Chapter 2

Theoretical Background

Modern medicine applies variety of imaging techniques of the human body. The group of electrobiological measurements comprises items as Electrocardiography (ECG, heart), Electromyography (EMG, muscular contractions), Electroencephalography (EEG, brain), Magnetoencephalography (MEG, brain), Electrogastrography (EGG, stomach), Electrooptigraphy (EOG, eye dipole field) [TEP02]. This research discusses the design of a system to analyze, detect and classify abnormalities in the brain known as Epileptic seizure using EEG signals.

In this chapter we present the main concepts related to Epilepsy and EEG and its characteristics. A review of Wavelet Transforms, including a mathematical description, and decomposition and reconstruction processes carried out are also presented here. This chapter also presents a review of Wavelet Neural Networks.

2.1 Epilepsy

Epilepsy is one of the most common neurological disorders and occurs with an incidence of 68.8/100,000 person-years [SRI14]. It is a heterogeneous condition characterized by recurrent spontaneous epileptic seizures. There are no age, sex or race discriminations, although incidence of the first epileptic seizure has a bimodal dis-

tribution, with peaks in childhood and after 60 years of age [ENG08].

Epilepsy is a neurological disorder and is defined as a symptom where a sudden and transient disturbance occurs in the normal electrical activity of the brain. Epileptic seizures are manifestations of epilepsy; an epileptic seizure is caused by the massive synchronization of neuronal electrical activity. During the seizure, groups of neurons discharge synchronously, creating a large amplitude signal and leading to uncontrollable oscillations. Multiples factors such as tumors, infections, brain injury, disease, trauma, light stimulation, genetics or metabolic and toxic disorders may be responsible for the synchronized discharges causing the epileptic seizures [BLI06]. These seizures are seen as a sudden abnormal function of the body, often with loss of consciousness, drop attacks, an increase in muscular activity, facial muscles and eye movements, aggressive outbursts, prolonged confusional states, and flexor spasms of a whole body [SUN12]. Seizure types can be divided into three main categories [BLI06]:

1. *Local.* The synchronized electrical activity starts in a well localized zone of the brain. The seizure, lasting a few seconds, is accompanied by jerking or spasms, as well as by a loss of consciousness.

2. *Generalized.* The EEG patterns are bilaterally symmetrical and roughly synchronous; the epileptic activity is spread over wide areas of both hemispheres simultaneously from the onset of attack.

3. *Unclassifiable.* Different from those described above.

EEG is very sensitive to the action of a wide range of pharmacological substances, especially psychotropic drugs, anaesthetics, and anticonvulsants. It is also affected by some drugs targeted to organs other than the central nervous system (CNS), such as antihistamines and antihypertensives. Influence of drugs on EEG primarily include changes in its spectral content and topographic characteristics. Effects of psychoactive drugs on EEG could be used to assess their action on the CNS

[BLI06]. Despite the introduction of new anti-epileptic drugs in the last decades, one-four of people with epilepsy continue to have seizures despite treatment. However, even when seizures are well controlled, self-reported quality of life is significantly lowered by the anxiety associated with the unpredictable nature of seizures and the consequences therefrom. Some of the difficulties in managing treatment-refractory epilepsy can be ameliorated by the ability to detect clinical seizures. This information might be useful both in developing accurate seizure diaries and in providing therapies during times of greatest seizure susceptibility. The ability to rapidly and accurately detect seizures could promote therapies aimed at rapidly treating seizures. The capability to detect seizures early and anticipate their onset prior to presentation would provide even greater advantages [SRI14]. An electroencephalogram has been an especially valuable clinical tool for the evaluation an treatment of epilepsy.

2.1.1 Electroencephalogram (EEG)

Electroencephalography is the neurophysiologic measurement of the electrical activity of the brain by recording from electrodes placed on the scalp, or in the special cases on the cortex. The resulting traces are known as an electroencephalogram (EEG) and represent so-called brainwaves. In neurology, the main diagnostic application of EEG is in the case of epilepsy, as epileptic activity can create clear abnormalities on a standard EEG study. A secondary clinical use of EEG is in the diagnosis of coma, encephalopathies, and brain death. EEG used to be a first-line method for the diagnosis of tumors, stroke and other focal brain disorders, but this use has decreased with the advent of anatomical imaging techniques such as Magnetic Resonance Imaging (MRI) and Computed Tomography (CT). Some researchers claim that the EEG can be used to predict abnormal development and aid in the evaluation of nonspecific symptoms such as behavioral disorders, anxiety, or learning disabilities.

EEG is a record of the electric signal generated by the cooperative action of

brain cells, or more precisely, the time course of extracellular field potentials generated by their synchronous action. Electroencephalogram derives from the Greek words *enkephalo (brain)* and *graphein (to write)*. EEG can be measured by means of electrodes placed on the scalp or directly on the cortex. In the latter case, it is called electrocorticogram (ECoG). Electric fields measured intracortically were named Local Fields Potentials (LFP). EEG recorded in the absence of an external stimulus is called spontaneous EEG; EEG generated as a response to external or internal stimulus is called an event-related potential (ERP). The amplitude of EEG of a normal subject in the awake state recorded with the scalp electrodes is 10-100 mV. In case of epilepsy, the EEG amplitudes may increase by almost one order of magnitude. In the cortex, amplitudes are in the range 500-1500 mV [BLI06].

2.1.2 Recording EEG

EEG is usually registered by means of electrodes placed on the scalp. They can be secured by an adhesive (like collodion) or embedded in a special snug cap. The resistance of the connection should be less than 5 KΩ, so the recording site is first cleaned with diluted alcohol, and conductive electrode paste applied to the electrode cup. Knowledge of exact positions of electrodes is very important for both interpretation of a single recording as well as comparison of results, hence the need for standardization [BLI06].

The traditional 10-20 electrode system states positions of 19 EEG electrodes (plus two electrodes placed on earlobes A1/A2) related to specific anatomic landmarks, such that 10-20 % of the distances between them are used as the electrode interval (see Figure 2.1 and Figure 2.2). The first part of derivation's name indexes the array's row-from the front of head: Fp, F, C, P, and O. The second part is defined with even numbers to the left and odd numbers to the right side, with identifier "z" located in the center. Progress in topographic representation of EEG recordings brought demand for a larger amount of derivations. Electrode sites halfway between

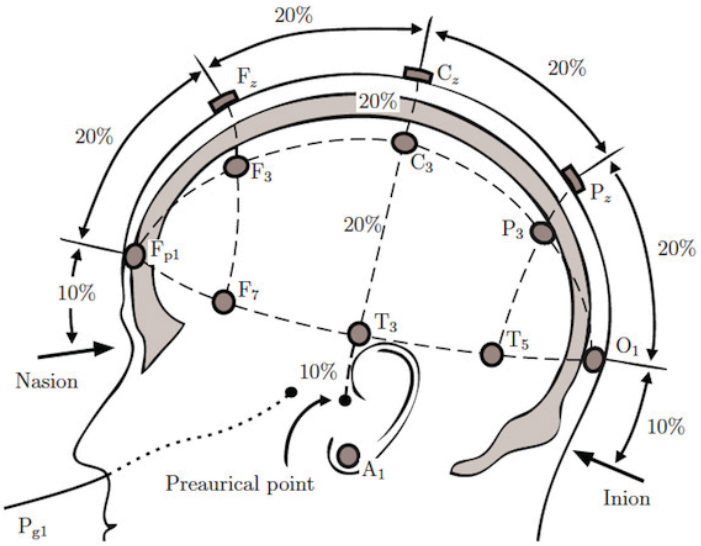

Figure 2.1: A sagittal view of electrode placement according to the international 10-20 system [TAN11].

those defined by the standard 10-20 system were introduced in the extended 10-20 system [BLI06].

Electrical signals detected along the scalp by an EEG contains useful information along with redundant or noise information. The information that originate from non-cerebral signals are called artifacts. EEG signal is almost always contaminated by such artifacts. The amplitude of artifacts can be quite large with respect to the size of amplitude of the cortical signals of interest. This is one of the reasons why it takes considerable experience to correctly interpret EEGs clinically. Some of the most common types of biological artifacts include:

- Eye-induced artifacts (includes eye blinks, eye movements and extra-ocular muscle activity).
- Electrocardiogram artifacts.
- Muscle activation-induced artifacts.
- Glossokinetic artifacts.

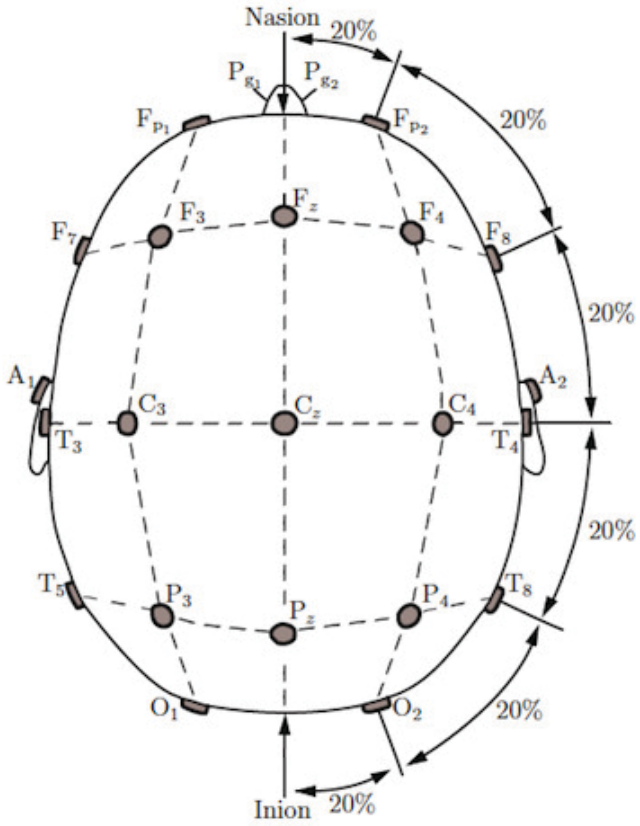

Figure 2.2: An axial view of electrode placement according to the international 10-20 system [TAN11].

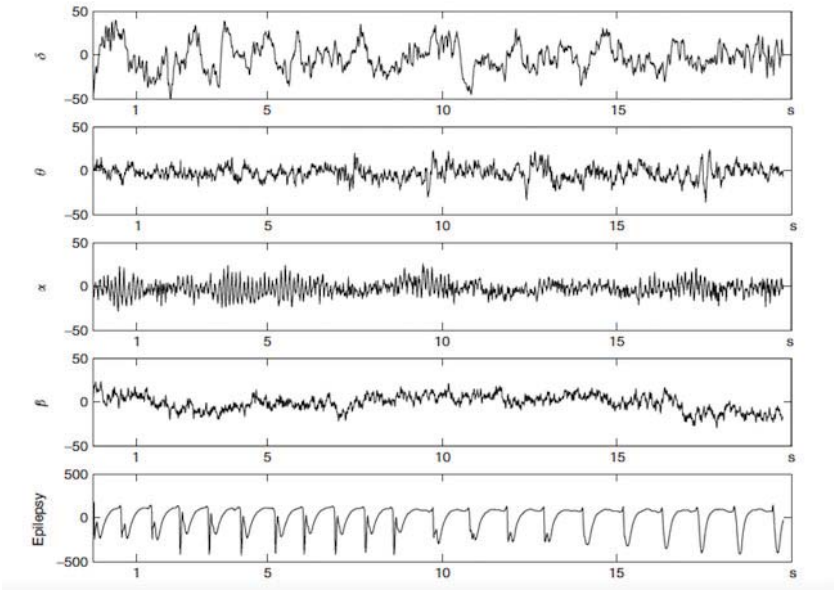

Figure 2.3: Characteristic EEG rhythms, from the top: δ (0-4 Hz), θ (4-8 Hz), α (8-12 Hz), β (12-30 Hz). The lowest trace-EEG during epileptic seizure, note that the amplitude scale is an order of magnitude bigger than non-epileptic states [BLI06].

The EEG signals used in the experiments reported in this thesis are from a free-available EEG database provided by the University of Bonn [BON16], [AND01]. In this context, a filtering of these EEG signals was performed to remove unwanted frequencies and noise added during recording (see Section 3.3).

2.1.3 EEG Sub-bands

An EEG is typically described in terms of rhythmic activity or sub-bands and transients. The rhythmic activity is divided into bands by frequency, the following rhythms have been distinguished in EEG: Delta (δ) of 0-4 Hz, Theta (θ) in 4-8 Hz, Alpha (α) in 8-12 Hz, Beta (β) in 12-30 Hz and Gamma (γ) above 30 Hz. These sub-bands are shown in Figure 2.3.

Gamma components are difficult to record by scalp electrodes because their frequency does not exceed 45 Hz; ECoG components, may register up to 100 Hz,

or even higher. The contribution of different rhythms to the EEG depends on the age and behavioral state of the subject, mainly the level of alertness. Considerable inter-subject differences in EEG characteristics also exist. EEG pattern is influenced by neuro-pathological conditions, metabolic disorders, and drug action [BLI06].

Each frequency band is associated with specific characteristics. **Delta** (δ) sub-band is a predominant feature in EEGs recorded during deep sleep. In this stage, delta waves usually have large amplitudes (75-200 mV) and show strong coherence all over the scalp; the frequency range of this sub-band is of 0-4 Hz [BLI06]. **Theta** (θ) waves rarely occur in adult humans. In this case, the frequency range is of 4-8 Hz and waves have a high amplitude and characteristic sawtooth shape. The activity in the theta sub-band may occur in emotional or some cognitive states; it can be also connected with the slowing of alpha rhythms caused by a pathology [BLI06]. **Alpha** (α) sub-band is predominant during wakefulness and it is most pronounced in the posterior regions of the head. These waves are best observed when the eyes are closed and the subject is in a relaxed state. They are blocked or attenuated by attention (especially visual) and by mental effort; the frequency range of this sub-band is of 8-12 Hz. **Mu** rhythms have a frequency band similar to alpha, but their topography and physiological significance are different each other. They are related to the function of motor cortex and they are prevalent in the central part of the head. Mu rhythms are blocked by motor functions [BLI06]. **Beta** (β) activity is characteristic for the states of increased alertness and focused attention; the frequency range of this sub-band is of 12-30 Hz [BLI06]. **Gamma** (γ) activity is connected with information processing, for example recognition of sensory stimuli, and the onset of voluntary movements. In general, it can be summarized that the slowest cortical rhythms are related to an idle brain and the fastest to information processing; Gamma waves show frequencies above 30 Hz [BLI06]. A summary of the characteristics of these sub-bands is shown in Table 2.1.

In this thesis, the decomposition of EEG into sub-bands was carried out using the popular methods of Discrete Wavelet Transform (DWT) and the Maximal

Table 2.1: Frequency sub-bands from an EEG signal.

Name	Frequency range (Hz)	Subject's conditions
Delta (δ)	0 - 4	Deep sleep, infancy, brain diseases.
Tetha (θ)	4 - 8	Parental and temporal regions in children, emotional stress in adults.
Alpha (α)	8-12	Awake but in quiet, resting and relaxed condition in healthy subjects.
Beta (β)	12 - 30	Intense mental activity (thinking, attention or problem solving).
Gamma (γ)	> 30	Information processing and the onset of voluntary movements.

Overlap Discrete Wavelet Transform (MODWT) (see Section 3.4). Delta (δ) and alpha (α) sub-bands were used to characterize the EEG, because according to the researches Sunhaya, S. et al. [SUN12], and Ravish, D. and Devi S. [RAV12], the rhythmic activity associated with the onset of seizure activity is composed of strong frequency components of these sub-bands. Next, we briefly describe these methods (see Sections 2.2.4 and 2.2.5).

2.2 Wavelet Transforms

In recent years, time-frequency domain analysis has been increasingly used for feature extraction of epileptic EEG. Wavelet Transforms have emerged as an important mathematical tool for EEG analysis. Wavelet transform can be used for analyzing signals in different sub-bands in a selective way, which is suitable for extracting epileptic characteristics and increases detection performance of the system. Contrary to Fourier transform, wavelet transform supplies a more flexible approach of time-frequency representation of a signal, using analysis windows with different sizes. An important characteristic of wavelet transform is that it supplies precise time information at high frequencies and precise frequency information at low frequencies.

This characteristic is of great importance, since signals in biomedical applications usually include low frequency information with long time duration and high frequency information with short time duration. By means of wavelet transform, transient characteristics are accurately captured and it is localized in both time and frequency domains [YUE11].

2.2.1 The Fourier Transform (FT)

Traditional mathematical tools study the statistical properties of a signal either in the time domain or in the frequency domain. For example, the Fourier Transform (FT) is a classical method for the analysis of stationary signals, that is, signals whose statistical properties do not evolve in time. The FT describes the signal $x(s)$ with a basis consisting of sines and cosines of different frequencies [ALA03]. The coefficients of these basis functions are computed by taking the inner product in $L^2(\mathbb{R})$ (set of square integrable real-valued functions) of the signal with sine wave basis functions of infinite duration, namely,

$$FT(f) = \int_{-\infty}^{+\infty} x(s)e^{-i2\pi fs}ds, \tag{2.1}$$

where $x(s)$ denotes the signal and $FT(f)$ is its Fourier Transform. The Fourier transform works well if $x(s)$ is composed of a few stationary components. However, any abrupt change in time in a non-stationary signal $x(s)$ is spread out over the whole frequency axis in $FT(f)$ [ALA03]. To overcome these drawbacks, the FT in Equation (2.1) has been slightly modified yielding the Short-Time Fourier Transform (STFT). The STFT applies a moving window over the data to determine the localized spectra. The signal $x(s)$ is restricted to an interval by multiplying it by a fixed window function, then the Fourier analysis of the product is carried out, namely,

$$STFT(t, f) = \int_{-\infty}^{+\infty} x(s)g^*(s-t)e^{-i2\pi fs}ds, \tag{2.2}$$

where $*$ denotes complex conjugate and $g(s)$ is known as the analysis window. The window $g(s-t)$ is a localized function which is shifted over time axis to compute the transform at several positions t. The STFT has become time dependent and thus forms a time-frequency domain description of a signal. A drawback of the STFT is that the time-frequency window is fixed whereby the window cannot adapt to the characteristics of the signal at a certain point [ALA03].

2.2.2 The Continuous Wavelet Transform (CWT)

The continuous wavelet transform (CWT) follows the ideas described earlier. The CWT of a signal $x(s) \in L^2(\mathbb{R})$ is defined as

$$CWT(t,f) = \sqrt{\left|\frac{f}{f_0}\right|} \int_{-\infty}^{+\infty} x(s)\psi^* \left(\frac{f}{f_0}(s-t)\right) ds, \tag{2.3}$$

where the analyzing wavelet $\psi(s)$ corresponds to the window function $g(s)$ of the STFT, and f_0 denotes the center frequency of the analyzing wavelet [ALA03]. Let $d = f_0/f$ denote a scale factor. Then, Equation (2.3) may be expressed as:

$$CWT(t,d) = \langle x(s), \psi_{d,t}(s)\rangle = \frac{1}{\sqrt{|d|}} \int_{-\infty}^{+\infty} x(s)\psi^* \left(\frac{1}{d}(s-t)\right) ds, \tag{2.4}$$

with $d \in \mathbb{R}^+, t \in \mathbb{R}, d \neq 0$, where d denotes the scale (or dilation) parameter, t denotes the translation parameter and $\langle \cdot, \cdot \rangle$ denotes the L^2-inner product [ALA03]. The wavelet transform is referred to as time-scale representation, as the frequency f act as part of the scaling process. The wavelets are thus scaled and translated version of the *mother wavelet* denoted by $\psi(s) \in L^2(\mathbb{R})$ [ALA03].

The wavelet analysis results in a set of coefficients, which indicate how close the signal is to a particular basis function. Such analysis is done by using Equation (2.4), whereas the synthesis consists of summing up all the orthogonal projections of the signal onto the wavelets [DAU92]:

$$x(s) = C_\psi^{-1} \int_0^{+\infty} \int_{-\infty}^{+\infty} CWT(t,d) \frac{1}{\sqrt{|d|}} \psi_{d,t}^* \left(\frac{s-t}{d}\right) \frac{dd\, dt}{d^2}, \tag{2.5}$$

where C_ψ denotes an *admissibility* constant that depends only on $\psi(s)$. From Equation (2.5), $x(s)$ can be described as an integration of the basis functions $\psi_{d,t}(s)$ with weights $CWT(t,d)$. The *admissibility condition* for a function $\psi(s)$, with Fourier transform $\Psi(f) = \int \psi(s)e^{-i2\pi fs}ds$, is given by [DAU92]

$$C_\psi = \int_{-\infty}^{+\infty} \frac{|\Psi(f)|^2}{f} df < +\infty \tag{2.6}$$

Equation (2.6) implies that $\Psi(f)$ should decay sufficiently fast as $f \to \infty$ and $\Psi(f)$ should be sufficiently flat near $f = 0$. A wavelet $\psi(s)$ is thus considered to be a band-pass window function since

$$\Psi(0) = 0 \Leftrightarrow \int_{-\infty}^{+\infty} \psi(s)ds = 0 \tag{2.7}$$

The wavelets thus must be oscillatory and have amplitudes, which quickly decay to zero in both positive and negative directions. These conditions must be simultaneously satisfied for the function to be a *small wave* or *wavelet*. The above property is related to the number of vanishing moments of $\psi(s)$. The vanishing moment is a criterion about how a function decays toward infinity. A wavelet is said to have v vanishing moments if [MAL99]:

$$\left.\frac{d^k\Psi(f)}{df^k}\right|_{f=0} = \int_{-\infty}^{+\infty} s^k\psi(s)ds = 0, \qquad \text{for } k = 0, 1, ..., v \text{ and } k \in \mathbb{Z} \tag{2.8}$$

The parameter k indicates the rate of decay. Equation (2.8) states a connection between the smoothness (differentiability) of $\psi(s)$ and the number of vanishing moments [DAU92], [MAL99].

2.2.3 Translation and Scaling Parameters

The continuous wavelet transform considers a family

$$\psi_{d,t}(s) = \frac{1}{\sqrt{|d|}}\psi\left(\frac{s-t}{d}\right), \tag{2.9}$$

where $d \in \mathbb{R}^+$, $t \in \mathbb{R}$, with $d \neq 0$, and $\psi(\cdot)$ satisfies the *admissibility condition*. Since two scales $d_0 < d_1$ roughly correspond to two frequencies $f_0 > f_1$, a natural way to

discretize the translation and scaling parameters is by sub-sampling the coefficients at scale d_1 at $(f_0/f_1)^{th}$ the rate of the coefficients at scale d_0 according to Nyquist's rule [DAU92]. For discrete wavelets the scale (or dilation) and translation parameters in Equation (2.9) are chosen such that at level j the wavelet $d_0^j\,\psi(d_0^{-j}\,s)$ is d_0^j times the width of $\psi(s)$. That is, the scaling parameter is $\{d = d_0^j \ : \ j \in \mathbb{Z}\}$ and the translation parameter is $\{t = n\,t_0\,d_0^j \ : \ j, n \in \mathbb{Z}\}$. This family of wavelets is thus given by

$$\psi_{j,n}(s) = d_0^{-j/2}\psi\left(d_0^{-j}\,s - n\,t_0\right), \tag{2.10}$$

so the discrete version of a wavelet transform is

$$d_{j,n} = \langle x(s), \psi_{j,n}(s)\rangle = d_0^{-j/2}\int_{-\infty}^{+\infty} x(s)\psi\left(d_0^{-j}\,s - n\,t_0\right)\,ds, \tag{2.11}$$

where $\langle\cdot,\cdot\rangle$ denotes the L^2-inner product. To recover $x(s)$ from coefficients $\{d_{j,n}\}$, the following stability condition should exist [DAU92],

$$A\|x(s)\|^2 \leq \sum_{j\in\mathbb{Z}}\sum_{n\in\mathbb{Z}} |\langle x(s), \psi_{j,n}(s)\rangle|^2 \leq B\|x(s)\|^2, \tag{2.12}$$

with $A > 0$ and $B < \infty$ for all signals $x(s)$ in $L^2(\mathbb{R})$ denoting the frame bounds. These frame bounds can be computed from d_0, t_0 and $\psi(s)$ [DAU92]. The reconstruction formula is thus given by

$$x(s) \approx \frac{2}{A+B}\sum_{j\in\mathbb{Z}}\sum_{n\in\mathbb{Z}} \langle x(s), \psi_{j,n}(s)\rangle \tag{2.13}$$

Note that the closer A and B, the more accurate the construction. When $A = B = 1$, the family of wavelets then forms an orthonormal basis [DAU92].

2.2.4 The Discrete Wavelet Transform (DWT)

The idea of multiresolution analysis is to write a signal $x(s)$ as a limit of successive approximations; the differences between two successive smooth approximations at resolution 2^{j-1} and 2 give the detail signal at resolution 2^j, where j is the resolution

level of the DWT. In other words, after choosing an initial resolution J, any signal $x(s) \in L^2(\mathbb{R})$ can be expressed as [MAL99], [DAU92]:

$$x(s) = \sum_{n\in\mathbb{Z}} c_{J,n}\, \phi_{J,n}(s) + \sum_{j=J}^{\infty} \sum_{n\in\mathbb{Z}} d_{j,n}\, \psi_{j,n}(s), \tag{2.14}$$

where $\phi(s)$ denotes the *scaling function*, $\psi(s)$ denotes the *wavelet function* and the detail or *wavelet coefficients* $\{d_{j,n}\}$ are defined by

$$d_{j,n} = 2^{-j/2} \int_{-\infty}^{+\infty} x(s)\psi_{j,n}(2^{-j}s - n)ds, \tag{2.15}$$

while the approximation or *scaling coefficients* $\{c_{j,n}\}$ are defined by

$$c_{j,n} = 2^{-j/2} \int_{-\infty}^{+\infty} x(s)\phi_{j,n}(2^{-j}s - n)ds. \tag{2.16}$$

Equations (2.15) and (2.16) express that signal $x(s)$ is decomposed in details $\{d_{j,n}\}$ and approximations $\{c_{j,n}\}$ to form a multiresolution analysis of the signal [MAL99].

An attractive property of wavelets is that there exists a recursive relationship between scaling $\{c_{j,n}\}$ and wavelet coeffcients $\{d_{j,n}\}$ at successive levels of resolution [ALA09]. That is, using Equation (2.16) and the dilation equation yields

$$c_{j,n} = \sum_{l\in\mathbb{Z}} g_l\, c_{j-1,\,2n-l} \tag{2.17}$$

$$d_{j,n} = \sum_{l\in\mathbb{Z}} h_l\, c_{j-1,\,2n-l} \tag{2.18}$$

where (g_l) represents the coefficients of a low-pass filter (or scaling filter) and (h_l) denotes the coefficients of a band-pass filter (or wavelet filter). Note that the sequences $\{c_{j,n}\}$ and $\{d_{j,n}\}$ are generated by down-sampling by a factor of two the output of the corresponding filters [ALA09].

Equations (2.17) and (2.18) denote approximation and details coefficients respectively at level of resolution j; these are obtained from approximation coefficients at a finer level of resolution $j-1$. In the opposite direction, approximation coefficients at level of resolution $j-1$ are computed from scaling and wavelet coefficients

at the coarser level of resolution j, according to:

$$c_{j-1,n} = \sum_{l\in\mathbb{Z}} g_l \, c_{j,\,2n-l} + \sum_{l\in\mathbb{Z}} h_l \, d_{j,\,2n-l} \tag{2.19}$$

The sequence $\{c_{j-1,n}\}$ is generated by up-sampling the output of the corresponding filters. This operation is achieved by inserting a zero every two samples. Equations (2.17) and (2.19) can be computed by a pyramid algorithm [MAL99] called Discrete Wavelet Transform (DWT)[ALA09].

The DWT has some limitations. For example, DWT requires the sample size N to be an integer multiple of 2^J. In addition, the number $\{N_j\}$ of scaling and wavelet coefficients at each level of resolution j decreases by a factor of two, due to the decimation process that needs to be applied at the output of the corresponding filters. This can introduce ambiguities in time domain. On the other hand, the number of scaling and wavelet coefficients is the same in all levels of MODWT. The down-sampling process can be avoided by using the Maximal Overlap Discrete Wavelet Transform (MODWT), which is also known as the undecimated discrete wavelet transform [MAL99], [PER00]. The MODWT may be computed for a time series of arbitrary lenght. Note, however, that the MODWT requires $\mathcal{O}(N\,log_2\,N)$ multiplications, whereas the DWT can be computed in $\mathcal{O}(N)$ multiplications. There is, thus, an increase in computational complexity when using the MODWT. However, its computational burden is the same as the widely used fast Fourier transform algorithm and hence quite acceptable [ALA09], [PER00].

2.2.5 The Maximal Overlap Discrete Wavelet Transform (MODWT)

To build the MODWT a rescaling of the defining filters is required to conserve energy, that is, $\tilde{g}_l = g_l/\sqrt{2}$ and $\tilde{h}_l = h_l/\sqrt{2}$, so that $\sum_{l=0}^{L-1} \tilde{g}_l^2 = 1/2$, and therefore the filters are still quadrature mirror filters (QMFs). The wavelet filter must satisfy

the following properties [PER00]:

$$\sum_{l=0}^{L-1} \tilde{h}_l = 0 \tag{2.20}$$

$$\sum_{l=0}^{L-1} \tilde{h}_l^2 = 1/2 \qquad \text{and} \qquad \sum_{l=-\infty}^{\infty} \tilde{h}_l \tilde{h}_{l+2r} = 0, \tag{2.21}$$

for all non zero integers r, where (L) denotes the length of the wavelet filter. The scaling filter $\{\tilde{g}_l\}$ is also required to satisfy Equation (2.21) and $\sum_{l=0}^{L-1} \tilde{g}_l = 1$. Now let $c_{0,n}^{(M)}$ be the time series, the MODWT pyramid algorithm generates the wavelet coefficients $\{d_{j,n}^{(M)}\}$ and the scaling coefficients $\{c_{j,n}^{(M)}\}$ from $\{c_{j-1,n}^{(M)}\}$, where (M) stands for MODWT [PER00]. That is, with non-zero coefficients divided by $\sqrt{2}$, the convolutions can be written as follows [PER00]:

$$\begin{aligned} d_{j,n}^{(M)} &= \sum_{l=0}^{L-1} \tilde{h}_l \, c_{j-1,(n-2^{j-1}l) \bmod N}^{M} \\ c_{j,n}^{(M)} &= \sum_{l=0}^{L-1} \tilde{g}_l \, c_{j-1,(n-2^{j-1}l) \bmod N}^{M} \end{aligned} \tag{2.22}$$

where $n = 0, 1, ..., N-1$ and (N) denotes the length of the time series to be analyzed. Equations (2.22) can also be formulated as circular filter operations of the original time series $\{x_n\}$, using the filters $\{\tilde{h}_{j,l} = h_{j,l}/2^{j/2}\}$ and $\{\tilde{g}_{j,l} = h_{j,l}/2^{j/2}\}$ (see Figure 2.4), namely,

$$\begin{aligned} d_{j,n}^{(M)} &= \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} \, x_{n-l \bmod N} \\ c_{j,n}^{(M)} &= \sum_{l=0}^{L_j-1} \tilde{g}_{j,l} \, x_{n-l \bmod N} \end{aligned} \tag{2.23}$$

As in the DWT, the MODWT wavelet coefficients $\{d_{j,n}^{(M)}\}$ at level of resolution j are associated to the same nominal frequency band given by $|f| \in [1/2^{j+1}, 1/2^j]$. The original signal can be recovered from $d_j^{(M)}$ and $c_j^{(M)}$ using the inverse pyramid algorithm [PER00],

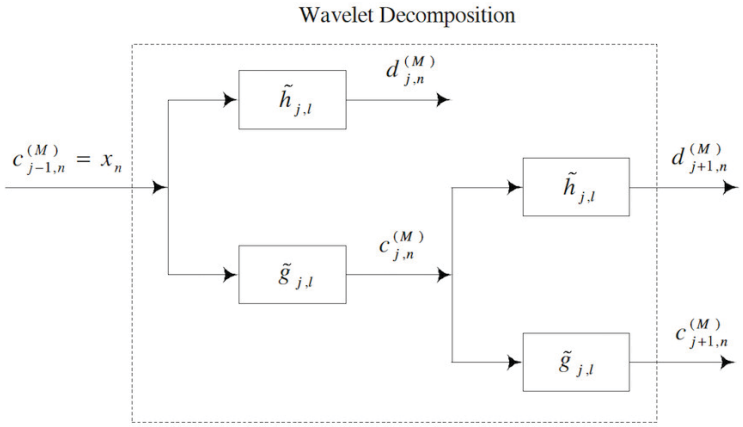

Figure 2.4: Wavelet decomposition of MODWT. The wavelet coefficients $\{d^{(M)}_{j,n}\}$ and scaling coefficients $\{c^{(M)}_{j,n}\}$ are computed by cascading convolutions with filter $\{\tilde{h}_{j,l}, \tilde{g}_{j,l}\}$ [ALA09].

$$c^{(M)}_{j-1,n} = \sum_{l=0}^{L-1} \tilde{h}_l \, d^{(M)}_{j,n+2^{j-1}l \bmod N} + \sum_{l=0}^{L-1} \tilde{g}_l \, c^{(M)}_{j,n+2^{j-1}l \bmod N} \tag{2.24}$$

where $n = 0, 1, ..., N-1$.

2.3 Wavelet Neural Networks

The idea of combining both wavelets and neural networks has resulted in the formulation of wavelet networks, a feed-forward neural network with one hidden layer of nodes, whose basis functions are drawn from a family of orthonormal wavelets. The similarity between the Discrete Inverse Wavelet Transform and a one-hidden-layer Neural Network, universal approximation properties of neural networks, and a rich theoretical basis of wavelet and neural networks have resulted in heightened activity in wavelet networks research and applications [SIT02].

Wavelet neural networks combine some of the useful classification properties of neural networks with the localization and feature extraction properties of wavelets. Wavelet neural networks replace the global sigmoidal activation units of the classical feedforward artificial neural networks (ANNs) with wavelets, while preserving the

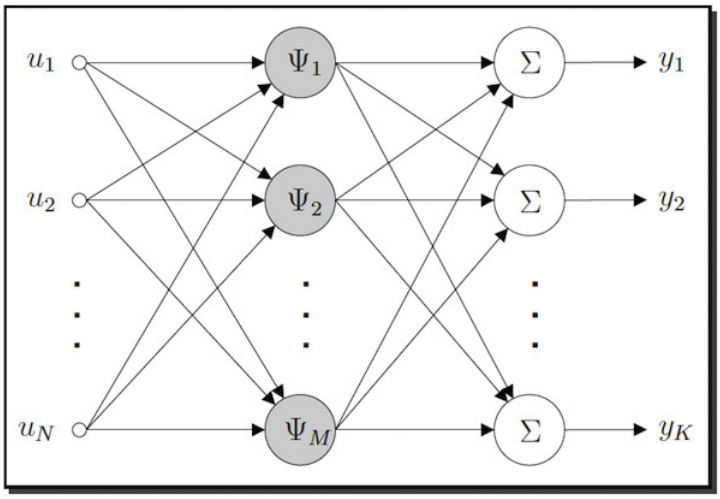

Figure 2.5: Structure of a Wavelet Neural Network [VEI05].

network's universal approximation property. One application of wavelet neural networks is function approximation. Given a series of observed values of a function, a wavelet network can be trained to learn the composition of that function, and hence calculate an expected value for a given input [VEI05], [SIT02].

The structure of a wavelet neural network is very similar to that of a $(1+1/2)$ layer neural network. That is, a Feed-Forward Neural Network, taking one or more inputs, with one hidden layer and whose output layer consists of one or more *linear combiners* or *adders* (see Figure 2.5). The hidden layer consists of neurons, whose activation functions are drawn from a wavelet basis. These wavelet neurons are usually referred to as *wavelons*.

There are two main approaches to creating Wavelet Neural Networks [VEI05].

- In the first the wavelet and the neural network processing are performed separately. The input signal is first decomposed using some wavelet basis by the neurons in the hidden layer. The wavelet coefficients are the output of one or more adders whose input weights are modified in accordance with some learning algorithm.

- The second type combines the two theories. In this case the translation and

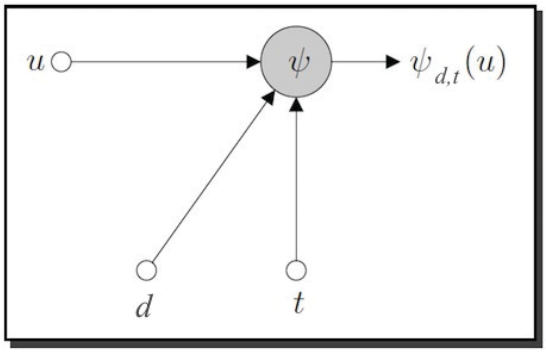

Figure 2.6: A Wavelet Neuron [VEI05].

dilation of the wavelets along with the adder weights are modified in accordance with some learning algorithm.

In general, when the first approach is used, only dyadic dilations and translations of the mother wavelet form the wavelet basis. This type of wavelet neural network is usually referred to as *wavenet*. The second type is referred to as a *wavelet network* [VEI05].

2.3.1 One-Dimensional Wavelet Neural Network.

The simplest form of wavelet neural network is one with a single input and a single output. The hidden layer of neurons consist of wavelons, whose input parameters (possibly fixed) include the wavelet dilation and translation coefficients. These wavelons produce a non-zero output when the input lies within a small area of the input domain. The output of a wavelet neural network is a linear weighted combination of the wavelet activation functions [VEI05].

Figure 2.6 shows the form of a single-input wavelon. The output is defined as:

$$\psi_{d,t}(u) = \left(\frac{u-t}{d}\right) \tag{2.25}$$

where d and t are the dilation and translation parameters respectively.

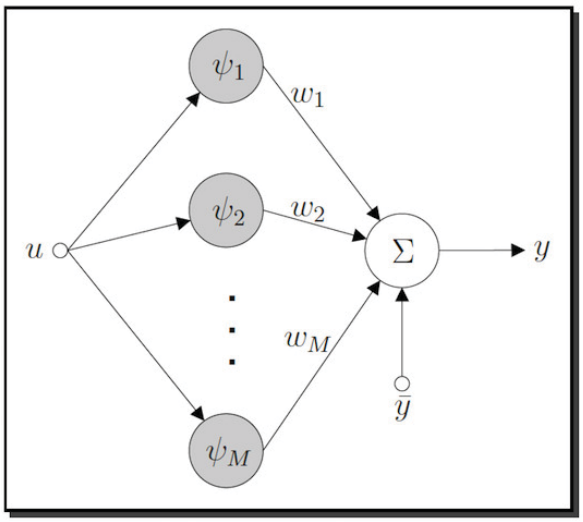

Figure 2.7: A Wavelet Network [VEI05].

Wavelet Network. The architecture of a single input single output wavelet network is shown in Figure 2.7. The hidden layer consists of M wavelons. The output neuron is an *adder*. It outputs a weighted sum of the wavelon outputs [VEI05]. The structure of this Wavelet Network is used in this research.

$$y(u) = \sum_{i=1}^{M} w_i \, \psi_{d_i, t_i}(u) + \bar{y} \tag{2.26}$$

The addition of the $\bar{y}$ value is to deal with functions whose mean is nonzero (since the wavelet function $\psi(u)$ is zero mean). The $\bar{y}$ value is a substitution for the scaling function $\phi(u)$, at the largest scale, from wavelet multiresolution analysis. In a wavelet network all parameters $\bar{y}$, w_i, t_i and d_i are adjustable by some learning procedure [VEI05].

Wavenet. The architecture for a wavenet is the same as for a wavelet network (see Figure 2.7), but the t_i and d_i parameters are fixed at initialisation and not altered by any learning procedure. One of the main motivations for this restriction comes from wavelet analysis. That is, a function $f(\cdot)$ can be approximated to an arbitrary level of detail by selecting a sufficiently large J such that

$$f(u) \approx \sum_n \langle f,\, \phi_{J,n} \rangle \; \phi_{J,n}(u) \tag{2.27}$$

where $\phi_{J,n}(u) = 2^{J/2}\, \phi(2^J\, u - n)$ is a scaling function dilated by 2^J and translated by dyadic intervals 2^{-J}.

The output of a wavenet is therefore

$$y(u) = \sum_{i=1}^{M} w_i\, \phi_{d_i,\, t_i}(u) \tag{2.28}$$

where M is sufficiently large to cover the domain of the function we are analysing. Note that an adjustment of $\bar{y}$ is not needed since the mean value of a scaling function is nonzero.

2.3.2 MULTIDIMENSIONAL WAVELET NEURAL NETWORK.

The input in this type of network is a multidimensional vector and the wavelons consist of multidimensional wavelet activation functions. They will produce a non-zero output when the input vector lies within a small area of the multidimensional input space. The output of the wavelet neural network is one or more linear combinations of these multidimensional wavelets. Figure 2.8 shows the form of a wavelon. The output is defined as

$$\Psi(u_1, ..., u_N) = \prod_{n=1}^{N} \psi_{d_n,\, t_n}(u_n) \tag{2.29}$$

This wavelon is equivalent to a multidimensional wavelet.

The architecture of a multidimensional wavelet neural network was shown in Figure 2.5. The hidden layer consists of M wavelons. The output layer consists of K adders. The output of the network is defined as

$$y_j = \sum_{i=1}^{M} w_{ij}\, \Psi(u_1, ..., u_N) + \bar{y}_j \quad \text{for } j = 1, ..., K \tag{2.30}$$

where the $\bar{y}_j$ allows to deal with functions of nonzero mean.

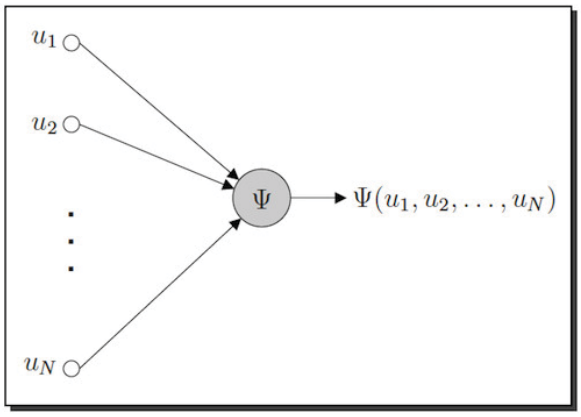

Figure 2.8: A Wavelet Neuron with a Multidimensional Wavelet Activation Function [VEI05].

Therefore the input-output mapping of the network is defined as:

$$\mathbf{y}(\mathbf{u}) = \sum_{i=1}^{M} \mathbf{w}_i \, \mathbf{\Psi}(\mathbf{u}) + \bar{\mathbf{y}} \quad \text{where} \begin{cases} \mathbf{y} = (y_1, ..., y_K) \\ \mathbf{w}_i = (w_{i1}, ..., w_{iK}) \\ \mathbf{u} = (u_1, ..., u_N) \\ \bar{\mathbf{y}} = (\bar{y}_1, ..., \bar{y}_K) \end{cases} \tag{2.31}$$

2.3.3 Learning Algorithm.

As we stated before, one application of wavelet neural networks is function approximation. Zhang and Benveniste [ZHA92] proposed an algorithm for adjusting the network parameters for this application. Learning is performed from a random sample of observed input-output pairs $\{u, f(u) = g(u) + \epsilon\}$ where $g(u)$ is the function to be approximated and ϵ is the measurement noise. Zhang and Benveniste suggested the use of a stochastic gradient type algorithm for the learning [VEI05], [ZHA92]. In this research, we concentrated on the multidimensional wavelet neural network, and we show the learning algorithms for both types of wavelet neural networks described in Section 2.3.1.

Stochastic Gradient Algorithm for the Wavelet Network. The parameters $\bar{y}$, w_i's, t_i's and d_i's should be formed into one vector θ. Now $y_\theta(u)$ refers to the wavelet network, defined by Equation (2.26) (shown below for convenience), with parameter vector θ.

$$y_\theta(u) = \sum_{i=1}^{M} w_i \, \psi\left(\frac{u - t_i}{d_i}\right) + \bar{y}$$

The objective function to be minimized is then

$$C(\theta) = \frac{1}{2} E\{(y_\theta(u) - f(u))^2\} \tag{2.32}$$

The minimization is performed using a *stochastic gradient algorithm*. This recursively modifies θ, after each sample pair $\{u_k,\, f(u_k)\}$, in the opposite direction of the gradient of [VEI05]

$$c(\theta,\, u_k,\, f(u_k)) = \frac{1}{2}(y_\theta(u_k) - f(u_k))^2 \tag{2.33}$$

The gradient for each parameter of θ can be found by calculating the partial derivatives of $c(\theta, u_k, f(u_k))$ as follows:

$$\begin{aligned}
\frac{\partial c}{\partial \bar{y}} &= e_k \\
\frac{\partial c}{\partial w_i} &= e_k \, \psi(z_{ki}) \\
\frac{\partial c}{\partial t_i} &= -e_k w_i \frac{1}{d_i} \psi'(z_{ki}) \\
\frac{\partial c}{\partial d_i} &= -e_k w_i \left(\frac{u_k - t_i}{d_i^2}\right) \psi'(z_{ki})
\end{aligned} \tag{2.34}$$

where $e_k = y_\theta(u_k) - f(u_k)$, $z_{ki} = \frac{u_k - t_i}{d_i}$ and $\psi'(z) = \frac{d\psi(z)}{dz}$.

To implement this algorithm, a *learning rate* value and the number of *learning iterations* need to be chosen. The learning rate $\eta \in (0, 1]$ determines how fast the algorithm attempts to converge. The gradients for each parameter are multiplied by γ before being used to modify that parameter. The learning iterations determine how many times the training data should be fed through the learning process. The

larger this value is, the closer the convergence of the network to the function should be, but the computation time will increase [VEI05].

Stochastic Gradient Algorithm for the Wavenet. As for the wavelet network, we can group the parameters $\bar{y}$, w_i's, t_i's and d_i's together into one vector θ. In the wavenet model, however, the t_i and d_i parameters are fixed at initialization of the network. The function $y_\theta(u)$ now refers to the wavenet defined by Equation (2.28), with parameter vector θ.

$$y_\theta(u) = \sum_{i=1}^{M} w_i \sqrt{d_i}\, \phi(d_i u - t_i) \tag{2.35}$$

The objective function to be minimized is as Equation (2.32), and this is performed using a *stochastic gradient algorithm.* After each $\{u_k, f(u_k)\}$ the w_i's in θ are modified in the opposite direction of the gradient of $c(\theta, u_k, f(u_k))$ (see Equation (2.33)). This gradient is found by calculating the partial derivative

$$\frac{\partial c}{\partial w_i} = e_k \sqrt{d_i}\, \phi(z_{ki}) \tag{2.36}$$

where $e_k = y_\theta(u_k) - f(u_k)$ and $z_{ki} = d_i u_k - t_i$.

As for the wavelet network, a learning rate and the number of learning iterations need to be chosen to implement this algorithm [VEI05].

2.4 Discussion

Epilepsy is a global disease with considerable incidence due to recurrent unprovoked seizures. These seizures can be non invasively diagnosed using electroencephalogram (EEG), a measure of neural electrical activity in brain recorded along scalp. Finding efficient and effective automatic methods for the identification of epileptic seizures is highly desired, due to the relevance of this brain disorder. Among several strategies, the monitoring of the electroencephalographic activity of the brain is considered one

of the most promising options for building epilepsy predictors. Despite the large amount of research going on in identification solutions, still it is required to find confident methods suitable to be used in real applications.

Over the past 10 years, the EEG analysis has been mostly focused on epilepsy seizure detection using different methodologies of computing, such as neural networks, wavelets, Fourier transforms, chaos theory, and others. This research discusses the design of a system to analyze, detect and classify stages of epilepsy using EEG signals, Wavelet Transforms and Wavelet Neural Networks.

In order to keep in mind some basic concepts of issues related to this research, in this chapter we have presented a background information of Epilepsy and Electroencephalogram (EEG) and its characteristics. Later on, an overview of the Wavelet Transforms was presented as well as an introduction about Wavelet Neural Networks.