movements of a paralyzed limb.

More details can be seen in Table 2.

7 Conclusion

This technical report has provided an overview of what bio-control systems have to offer, in particular on EEG and EMG based control systems. The design of bio-control systems has four stages: data acquisition & segmentation, feature extraction, classification and control. It is clear that bio-control technologies will begin to converge to enhance our human-machine interaction. The technology is extremely useful for improving the quality of life of disabled and elderly people.

In the near future we will see highly robust and flexible bio-control systems, which are based on various bio-signals such as voice, muscle contractions, brain waves and gestures. These control systems will become increasingly simple and intuitive, and no training is required, namely plug and play. These bio-control systems will have ability to understand human intentions and emotions, and adept the dynamic changes in the real-world. It is no doubt that these big inventions will change our life style forever in the 21st century just as the computers did in the 20th century.

[24] Insomnia: disorders in initiating or maintaining sleep; Hypersomnia: disorders causing excessive sleep and somnolence; Circadian rhythm disorders: disorders in the sleep-wake schedule; Parasomnia: deviations in the normal sleep pattern [80].

References

[1] A.B. Ajiboye and R.F. Weir. A heuristic fuzzy logic approach to emg pattern recognition for multifunctional prosthesis control. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 13(3):280–291, 2005.

[2] E. Alpaydin. *Introduction to machine learning*. The MIT Press, 2004.

[3] M. Asghari Oskoei and H. Hu. Myoelectric control systems–A survey. *Biomedical Signal Processing and Control, Elsevier*, 2(4):275–294, 2007.

[4] R. Aufrichtigl, S.B. Pedersen, and P. Jennum. Adaptive segmentation of EEG signals. In *International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 453–454, 1991.

[5] S. Balakrishnama and A. Ganapathiraju. Linear discriminant analysis-a brief tutorial. *Institute for Signal and information Processing*, 1998.

[6] A. Bashashati, M. Fatourechi, R.K. Ward, and G.E. Birch. A survey of signal processing algorithms in brain–computer interfaces based on electrical brain signals. *Journal of Neural engineering*, 4:R32, 2007.

[7] J.V. Basmajian. Muscles alive. Their functions revealed by electromyography. *Academic Medicine*, 37(8):802, 1962.

[8] O. Bida. *Influence of electromyogram (EMG) amplitude processing in EMG-torque estimation*. PhD thesis, Worcester Polytechnic Institute, 2005.

[9] R. Biscay, M. Lavielle, A. González, I. Clark, and P. Valdés. Maximum a posteriori estimation of change points in the EEG. *International journal of bio-medical computing, Elsevier*, 38(2):189–196, 1995.

[10] S.T. Bow. *Pattern recognition and image preprocessing*. New York Marcel Dekker, 2002.

[11] J.D. Bronzino. *The biomedical engineering handbook*. CRC Pr I Llc, 2000.

[12] N. Bu, M. Okamoto, and T. Tsuji. A hybrid motion classification approach for emg-based human–robot interfaces using bayesian and neural networks. *Robotics, IEEE Transactions on*, 25(3):502–511, 2009.

[13] A.D.C. Chan and K.B. Englehart. Continuous myoelectric control for powered prostheses using hidden markov models. *Biomedical Engineering, IEEE Transactions on*, 52(1):121–124, 2005.

[14] S. Chandaka, A. Chatterjee, and S. Munshi. Cross-correlation aided support vector machine classifier for classification of EEG signals. *Expert Systems with Applications, Elsevier*, 36(2):1329–1336, 2009.

[15] W.A. Chaovalitwongse, Y.J. Fan, and R.C. Sachdeo. On the time series k-nearest neighbor classification of abnormal brain activity. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 37(6):1005–1016, 2007.

[16] K. Choi and A. Cichocki. Control of a wheelchair by motor imagery in real time. *Intelligent Data Engineering and Automated Learning–IDEAL 2008, Springer*, pages 330–337, 2008.

[17] C.I. Christodoulou and C.S. Pattichis. Unsupervised pattern recognition for the classification of EMG signals. *Biomedical Engineering, IEEE Transactions on*, 46(2):169–178, 1999.

[18] J.U. Chu, I. Moon, and M.S. Mun. A real-time EMG pattern recognition system based on linear-nonlinear feature projection for a multifunction myoelectric hand. *Biomedical Engineering, IEEE Transactions on*, 53(11):2232–2239, 2006.

[19] R.R. Coifman and M.V. Wickerhauser. Entropy-based algorithms for best basis selection. *Information Theory, IEEE Transactions on*, 38(2):713–718, 1992.

[20] B. Crawford, K. Miller, P. Shenoy, and R. Rao. Real-time classification of electromyographic signals for robotic control. volume 20, page 523. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2005.

[21] C.J. De Luca. Surface electromyography: Detection and recording. *DelSys Incorporated*, 2002.

[22] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern classification*, volume 2. Wiley New York:, 2001.

[23] I. Dvořák and A.V. Holden. *Mathematical approaches to brain functioning diagnostics*. Manchester Univ Pr, 1991.

[24] B. E. Brodsky, B. S. Darkhovsky, A. Ya. Kaplan, and S. L. Shishkin. A nonparametric method for the segmentation of the eeg. *Computer methods and programs in biomedicine, Elsevier*, 60(2):93–106, 1999.

[25] K. Englehart. *Signal representation for classification of the transient myoelectric signal*. PhD thesis, University of New Brunswick, 1998.

[26] K. Englehart and B. Hudgins. A robust, real-time control scheme for multifunction myoelectric control. *Biomedical Engineering, IEEE Transactions on*, 50(7):848–854, 2003.

[27] K. Englehart, B. Hudgins, P.A. Parker, and M. Stevenson. Classification of the myoelectric signal using time-frequency based representations. *Medical engineering & physics, Elsevier*, 21(6-7):431–438, 1999.

[28] S.M.P. Firoozabadi, M.A. Oskoei, and H. Hu. A Human-Computer Interface based on Forehead Multi-channel Bio-signals to control a virtual wheelchair. In *Proceedings of the 14th Iranian Conference on Biomedical Engineering (ICBME)*, pages 272–277. Citeseer, 2008.

[29] K. Fukunaga. *Introduction to statistical pattern recognition*. Academic Press Professional, 1990.

[30] D. Gabor. Theory of communication. Part 1: The analysis of information. *Electrical Engineers-Part III: Radio and Communication Engineering, Journal of the Institution of*, 93(26):429–441, 1946.

[31] S. Günes, M. Dursun, K. Polat, and S. Yosunkaya. Sleep spindles recognition system based on time and frequency domain features. *Expert Systems with Applications, Elsevier*, 2010.

[32] K. Gurmanik, A.S. Arora, and V.K. Jain. EMG Diagnosis via AR Modeling and Binary Support Vector Machine Classification. *International Journal of Engineering Science and Technology*, 2, 2010.

[33] R. Gut and G.S. Moschytz. High-precision EMG signal decomposition using communication techniques. *Signal Processing, IEEE Transactions on*, 48(9):2487–2494, 2000.

[34] S. Halder, M. Bensch, J. Mellinger, M. Bogdan, A. Kübler, N. Birbaumer, and W. Rosenstiel. Online artifact removal for brain-computer interfaces using support vector machines and blind source separation. *Computational Intelligence and Neuroscience*, 2007:1–9, 2007.

[35] J.S. Han, W.K. Song, J.S. Kim, W.C. Bang, H. Lee, and Z. Bien. New EMG pattern recognition based on soft computing techniques and its application to control of a rehabilitation robotic arm. In *Proc. of 6th International Conference on Soft Computing (IIZUKA2000)*, pages 890–897, 2000.

[36] S.S. Haykin. *Neural networks and learning machines*, volume 10. Prentice Hall Upper Saddle River, NJ, 2009.

[37] M. Hollander and D.A. Wolfe. Nonparametric statistical methods. *New York: John Wiley & Sons*, 1973.

[38] N.J. Huan and R. Palaniappan. Neural network classification of autoregressive features from electroencephalogram signals for brain–computer interface design. *Journal of neural engineering*, 1:142, 2004.

[39] H.P. Huang and C.Y. Chen. Development of a myoelectric discrimination system for a multi-degree prosthetic hand. In *Robotics and Automation, 1999. Proceedings. 1999 IEEE International Conference on*, volume 3, pages 2392–2397. IEEE, 1999.

[40] Z. Iscan, Z. Dokur, and T. Demiralp. Classification of electroencephalogram signals with combined time and frequency features. *Expert Systems with Applications, Elsevier*, 2011.

[41] A.K. Jain, R.P.W. Duin, and J. Mao. Statistical pattern recognition: A review. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(1):4–37, 2000.

[42] C.J. James, R.D. Jones, P.J. Bones, and G.J. Carroll. Detection of epileptiform discharges in the eeg by a hybrid system comprising mimetic, self-organized artificial neural network, and fuzzy logic stages. *Clin. Neurophysiol*, 110(12):2049–2063, 1999.

[43] H.H. Jasper. The ten twenty electrode system of the International Federation. *Electroencephalography and clinical neurophysiology*, 10:371–375, 1958.

[44] H. Jeong, J.S. Kim, and J.S. Choi. A Study of an EMG-controlled HCI Method by Clenching Teeth. In *Computer Human Interaction*, pages 163–170. Springer, 2004.

[45] A. Kaplan, J. Roschke, B. Darkhovsky, and J. Fell. Macrostructural EEG characterization based on nonparametric change point segmentation: application to sleep analysis. *Journal of neuroscience methods, Elsevier*, 106(1):81–90, 2001.

[46] A.Y. Kaplan and S.L. Shishkin. Application of the change-point analysis to the investigation of the brain's electrical activity. *Non-Parametric Statistical Diagnosis: Problems and Methods*, pages 333–388, 2000.

[47] B. Karlik. Differentiating type of muscle movement via ar modeling and neural network classification. *Turk J Elec Engin*, 7(1-3), 1999.

[48] G. Kaur, A.S. Arora, and VK Jain. Comparison of the techniques used for segmentation of EMG signals. In *Proceedings of the 11th WSEAS international conference on Mathematical and computational methods in science and engineering*, pages 124–129. World Scientific and Engineering Academy and Society (WSEAS), 2009.

[49] K. Kiguchi, T. Tanaka, and T. Fukuda. Neuro-fuzzy control of a robotic exoskeleton with emg signals. *Fuzzy Systems, IEEE Transactions on*, 12(4):481–490, 2004.

[50] D.H. Kil and F.B. Shin. *Pattern recognition and prediction with applications to signal characterization*. Amer Inst of Physics, 1996.

[51] R. Leeb, D. Friedman, G.R. Müller-Putz, R. Scherer, M. Slater, and G. Pfurtscheller. Self-paced (asynchronous) BCI control of a wheelchair in virtual environments: a case study with a tetraplegic. *Computational intelligence and neuroscience*, 2007:1–12, 2007.

[52] F. Lotte. The use of fuzzy inference systems for classification in eeg-based brain-computer interfaces. In *Proc. of the 3rd international Brain-Computer Interface workshop*, pages 12–13. Citeseer, 2006.

[53] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi. A review of classification algorithms for EEG-based brain–computer interfaces. *Journal of neural engineering*, 4:R1, 2007.

[54] M.F. Lucas, A. Gaufriau, S. Pascual, C. Doncarli, and D. Farina. Multi-channel surface emg classification using support vector machines and signal-based wavelet optimization. *Biomedical Signal Processing and Control, Elsevier*, 3(2):169–174, 2008.

[55] J. Malmivuo and R. Plonsey. *Bioelectromagnetism: principles and applications of bioelectric and biomagnetic fields.* Oxford University Press, USA, 1995.

[56] D.J. McFarland, L.M. McCane, S.V. David, and J.R. Wolpaw. Spatial filter selection for EEG-based communication. *Electroencephalography and clinical Neurophysiology, Elsevier*, 103(3):386–394, 1997.

[57] R. Merletti and P.A. Parker. *Electromyography: Physiology, engineering, and noninvasive applications*. Wiley-IEEE Press, 2004.

[58] T.M. Mitchell. *Machine learning.* Mac Graw Hill, 1997.

[59] D. Novák, YHT Al-Ani, and L. Lhotska. Electroencephalogram processing using hidden markov models. *Report Transdisciplinary Biomedical Engineering Research*, 2003.

[60] D. Novák, L. Lhotská, V. Eck, and M. Sorf. EEG and VEP signal processing. *Cybernetics, Faculty of Electrical Eng*, pages 50–53, 2004.

[61] P.L. Nunez. *Electric fields of the brain: the neurophysics of EEG*. Oxford University Press, New York, 1981.

[62] B. Obermaier, GR Muller, and G. Pfurtscheller. Virtual keyboard controlled by spontaneous eeg activity. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 11(4):422–426, 2003.

[63] M. Okamoto, Y. Matsubara, K. Shima, and T. Tsuji. Emg pattern classification using hierarchical network based on boosting approach. *International Journal of innovative computing, information and control*, 5(12(B)):4935–4943, 2009.

[64] A.H. Omidvarnia, F. Atry, S.K. Setarehdan, and B.N. Arabi. Kalman filter parameters as a new EEG feature vector for BCI applications. In *Proceedings of the 13th European Signal Processing Conference Eusipco2005*. Citeseer, 2005.

[65] M.A. Oskoei and H. Hu. GA-based feature subset selection for myoelectric classification. In *Proc. Int. Conf. Robot. Biomimetics*, pages 1465–1470. Citeseer, 2006.

[66] M.A. Oskoei and H. Hu. Support vector machine-based classification scheme for myoelectric control applied to upper limb. *Biomedical Engineering, IEEE Transactions on*, 55(8):1956–1965, 2008.

[67] Ramaswamy Palaniappan. *Biological Signal Analysis*. BookBoon, 2010.

[68] D. Peleg, E. Braiman, E. Yom-Tov, and G.F. Inbar. Classification of finger activation for use in a robotic prosthesis arm. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 10(4):290–293, 2002.

[69] W. Penfield and T. Rasmussen. *The cerebral cortex of man: a clinical study of localization of function*. Macmillan, 1950.

[70] J. Perry and G.A. Bekey. EMG-force relationships in skeletal muscle. *Crit Rev Biomed Eng*, 7(1):1–22, 1981.

[71] A. Phinyomark, C. Limsakul, and P. Phukpattaranont. A novel feature extraction for robust EMG pattern recognition. *Journal of Computing*, 1(1):71–80, 2009.

[72] J.G. Proakis and D.G. Manolakis. *Digital signal processing: principles, algorithms, and applications*, volume 3. Prentice Hall Upper Saddle River, NJ:, 1996.

[73] M.B.I. Reaz, MS Hussain, and F. Mohd-Yasin. Techniques of EMG signal analysis: detection, processing, classification and applications. *Biological procedures online, Springer*, 8(1):11–35, 2006.

[74] S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall, 2003.

[75] M. Sabeti, R. Boostani, SD Katebi, and GW Price. Selection of relevant features for EEG signal classification of schizophrenic patients. *Biomedical Signal Processing and Control, Elsevier*, 2(2):122–134, 2007.

[76] Y. Si, J. Gotman, A. Pasupathy, D. Flanagan, B. Rosenblatt, and R. Gottesman. An expert system for eeg monitoring in the pediatric intensive care unit. *Electroencephalography and clinical neurophysiology, Elsevier*, 106(6):488–500, 1998.

[77] S. Siegel. Nonparametric statistics. *The American Statistician*, 11(3):13–19, 1957.

[78] SN Sivanandam, S. Sumathi, and SN Deepa. *Introduction to fuzzy logic using MATLAB*. Springer Verlag, 2007.

[79] S. Solhjoo, A.M. Nasrabadi, and M.R.H. Golpayegani. Classification of chaotic signals using hmm classifiers: Eeg-based mental task classification. In *Proceedings of the European Signal Processing Conference*, 2005.

[80] L. Sörnmo and P. Laguna. *Bioelectrical signal processing in cardiac and neurological applications*. Elsevier. Academic Press, 2005.

[81] A. Subasi and M.I. Gursoy. EEG signal classification using PCA, ICA, LDA and support vector machines. *Expert Systems With Applications, Elsevier*, 2010.

[82] A. Subasi, M. Yilmaz, and H.R. Ozcalik. Classification of emg signals using wavelet neural network. *Journal of neuroscience methods, Elsevier*, 156(1-2):360–367, 2006.

[83] H. Tamura, T. Manabe, T. Goto, Y. Yamashita, and K. Tanno. A Study of the Electric Wheelchair Hands-Free Safety Control System Using the Surface-Electromygram of Facial Muscles. *Intelligent Robotics and Applications, Springer*, pages 97–104, 2010.

[84] K. Tanaka, K. Matsunaga, and H.O. Wang. Electroencephalogram-based control of an electric wheelchair. *Robotics, IEEE Transactions on*, 21(4):762–766, 2005.

[85] T. Tsuji, N. Bu, O. Fukuda, and M. Kaneko. A recurrent log-linearized gaussian mixture network. *Neural Networks, IEEE Transactions on*, 14(2):304–316, 2003.

[86] M. van Gerven, J. Farquhar, R. Schaefer, R. Vlek, J. Geuze, A. Nijholt, N. Ramsey, P. Haselager, L. Vuurpijl, S. Gielen, et al. The brain-computer interface cycle. *Journal of Neural Engineering*, 6(4):1–10, 2009.

[87] C. Vidaurre, N. Kramer, B. Blankertz, and A. Schlogl. Time domain parameters as a feature for EEG-based brain-computer interfaces. *Neural Networks, Elsevier*, 22(9):1313–1319, 2009.

[88] M. Weeks. *Digital signal processing using MATLAB and wavelets*. Jones and Bartlett Publishers, LLC, second edition, 2011.

[89] L. Wei and H. Hu. EMG and visual based HMI for hands-free control of an intelligent wheelchair. In *Intelligent Control and Automation (WCICA), 2010 8th World Congress on*, pages 1027–1032. IEEE, 2010.

[90] G. Welch and G. Bishop. An introduction to the Kalman filter. *University of North Carolina at Chapel Hill, Chapel Hill, NC*, 7(1), 1995.

[91] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, and T.M. Vaughan. Brain-computer interfaces for communication and control. *Clinical neurophysiology, Elsevier*, 113(6):767–791, 2002.

[92] E. Yom-Tov and G.F. Inbar. Feature selection for the classification of movements from single movement-related potentials. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 10(3):170–177, 2002.

[93] M. Zecca, S. Micera, MC Carrozza, and P. Dario. Control of multifunctional prosthetic hands by processing the electromyographic signal. *Critical Reviews in Biomedical Engineering, Citeseer*, 30(4-6):459, 2002.

[94] S.M. Zhou, J.Q. Gan, and F. Sepulveda. Classifying mental tasks based on features of higher-order statistics from eeg signals in brain-computer interface. *Information Sciences, Elsevier*, 178(6):1629–1640, 2008.