Neurokiff "Taste-IT"

Case Study Report

Gabriela Hernández Larios and
Maíra Machado Ladeira

EM-DMKM Master Program

June 2015

Acknowledgements

We would like to thank Christophe Thovex and Alex Morel for their support during the whole development of this project.

Contents

List of abbreviations

EEG	Electroencephalogram/Electroencephalography
EMG	Electromyogram/Electromyograph
EOG	Electrooculogram/electrooculographic
IIR	Infinite Impulse Response
FIR	Finite Impulse Response
BSS	Blind signal separation
AR	Autoregressive
WDNN	Weighted Distance Nearest Neighbors
KNN	K-Nearest Neighbors
NN	Nearest Neighbors
LV1	Leave one out
STFT	Short Time Fourier Transform
FT	Fourier Transform
WT	Wavelet Transform
DWT	Discrete Wavelet Transform
SOBI	Second Order Blind Identification

Chapter 1

Introduction

1.1   Motivation

Detection of human emotions is a field that stimulates increasing interest in different scientific domains. The correct detection and classification of human reactions to task stimuli can help develop better consumer products. While different stimuli cause different reactions and different emotions and reactions, the same stimulus in distinct individuals is not necessary the same, but usually correlated.

One of the biggest challenges of emotion detection nowadays is the politeness of a supervised indivual: a person might manipulate its emotions for distinct reasons, such as etiquette or good manners. An automatic process to detect human emotions provoked by different stimuli can help to overcome this challenge. EEG (Electroencephalogram) signals are a record of electrical activity in the brain. These signals are measured with electroencephalograms and, when analyzed, can provide several information about the individualś brain, including emotions felt during the record of the signals. EEG signals are studied for over 100 years and are widely used by medicine for applications like detecting epilepsy and brain death [1]. In previous years, the EEG recording machines were very expensive which made it's use restrict to medical applications. However, during the past years, several low cost portable EEGs, like Muse EEG and Neurosky, have been released, which makes the use of EEG signals for several applications possible, including emotions classification.

When used for classifying emotions, these signals can provide a honest evaluation of an individual's emotions.

1.2   Project Structure

This project aims to use the signals collected by the Muse EEG[1] device to determine if the low cost device is able to correctly detect and classify the emotions provoked by taste stimuli. In Chapter 5 we explain how we collected data with the Muse device, then we clean this data using classical approaches for cleaning EEG signals, as shown in Chapter 6. Then, two different models are generated to classify the collected signals (Chapters 7 and 8) and a discussion is proposed (Chapter 9) Finally, Chapter 10 presents the conclusions of our work.

---

[1]http://www.choosemuse.com/

Chapter 2

Background

In this section, we present the necessary background for the correct understanding of this project.

2.1 EEG Signals

An EEG signal is a measurement of currents that flow during synaptic excitation of the dendrites of many pyramidal neurons in the cerebral cortex. When brain cells are activated, the synaptic currents are produced by the neurons and transported by the dendrites. These currents generate a magnetic field measurable by electromyogram (EMG) machines and a secondary electrical field over the scalp measured by EEG systems [2].

Electroencephalography (EEG) is the recording of electrical activities along the scalp using a pair or multiple electrodes that detect signals correlated to neural activity [3].

These electrical activities are rich in information on mental activity and emotional states [4] and several research projects are conducted in order to classify emotions from the EEG signals. For example, the work of [5] the EEG signals are used to classify positive and negative emotions from stimuli provoked by selected movies that are displayed to the subjects. The work has an accuracy result of around 85%. The work of [6] aims to classify the subjects emotions for stimuli provoked by the subject's' imagination (the subject imagines a situation that would provoke a certain emotion and this process is recorded by and EEG equipment) and by images and musics that are supposed to trigger a certain emotion. The study obtains results with accuracy of around 90-95% and also analyses the best electrodes to measure each emotion. The work of [7] also uses stimuli provoked by videos in order to detect subjects emotions in a unsupervised model using fuzzy-k-means and fuzzy-c-means clusters. Several further examples can be found in the literature.

2.1.1 Type of waves

The visual inspection of EEG signals allows diagnosing many brain disorders. They are represented by waves which, when analyzed by experts in the field, can reveal important information about one's brain. There are five major brain waves distinguished by their different frequency ranges: alpha, theta, beta, delta and gamma [2].

The **alpha** waves appear in the posterior half of the head and have a frequency range of $8\text{-}13Hz$. These waves have been thought to indicate a relaxed awareness without attention or concentration. Most subjects produce some alpha waves with their eyes closed and the waves are reduced or eliminated by opening the eyes, hearing an unfamiliar sound, anxiety or mental concentration or attention. Alpha

waves have higher amplitudes, but usually less than $50\mu V$.

Theta waves have a frequency within the range of 4-8Hz and have been associated to unconscious material, creative inspiration and deep meditation. The theta waves are frequent during the infancy and childhood and in adults are usually associated to pathological problems.

Beta waves have a frequency range of 14-26Hz and are usually associated to the activities of thinking, active attention, focus or solving concrete problems. Beta waves are found in normal adults and the amplitude of these waves are normally less than $30\mu V$.

Delta waves are in the frequency range between 0.5-4Hz and are primarily associated with deep sleep. It is very easy to confuse artifact signals caused by the large muscles of the neck and the jaw with the genuine delta response but signal analysis methods help to determine when the response is caused by artifacts and when it is an actual delta wave signal.

The frequencies above $30Hz$ correspond to the **gamma** waves. These waves have an amplitude very low and their occurrence is rare, usually confirming a certain type of brain diseases.

Figure 2.1 shows a representation of the four dominant brain normal rhythms.
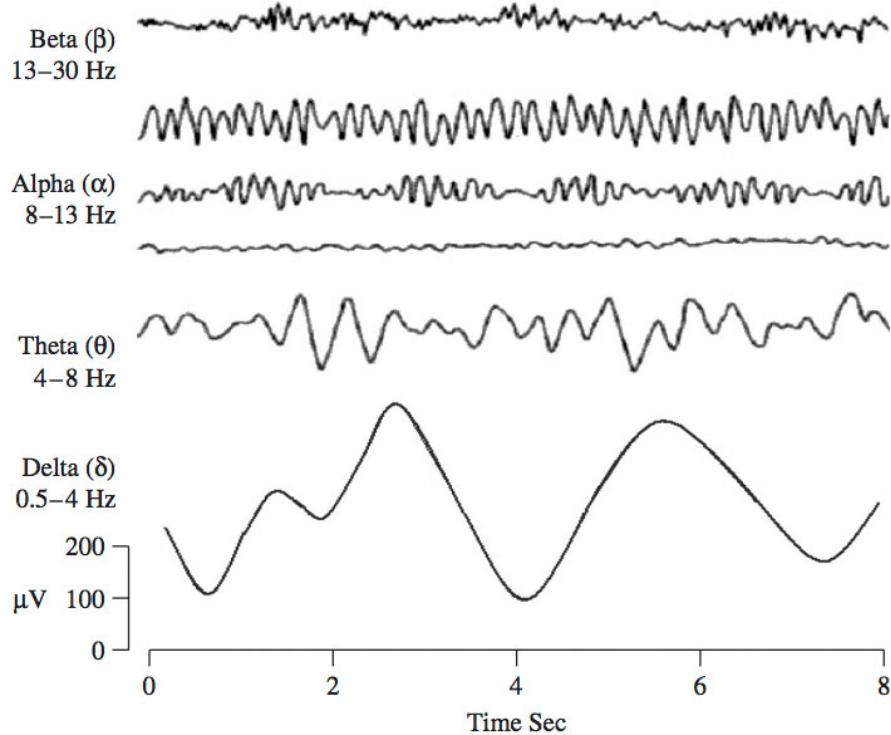
Figure 2.1: Representation of the 4 typically dominant types of brain waves from [2], for high and low frequencies.

The power of a signal is defined as the amount of energy consumed per unit of time. The power of a signal is given by the sum of the absolute squares of its time-domain samples divided by the signal length. The band power feature reflects the power at each electrode position [8]. In this project, the band power is used to measure the power of each EEG wave (alpha, beta, delta, theta and gamma) for each stimulus, allowing us to use the different EEG waves in a time-domain approach.

2.2 Noise Treatment

One of the biggest challenges to the analysis and interpretation of EEG signals is that the recording is highly susceptible to several sources and forms of noise.

The noise present on the data can be generated in many different ways. The blinking of an eye, the environmental sounds or even the electrical current are already enough to disturb a good source of data during the record of EEG waves.

Some techniques of noise handling have been developed and will be described below, according to the work of [9]:

- Elimination of noise source: The best way of dealing with noise problems on EEG signals is having no noise in the first place. Some external source of noise problems can be eliminated with a simple charging of the portable EEG device or by asking the subjects to not move during the performance of the experiments. Factors like the environmental lights and sounds should also have their influences minimized during the data collection.

- Signal Averaging:The key for this technique is the assumption that the noise in the signal is random and, for this reason, only random and symmetric noise can be eliminated with this technique.

- Rejection of noise data: when the noise is easily recognizable and not frequent, the best way of dealing with it is with the elimination of the noise data.

- Removal of noise: There are several ways to remove the noise of EEG data. Filtering, linear regression, data decomposition (such as ICA and FFT) and Blind signal separation are some of the options available.

2.2.1 Filtering

A filter is essentially a system or network that selectively changes the wave shape, amplitude-frequency and/or phase-frequency characteristics of a signal in a desired manner [10]. Digital filters can be divided in two categories:

- Infinite Impulse Response (IIR): are filters which have an impulse response that does not become exactly zero in a finite amount of time. IIR filters are represented by:

$$y(n) = \sum_{k=0}^{\infty} h(k)x(n-k) \tag{2.1}$$

- Finite Impulse Response (FIR): are the filters which have and impulse response with finite duration, settling to zero in a finite time. FIR filters are represented by:

$$y(n) = \sum_{k=0}^{N-1} h(k)x(n-k) \tag{2.2}$$

In the analysis of EEG signals, filters are usually used to remove some frequencies and keep others, in order to suppress interfering signals and reduce background noise. In this project, FIR filters are used to remove frequencies from the EEG signals that are either too low or too high in frequency.

2.2.2 Blind signal separation - BSS

Eye- and muscle movements produce electrical potentials that propagate over the scalp, creating significant electrooculographic (EOG) and electromyograph (EMG) artifacts in the recorded EEG [11] and [12]. These artifacts complicate the interpretation of the EEG and are traditionally corrected using regression-based methods that assume the recording of noisy signals as reference for the cleaning of the desired ones. However, there is a growing interest in separating a mixed signal without the aid of noisy

information. Blind Signal Separation (BSS) is a method to separate mixed signals that allow the removal of noise from recorded signals without the regression from a previously known noisy signal. BSS identify noisy components using their predominant characteristics, such as few predominant low-frequency components in EOG signals [11].

This project uses BSS techniques for the removal of EOG and EMG components from the EEG signals.

2.3 Fourier Transform (FT)

Fourier shows that any periodic function can be represented as infinite sum of sines and cosines (periodic complex exponential functions), as shown in figure 2.2. Roughly speaking, the Fourier analysis or harmonic analysis of a time series is a decomposition of the series into a sum of sinusoidal components (the coefficients of which are the discrete Fourier transform of the series). Fourier analysis is often used to describe any data analysis procedure that describes or measures fluctuation in a time series by comparing them with sinusoids.

The Fourier transform decomposes a function of time, like a signal, into the frequencies that compose it. The Fourier transform is the frequency domain representation of the original signal in time domain. Therefore, the term Fourier transform refers to both the frequency domain representation and the mathematical operation that associates the frequency domain representation to a function of time.
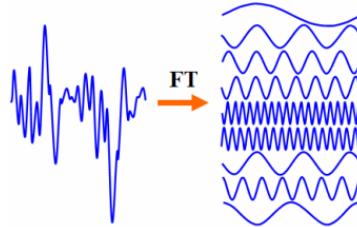

Figure 2.2: Fourier transform definition

The Fourier transform is defined as:

$$F(\omega) = 1/2\pi \int_{-\inf}^{inf} f(t)e^{-j\omega t}dt \tag{2.3}$$

where j represents the imaginary part, i.e. $j = \sqrt{-1}$ and ω the angular frequency in rad/s, which is equivalent to $2\pi f$.

In the above equation, we can see that the Fourier transform takes a function of time $f(t)$ and transforms it to a function which lives in the frequency domain $f(\omega)$, and changes the basis of the function to consines and sines.

Another way to write down this formula is by using sines and cosines, which can be done by substituting $e^{j\theta}$ for the Euler's formula 2.4.

Euler's formula says that:

$$e^{j\theta} = cos(\theta) + jsin(\theta) \tag{2.4}$$

In this study, we first attempt to analyze the EEG signals by using this analysis, however since it is crucial for our goal to know at what instance certain frequencies were risen, this analysis could not be longer used. Figure 2.3, shows the FT applied to the red wine experiment of one of the participants.
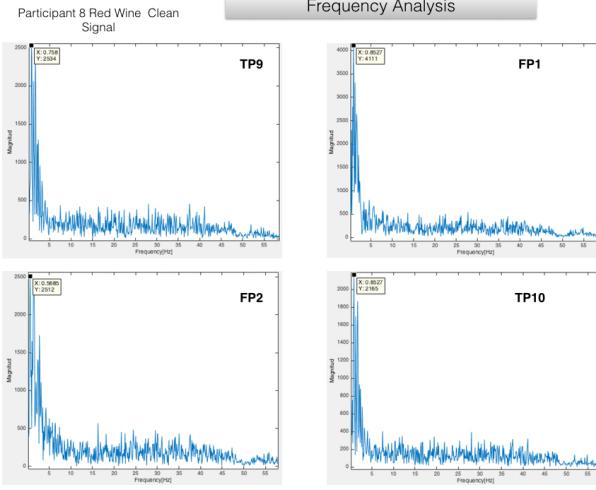
Figure 2.3: Fourier Transform applied to the signals of the Red Wine Experiment

Many works are using Fourier analysis for classifying EEG signals, mostly to identify neurological diseased or tumors in the brain, such as in [13], [14], [15]. However, in our case we are interested on the frequencies risen in the instant when the stimuli is given to the participants, therefore this method cannot give us all the information needed.

2.4   Short Time Fourier Transform (STFT)

In STFT, the signal is divided into small portions that can be assumed to be stationary, to accomplish this purpose, a window function is necessary. The width of the window function must be equal to the segment where its stationarity holds. The STFT is defined as:

$$STFT(l, w) = \int_t [f(t)w(t - l)]e^{-j\omega t}dt \qquad (2.5)$$

where w is a window function.

In each window the Fourier transform is performed to obtain the spectral component.

There is a trade-off between the width of the window and the resolution of the time and frequency. The narrower we make the window, the better we get the time resolution but a worse frequency resolution, and vice-versa; making the window wider results in a better frequency resolution but a worse time resolution, as shown in figure 2.4. This problem stems from the *Heisenberg's Uncertainty Principle* which states that it is not possible to know the exact time-frequency representation of a signal, i.e. we cannot know the spectral component existing at an instant. What we can know is time-intervals in which certain bands of frequencies are, raising a resolution problem.

In STFT the problem is the width of the window, therefore the best width is application dependent. In some studies this method has giving good results such as those reported on [14], [13].
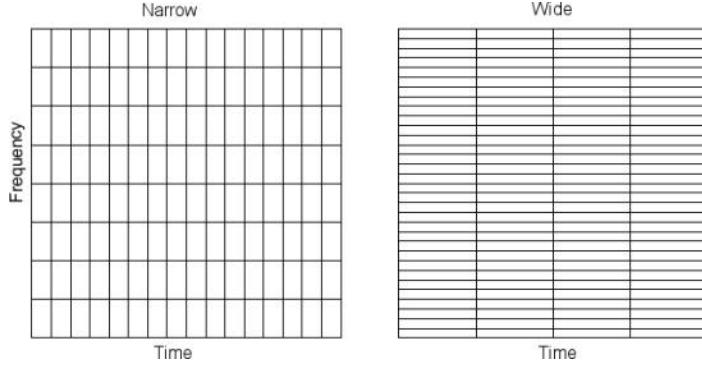
Figure 2.4: Comparison between a Narrow window and a Wide window

We used the STFT to validate with the video recorded during the experiments, whether the stimulus given at the instant t, rose some frequencies and the interval of seconds these risen frequencies have lasted. This was useful for filtering the signals, focusing on the time where the stimuli were given for each participant.

The figures 2.5 and 2.6 show the signal in time-domain and its STFT respectively. For this participant, we have validated, with the help of the video recorded, that the stimulus was giving in the second 5, so from the second 5 to the second 8 we can observe some changes in the band of frequencies.
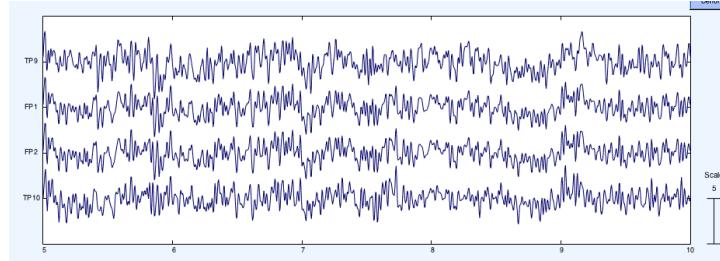
Figure 2.5: EEG signals of participant 10 during the chocolate experiment
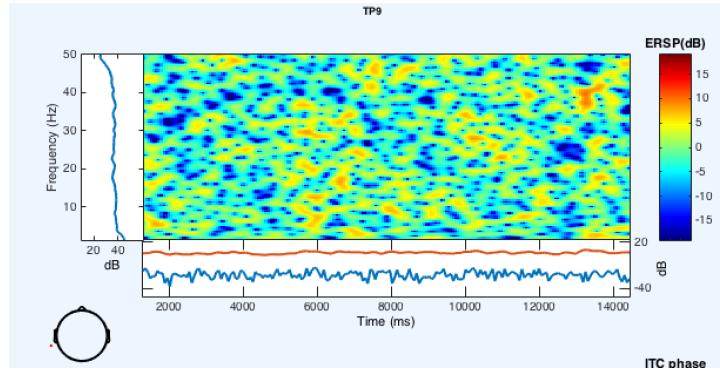
Figure 2.6: STFT applied to the EEG signal (channel TP9) of figure 2.5

2.5 Wavelet Transform

2.5.1 Wavelet Theory

In this work we are dealing with two main domains for analyzing in the signals: frequency and time.

While the Fourier transform is a very useful tool to analyze the frequency of the components in a signal, it cannot give information about at what instance a particular frequency rises. In this study, it is

crucial to know this information, in order to discover if the stimulus giving at time t, has rose a particular frequency in the participant. Short-time Fourier Transform (STFT), gives information of frequency and time by using sliding window to find a spectrogram. Nevertheless the length of the window tends to limit the resolution of the frequency, the narrower we make the window, we get better time resolution but poorer frequency resolution and vice versa.

Wavelet transform works on a multi-scale basis,giving information of time and frequency without limiting the resolution of the frequency, for achieving this, it analyzes the signal at different frequencies with different resolutions. Therefore, the multi-scale feature of the Wavelet transform permits the decomposition of a signal into two versions: The translated version wavelets allow us to locate where we are interesting on, whereas the scaled-version wavelets allow us to analyze the signal in different scales.

In this study, we have used the family of Daubechies wavelets. Analogous to the sines and cosines used in Fourier Analysis, wavelets are used as basis functions to represent or decompose the signal in other functions. Once the "mother wavelet" has been fixed, it is translated, obtaining translated wavelets, and scaled, obtaining the scaled wavelets. The term translation is related to the location of the window, as the window is shifted through the signal, thus corresponds to time information whereas scaling either dilates or compress a signal, in WT smaller scales correspond to compressed signals while large scales to dilated signals. The family of Daubechies wavelets can be observed on figure 2.7
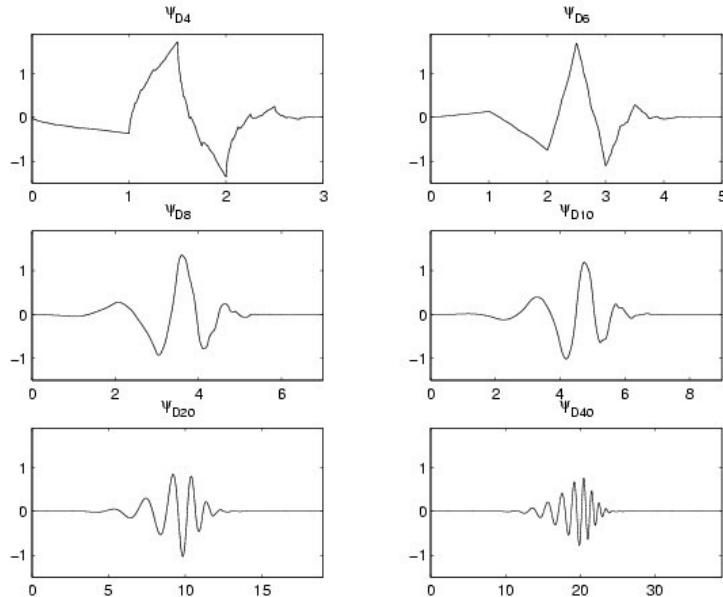
Figure 2.7: Daubechies Family of wavelets

In WT, we talk about a time-scale plane, figure 2.8, where the interest is the relation between scale and frequency, low scales correspond to high frequencies whereas high scales to low frequencies. The WT gives good time resolution and poor frequency resolution at high frequencies and good frequency resolution and poor time resolution at low frequencies. It means that this approach works well, when the signal has high frequency components for short durations and low frequency components for long durations which is the case in most biological signals, mainly EEG, EMG and ECG. For this study, we chose the daubechies wavelet of order 4 (db4) as a mother wavelet, this decision was based on several experiments performed by testing the HAAR wavelet (db2), db4 and db6, having the best results with db4. It is worth mentioning that in recent studies adaptive wavelets for the signals have been developed such as the Wavelet Kernel Learning algorithm proposed in [16], however during among this work, we only used daubechies wavelet 4 to extract features of the EEG signals.
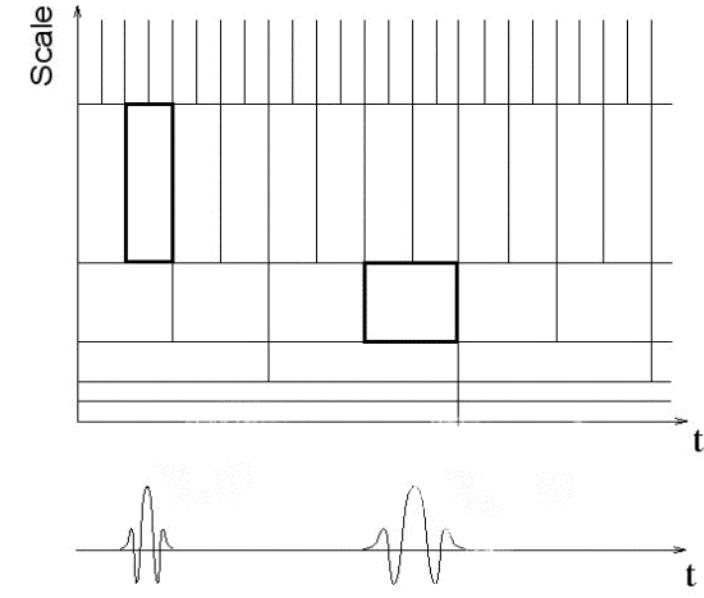
Figure 2.8: Time-scale plane

Conversely to STFT, in WT the Fourier transforms of the windowed signals are not taken and the width of the window is changed due to the fact that the transform is computed for every single spectral component.

For this work, we used Discrete Wavelet Transform (DWT), defined as follows:

$$C(a,b) = C(j,k) = \sum_{neZ} f(n)\psi_{j,k}(n) \tag{2.6}$$

where $\psi_{j,k}$ is a discrete wavelet (called mother wavelet) and it is defined as:

$$\psi_{j,l}(n) = 2^{-j/2}\psi(2^{-j}n - k) \tag{2.7}$$

The parameters a,b represent the scale and the translation respectively, where $a = 2^i$ and $b2^j k$

The coefficient C(a,b) represents the correlation of the wavelet with that portion of the signal, the higher C, the higher similarity. Figure 2.9 gives an intuition of this concept.
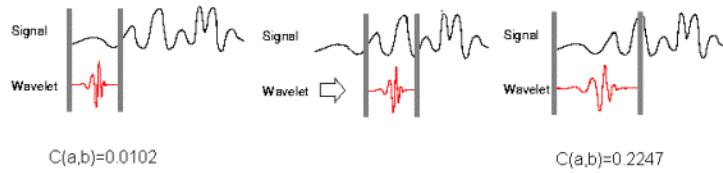
Figure 2.9: Interpretation of the Wavelet Transform coefficients

2.5.2 Wavelet Transform Computation

In order to be useful and easy to apply, wavelet transform uses fast algorithms such as algorithms based on multiresolution analysis, for finding the coefficients C(j,k) and for reconstructing the function they represent.

By using multiresolution analysis, the time-scale representation of a digital signal is obtained using digital filtering techniques.

The procedure of the multiresolution decomposition of a signal $x[n]$ is shown by the schema in figure 2.10. The sequence $h[n]$ is the low-pass filter and $g[n]$ is the high-pass filter. Filters belong to the family finite impulse response (FIR) filters. In other words, the first filter $g[n]$ corresponds to the discrete mother wavelet, high-pass in nature, whereas $h[n]$ is its mirror version which is low-pass in nature. The down-sampled output of the high-pass filter provides the detail D_n whilst the output of the low-pass filter gives the approximation A_n, the approximation is further decomposed at each level step, as we can observe in figure 2.10 where the first approximation A_1 is further decomposed and the process is continued.
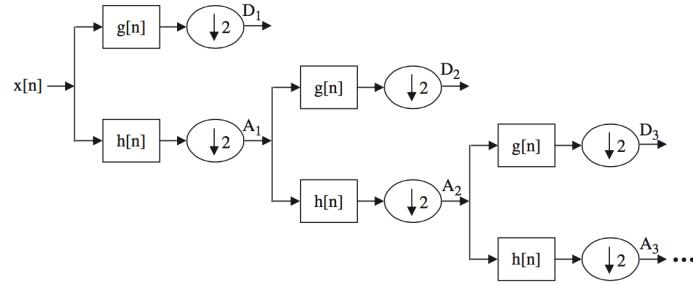
Figure 2.10: Subband decomposition of the discrete wavelet transform, where g[n] is the high-pass filter and h[n] represents the low-pass filter. [17]

The normalized wavelet 2.8 and scale basis 2.9 functions can be defined as:

$$\varphi_{i,l}(k) = 2^{i/2} h_i(k - 2^i l) \tag{2.8}$$

$$\phi_{i,l}(k) = 2^{i/2} g_i(k - 2^i l) \tag{2.9}$$

where k is the equally sampled discrete time, $2^{i/2}$ is an inner product normalization, and i and l are the scale parameter and translation parameter, respectively (please notice that before, in subsection 2.5.1 these parameters were defined as a and b respectively).

Finally, the discrete wavelet transform (DWT) can be described as:

$$a_{(i)}(l) = x(k) * \varphi_{i,l}(k) \tag{2.10}$$

$$d_{(i)}(l) = x(k) * \phi_{i,l}(k) \tag{2.11}$$

where $a_{(i)}(l)$ and $d_{(i)}(l)$ are the approximation coefficients and the detail coefficients respectively.

Thus, the discrete wavelet transformation can be summarized as a single line for the signal x[n]:

$X[n] -> (d^1, d^2, ..., d^k, a^k)$

where k represents the level or layer.

This can be easily seen in the figure 2.11, where there are three levels of decomposition and the leaves represent the coefficients of the Wavelet transform, i.e., D1, D2, D3, A3.
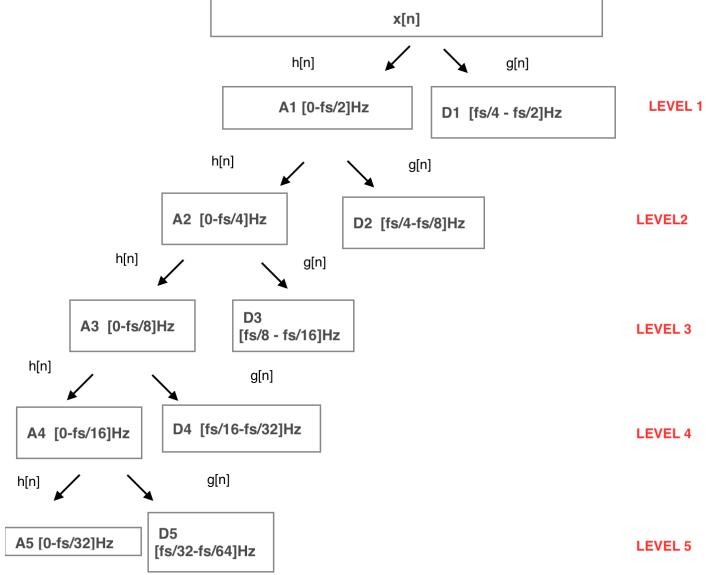
Figure 2.11: Wavelet transform decomposition in levels

2.6 Autoregressive coefficients (AR coefficients)

An autoregressive model (AR) is the a representation of a type of random process useful for representing time-varying processes, like EEGs. In a AR model each sample of data can be predicted from the previous weighted samples, where the number of coefficients denotes the model order [8]. The AR model is represented by:

$$x(t) = -\sum_{i=1}^{p} \hat{a}_i x(t-i) \tag{2.12}$$

where $\hat{a}_i$ denotes the AR model coefficients and p is the model order.

The Burg algorithm (well defined and explained in [18]) estimates the AR parameters determining reflection coefficients k that minimize the sum of forward and backward residuals[19].
The burg algorithm is computationally simple and guaranteed to be stable. The AR coefficients found with the Burg algorithm in general represent the signals in a more detailed way than mean and variance.

2.7 Higuchi fractal dimension - HFD

A fractal dimension can be interpreted simply as the degree of irregularity in a signal [8]. The Higuchi fractal dimension is a technique for calculating the fractal dimension D, of a time series [20].

In order to obtain the fractal dimension D, Higuchi [] considered a finite set of observations taken as regular interval:

$$X(1), X(2), X(3), ..., X(N). \tag{2.13}$$

From this one, a new one must be constructed:

$$X_k^m; X(m), X(m+k), X(m+2k), ..., X(m + [\frac{N-m}{k}]) \tag{2.14}$$

with $(m = 1, 2, ..., k)$;, both k and m integers, m and k indicate the initial time and the interval time,

respectively.

Higuchi [], defines the length of the curve associated to each time series, X_k^m as:

$$L_m(k) = \frac{1}{k}\left(\sum_{i=1}^{[\frac{N-m}{k}]}(X(m+ik)-X(m+(i-1)k))\right)\left(\frac{N-1}{[\frac{N-m}{k}k]}\right) \qquad (2.15)$$

where the term

$$\frac{N-1}{[\frac{N-m}{k}k]} \qquad (2.16)$$

represents a normalization factor. Higuchi takes the average value $< L(k) >$ of the lengths associated to the time series given by the equation 2.15. If the average value follows a power law:

$$< L(k) > \propto k^{-D} \qquad (2.17)$$

then the curve is fractal with dimension D.

2.8 Probabilistic Neural Network

Basically, a probabilistic neural network (PNN) is a feedforward neural network based on the Bayesian Networks and the statistical algorithm Kernel Fisher discriminant analysis. This is organized in a multilayered feedforward network with four layers: Input layer, Pattern layer, Summation layer and Output layer, as shown in figure 2.12.
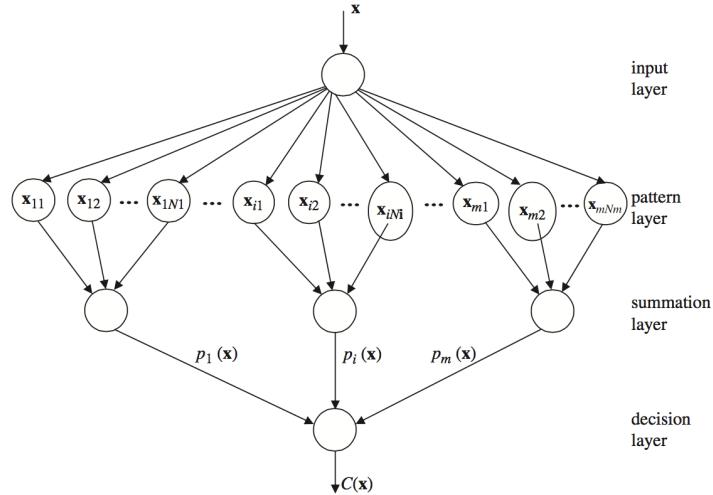
Figure 2.12: Basic architecture of the PNN [17]

In the input layer no computation is performed, this is only for distributing the input to the neurons in the pattern layer. When the pattern layer receives a vector $\mathbf{x}$ from the input layer, the output of each neuron x_{ij} of the pattern layer is defined as:

$$\phi_{ij}(x) = \frac{1}{(2\pi)^{d/2}\sigma^d}exp[-\frac{(\mathbf{x}-\mathbf{x}_{i,j})^T(\mathbf{x}-\mathbf{x}_{i,j})}{2\sigma^2}] \qquad (2.18)$$

where d corresponds to the dimension of the pattern vector $\mathbf{c}$, σ is the smoothing parameter and $\mathbf{x}_{ij}$

is the neuron vector.

This output is passed to the summation layer,where the neurons compute the maximum likelihood (ML) of pattern $\mathbf{x}$ of being classified into the class C_i. This is achieving by summing and averaging the output of all neurons that belong to the same class, and it is defined as:

$$p_i(x) = \frac{1}{(2\pi)^{d/2}\sigma^d}\frac{1}{N_i}\sum_{j=1}^{N_i}exp[-\frac{(\mathbf{x}-\mathbf{x}_{i,j})^T(\mathbf{x}-\mathbf{x}_{i,j})}{2\sigma^2}] \qquad (2.19)$$

where N_i denotes the total number of samples in class C_i. Please note, that in the equations 2.19 and 2.18 the T denotes the transpose of the term $\mathbf{x}-\mathbf{x}_{i,j}$

Finally, the decision layer, classifies the pattern $\mathbf{x}$ with the Bayes's decision rule based on the output coming from all the summation layer neurons, as stated in the following equation:

$$\hat{C}(\mathbf{x}) = argmax p_i(x) \qquad (2.20)$$

for $i = 1, 2, ..., m$.

$\hat{C}$ denotes the estimated class of the pattern $\mathbf{x}$ and m is the total number of classes in the training samples [17].

2.9   Weighted Distance Nearest Neighbors

K-nearest Neighbor algorithm (KNN) is one of the most well-know supervised learning algorithm in pattern classification, introduced in [21]. It retains the entire training set during the learning process and returns as classification a class represented by the majority label of its k-nearest neighbors in the training set. The Nearest Neighbor algorithm (NN) is a simplification of the KNN when $k = 1$ [22]. Weighted Distance Nearest Neighbors is a improvement to NN where the leave one out (LV1) classification error is minimized by the assignment of weights values to the training set [8].
By the assignment of small weights to low quality training samples and big weights to high quality ones, the influence of each sample in the feature space can be controlled and the quality of the model can be improved. This way, the WDNN rejects the assumption of the NN algorithm that all the samples have the same quality, which is specially true for the record or the EEG samples.

The model represented in 8 uses a combination of AR coefficients with band power measures of the alpha, beta, delta, theta and gamma waves and the HFD of the EEG signal to feed a WDNN classification method and return the best classification for each EEG recording.

Chapter 3

Overview

The first step of our work was to define the project structure and create a project overview. As our work involves a complete cycle of data mining for the emotion classification of EEG signals, our first task was get familiar with the Muse device and define a protocol of data collection in order to, with the help of Muse EEG low cost device, collect EEG signals generated with taste stimulus. With the collected signals, the next step was to determine the best way for pre-processing the signals in order to minimize artifacts influences and guarantee a clean version of our data. With the data cleaned, we decided to create two different models to classify our data, allowing a future comparison of them.

Probabilistic Neural Network and Weighted Distance Nearest Neighbors (WDNN) were selected as the algorithms to be implemented in the model. Each one of them was feed with different properties of our data. For the Probabilistic Neural Network, we calculated the wavelet coefficients and used their statistic values: media, standard derivation, minimum, maximum for feeding the neural network in a work similar to what was done in [17]. For the WDNN, we calculated the AR coefficients of the data, the band power of alpha, beta, delta, theta and gamma waves and the HFD of the EEG signals to feed the classifier in a work similar to [8]. The model of the WDNN algorithm was modified in order to first try to classify each signal and check the obtained result. When the obtained result is not the expected one, the signal is added to the training sample. With this approach we expect to improve the performance of the model over the time.

The final phase of our project was to perform the interpretation of the results and the comparison of the models. Figure 3.1 shows a graphical representation of the overview of our project, represent each phase and total flow of the work. Each phase of the project will be detailed in the following chapters.
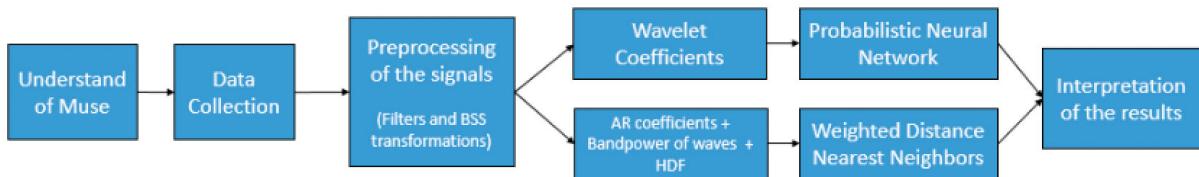
Figure 3.1: Overview of the project

Chapter 4

The Muse device

The Muse[1] headset is a portable EEG for the collection of brain waves data, Figure 4.1. Originally, Muse is developed together with an application that aims to help the user to control his/her brain for meditation purposes. However, Muse also has a development SDK program[2] that allows the collection of data and facilitates the research applicability of the device.

Figure 4.1: The Muse headset

4.1 Data Files

Data can be saved in CSV, MATLAB or native MUSE format. It provides several paths with different type of data recorded as: EEG raw data without any processing, pre-processed data by Muse SKD providing alpha, beta, delta, gamma and theta waves in their absolute values and also in relative values, weighting the waves with the values of the other wave types [3].MUSE encodes the raw EEG data as a value between 0.0 and 1682.815 micro-volts, and the EEG signals are oversampled and then downsampled to yield a selectable output sampling rate from 220Hz to 500Hz, for our experiments we used a sampling rate of 220 Hz [23].

Regardless the path of the data, in all the files, the header specifies the order in which the data from the channels have been recorded: TP9, FP1, FP2, TP10. The first entry line in each file is the time stamp indicating the year, month, day, hour, minute, second, millisecond and microsecond at which the data was received. The next values represent the data values at each of the electrode locations. Due to the already pre-processed alpha, beta, delta, gamma and theta data were not giving us good results and their frequency bands were not the ones found in the literature, we decided to use the raw data during this project with the goal of detecting emotions.

---

[1] http://www.choosemuse.com/
[2] http://www.choosemuse.com/developer-kit/
[3] https://sites.google.com/a/interaxon.ca/muse-developer-site/museio/osc-paths/osc-paths—v3-6-0

4.2   Sensor placement

In 1958, International Federation in Electroencephalography and Clinical Neuropsychophysiology adopted a standardization for electrode placement called 10-20. The 1020 system is an internationally recognized method to describe and apply the location of scalp electrodes in the context of an EEG test or experiment [24]. Electrode placements are labeled according to the brain areas as: F(frontal), C(Central), T(Temporal), P(Posterior) and O(Occipital). MUSE data is collected from the electrodes FP1 and FP2 (Prefrontal or forehead) and TP9 and TP10 (Temporal-Posterior). In the 10-20 electrode placement system shown in figure 4.2 the red colored electrodes highlight the ones used by MUSE. All the data collected with the headset is described according to these 4 electrodes.
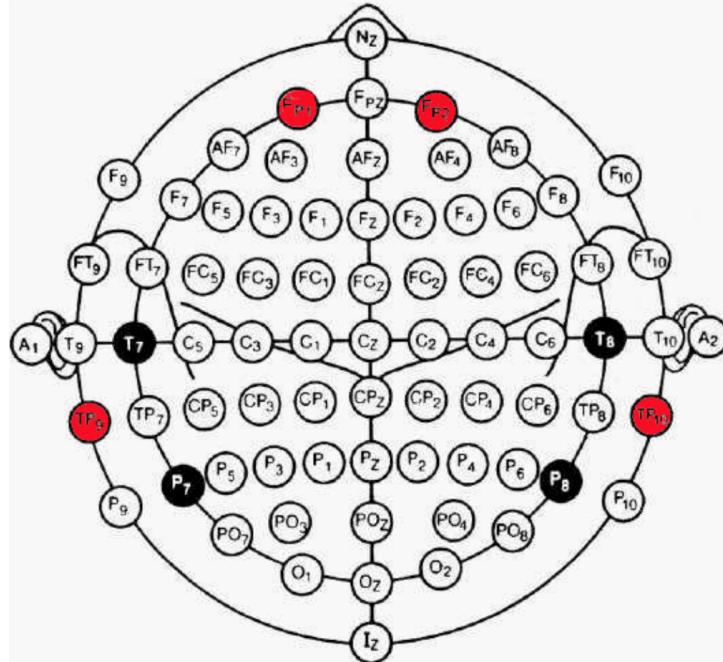
Figure 4.2: 10-20 Electrode Placement System

Besides these electrodes, MUSE has three additional reference sensors: a center sensor which is the system ground and the two outside sensors as references for computing the voltage differences, figure 4.3 shows the sensor locations in the MUSE device.
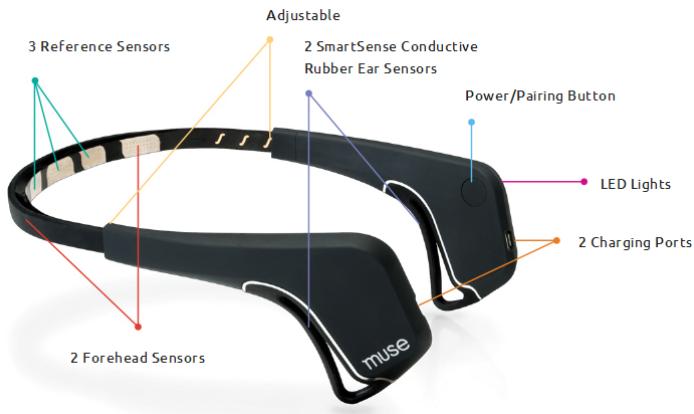
Figure 4.3: Muse Sensor Locations

Chapter 5

Data Collection

5.1 Experiment Protocol

For the data collection process, an experiment protocol was defined, in order to avoid noise and bias in the raw signals and to make the pre-processing and cleaning step easier. We based our protocol in previous studies with regard to food experiments and the emotions risen, as those stated by Park et al. in [25] and Rousmans et al. in [26].

5.1.1 Participants selection

The first step of the data collection protocol was to define the participants to be analyzed during the experiment. For this phase, adults without any mental diseases were selected. The target was 12 people between 23 and 59 years old.

5.1.2 Inclusion/Exclusion criteria

Inclusion criteria:

- All study participants must agreed in fill the survey questionnaire.

- All study participants must speak english

Exclusion criteria:

- Study participants must not take medications like anti-epileptic, sedatives, tranquilizers, barbiturates and antidepressants.

- Study participants must not eat and drink anything (except water) within 2 hours of the experiment appointment.

- Study participants must not have taste disorders.

- study participants must not have eating disorders.

- Study participants must not be allergic to the food they will taste.

- Study participants must not have mental disorders nor hyperactivity

5.1.3 Participants preparation for the experiment

EEG signals can be influenced by the consume of certain types of aliments in a period close to the data collection. In order to minimize the effects caused by it and this way have a better result on the experiments, the participants are required to not eat or drink anything, except water, for the 2 hours preceding the experiment as previously mentioned. The participants are also required to not consume alcohol and/or drugs during the day of the experiment and on the day before it. Finally, all the participants are queried to clean the parts of the head where the electrodes will be positioned.

5.1.4 Taste Stimulus

For the taste stimulus of the experiment, 4 different aliments are used to provoke different reactions on the subjects. The selection of the food according to the reactions stimulated on the brain was based on the experiments performed by Rousmans et al. in [25] and Park et al. in[26]. Rousmans proposed the solution of 0.3M sucrose to rise pleasant stimulus, and 0.15M NaCl for unpleasant stimulus, whereas Park et al. proposed milk chocolate and mustard for respectively pleasant and unpleasant stimulus.

For this study we have chosen 0.15M of NaCl as an unpleasant stimuli and milk chocolate as a pleasant stimuli. One of the target applications for this project is evaluate wines by measuring the brain waves of the participants to check whether the wine was liked or not. Having said that, we added red wine (Merlot - Pays D'Oc - La Réserve de Valentin) and White wine (Saumur- Appellation Saumur Contrôlée - Grande Réserve) as taste stimulus to check the emotions in the brain.

Questionnaires applied to all the subjects should define a grade between 1-5 for each stimulus.

5.1.5 Experiment description

Healthy subjects were asked to taste food/liquids and provide emotional feedback for each taste.First, the subjects were verbally informed about the procedure and the experiment, without revealing the tastes (for avoiding bias). Before the experiment started, the subjects were asked to clean their head where the sensors of the headset would be placed, to sit in a comfortable position and to avoid to move or blink. For each participant the EEG device was calibrated until it did not present noise anymore, this could be achieved with the muse-io and muse-lab modules which provides information of the noise in each channel. The experiments consisted on four trials for each participant regarding the different tastes. Each participant began the experiment by drinking water, followed by the milk chocolate solution, NaCl solution, red wine and ending with the white wine. After each different taste stimuli, the subject drank water and rested for 30 seconds before starting a new trial. During the trials, subjects were asked to not swallow the liquid, keeping it in their mouths for 4 seconds. The aliments were conduced to their mouths with the help of a spoon to minimize the influences of the movements on the recording of the EEG signals

When all the trials were performed, the participant was asked to fill a questionnaire describing his/her emotions in each trial, between a scale from 1 to 5, being 1 a very unpleasant emotion and 5 a very pleasant emotion.

The data is recorded for a period of approximately of 10s with MuseLabs and later converted to Matlab format with MusePlayer.

Each experiment was video-recorded with the agreement of the participants.

5.2   Experiment Results

The experiments were performed in a group of 12 participants that agreed with the consume of alcohol except for 2 that didn't agree. The questionnaire field by the participants can be visualized on the Table 5.1

Participant	Age	Salt Stimulus	Chocolate Stimulus	Red Wine	White Wine
1	24 years	1	4		
2	24 years	2	4		
3	26 years	1	3	5	4
4	30 years	2	4	4	5
5	23 years	2	4	2	3
6	27 years	2	5	4	4
7	25 years	1	3	3	4
8	49 years	1	4	5	4
9	52 years	1	4	5	4
10	26 years	1	3	5	4
11	26 years	1	3	5	4
12	31 years	1	4	5	5

Table 5.1: Participants questionnaire results

Unfortunately, when the data was processed and analyzed we found out that the data of some participants were still containing a lot of noise, mostly for face movements, that we could not avoid and we had to discard them. Leaving only the data of participants 10, 11 and 12 to work in the development of the model.

Chapter 6

Cleaning the data

This chapter describes the process of pre-processing of the EEG signals in order to remove the artifacts of the data and get a clean signal.

6.1 Filtering and Removal of Artifacts

After deciding to work with the raw signal, our first task was to filter the data remove the artifacts from it. In this phase of the work, we filtered the signal in a range of $1\text{-}50Hz$ and tried to remove artifacts like eye blink (EOG), eye movement (EOG), line noise and muscle activity (EMG) from our raw signal in order to obtain a clean EEG. In Fig 6.1 it is possible to visualize the expected forms of raw and artifacts signals. 6.1(a) shows what to expect from a clean EEG signal, 6.1(b) represents the noise generated by an eye blink, 6.1(c) shows the noise generated by eye's movement, 6.1(d) represents the line noise at 60Hz and 6.1(e) shows the noise generated by muscular activity. The muse EEG device already has a

(a) Clean EEG

(b) Eye Blink

(c) Eye Movement

(d) 60 Hz Line Noise

(e) Muscle Activity
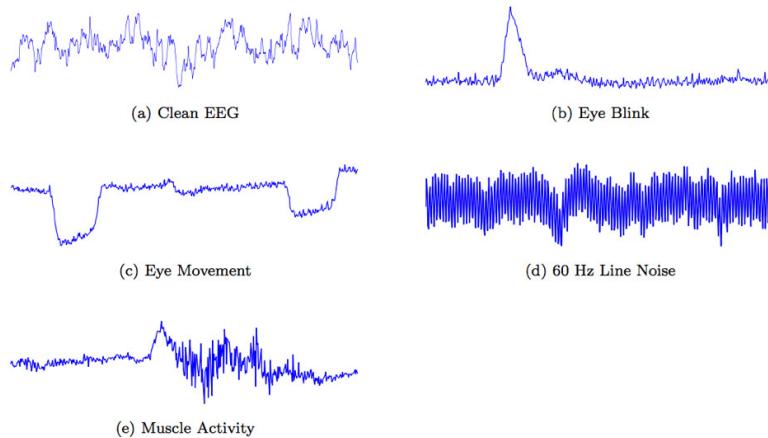
Figure 6.1: Raw signal before the artifacts removal (top) and after artifact removal (bottom) by [?]

line noise filter build in on it. For that reason, the clean of the line noise artifacts did not represent an important part of our work. The filtering and removal of artifacts activities of this project were both performed with EEGLab [1]. EEGLab is a toolbox for matlab for processing continuous EEG signals and includes features for visualize, analyze and transform the signals.

---

[1] http://sccn.ucsd.edu/eeglab/

With the collected data transformed for the format accepted by EEGLab, our first step was to apply two FIR filter of lower and higher frequency respectively. The filters are a default feature of EEGLab and they are applied separately, according to EEGLab's tutorials[2]. Figures 6.2 and 6.3 show the EEGLab windows for the FIR filters configuration and Figure 6.4 shows the resultant signal, after the application of the filters.
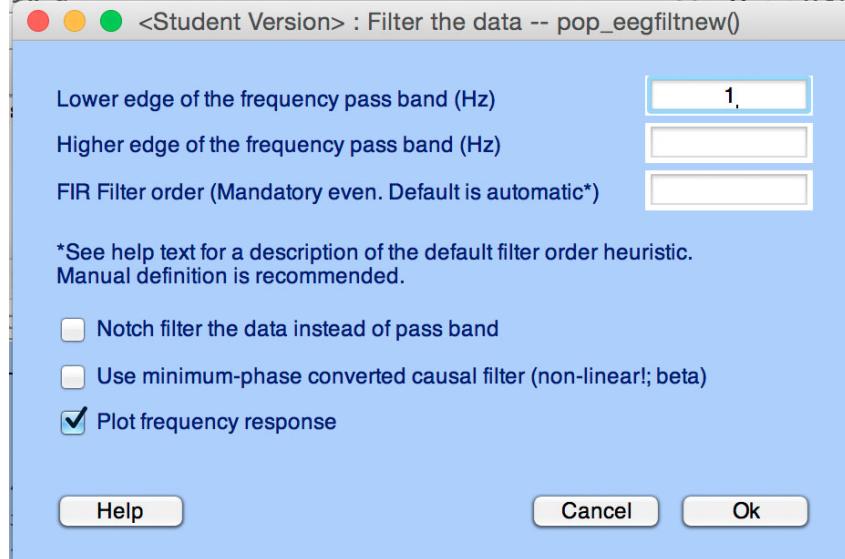
Figure 6.2: Filter of the Lower edge of the frequency pass band
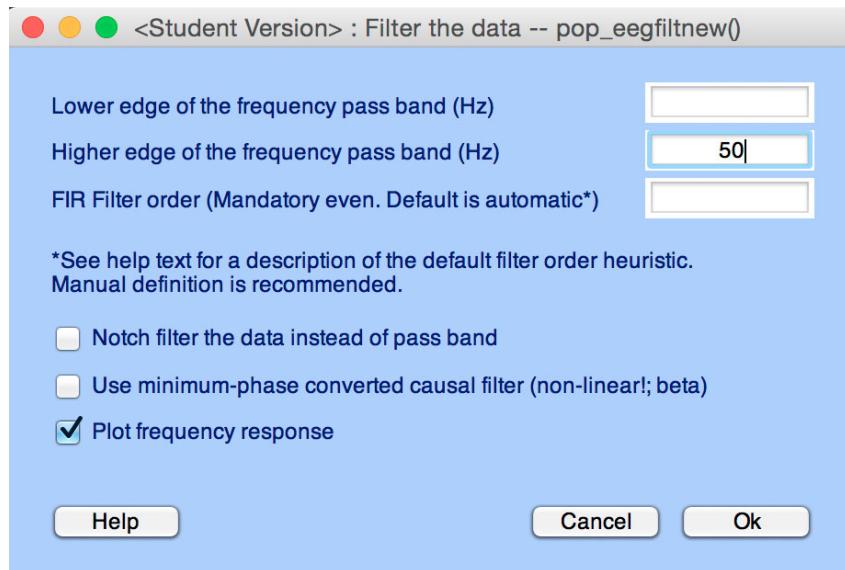
Figure 6.3: Filter of the Higher edge of the frequency pass band

With the Filtered data, we performed a visual rejection of the data, eliminating the most predominant noisy data. Figure 6.5 shows the process of recognize and reject noisy data from EEG signals. After this, an automatic removal of artifacts was performed with the AAR plugin[3] of EEGLab. The AAR plugin uses BSS techniques with Second Order Blind Identification (SOBI) for remove EOG and EMG artifacts from the EEG signals, according to the work of [11]. Figure 6.6 shows the resultant EEG signal after

---

[2]http://sccn.ucsd.edu/wiki/Chapter_04:_Preprocessing_Tools
[3]http://www.germangh.com/eeglab_plugin_aar/
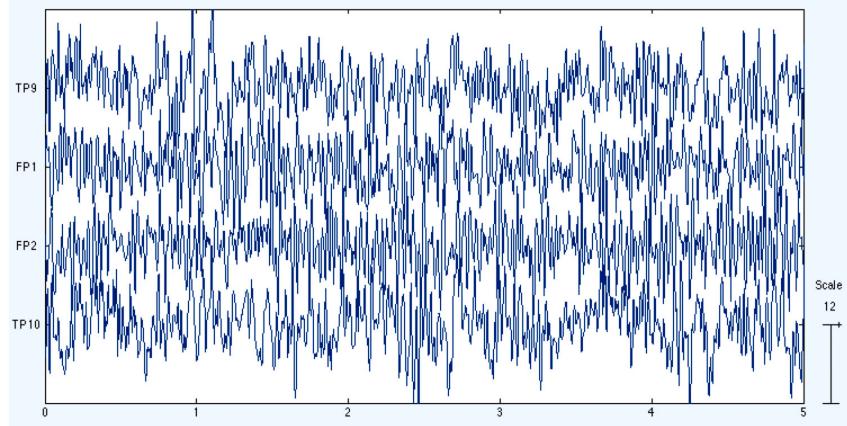
Figure 6.4: Filtered data resultant from the lower and higher edge filters

the removal of EOG and EMG artifacts. It is possible to check in the Figure the points where the visual analysis removed the most obvious noisy (marked with a red line).
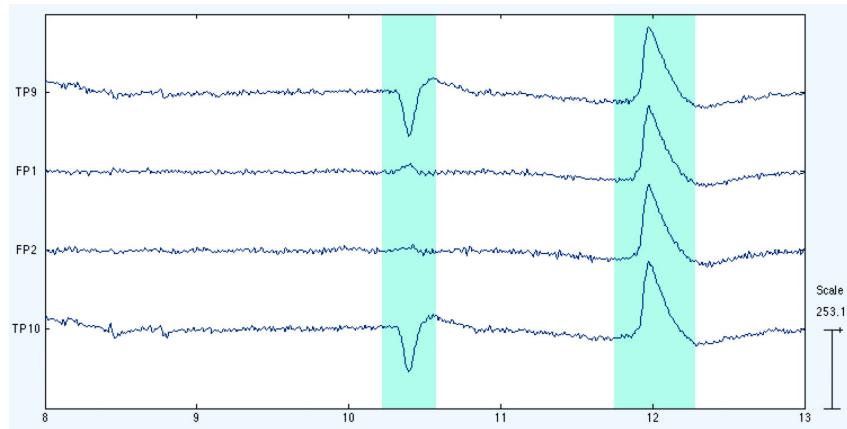
Figure 6.5: Visual rejection of noisy data with EEGLab.
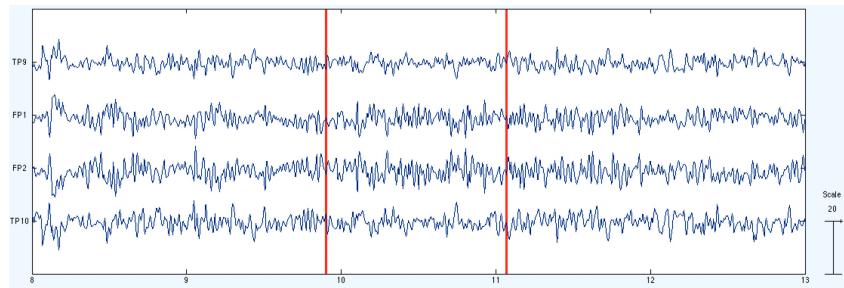
Figure 6.6: EEG signal after Filtering and removal of EOG and EMG artifacts.

Comparing Figures 6.4 and 6.6 it is easy to see that the result signal is closer to a clean EEG signal than the original after the filters. It is also possible to notice that the automatic techniques used on the process of signal cleaning is not removing all the noise from the raw data. This fact is acknowledged by this project but won't be treated in a first moment. The signal resultant from this phase of the project was then used in both models, developed in 7 and 8

Chapter 7

Model 1 - Weighted Distance Nearest Neighbors

This chapter describes the activities performed in the cleaned EEG data in order to perform a WDNN classification. The idea for this model is to classify each new EEG datum by useing the current model in order to obtain a predicted class. The predicted class is then compared with the expected one and when the results are different the model is updated, including the new data signal in the training sample. With this approach we expect to obtain a model that improves over time. The next sections will describe the activities performed both in the signal and in the model, in order to obtain the final model.

7.1   Data preparation

One big challenge of time-series data like EEG is the phase shift problem. Similar waves can have completely different results for starting in different time periods. Several approaches have been developed dealing with this problem representing different aspects of the data that can be interesting for different applications of the data.
In this first developed method, we deal with this problem by representing the EEG data with 3 different approaches:

- Autoregressive Coefficients

- Band power of alpha, beta, delta, theta and gamma waves

- Higuchi fractal dimension of the EEG signal

This approaches, selected according to the work of [8], were chosen because they investigate different aspects of the data. They are related to the power spectrum, frequency domain and complexity or irregularity of the EEG signals, respectively.
In this project, all the previous approaches were calculated with the help of Matlab functions and the process for obtain each of these properties will be described below:

The **AR coefficients** are calculated with the Burg method using the *arburg* function of the Signal Processing toolbox of Matlab. This project uses the order 4 to calculate the AR coefficients in order to have a smaller dimension of the data than the one described in [8].
The **Band power** of each wave type (alpha, beta, delta, theta and gamma) were calculated using the *bandpower* function, also from the Signal Processing toolbox of Matlab.

The **Higuchi fractal dimension** was calculated using the function *Higuchi_FD* from the complete higuchi fractal dimension algorithm[1]. In this project, the HFD is calculated with a max delay of 50 time series.
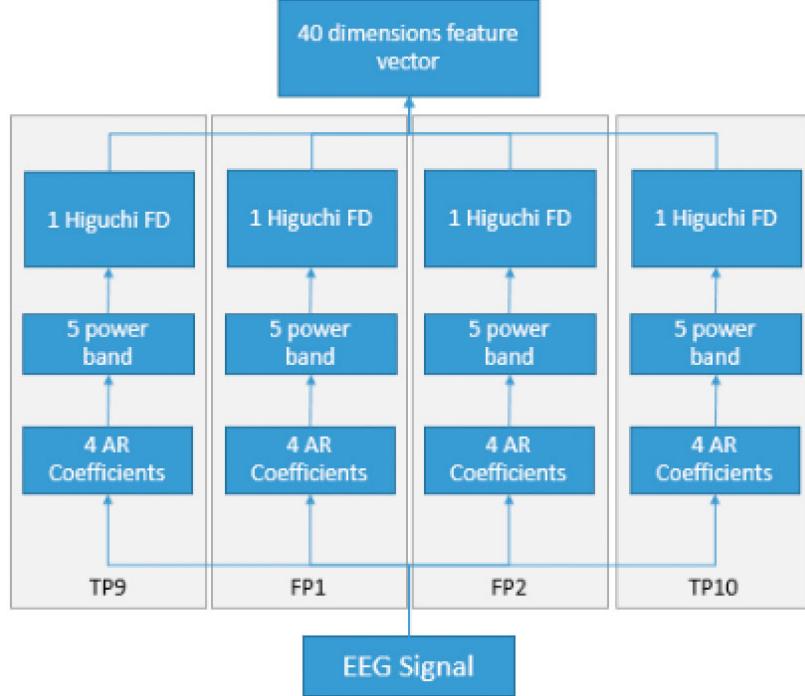
Figure 7.1: Composition of the data used to feed the WDNN algorithm. Each channel has its properties calculated and the final vector is the combination of all the channels.

Each of these properties is calculated for each channel of an EEG signal and the results of the channels are combined in order to generate the data point for being used in the WDNN algorithm. Figure 7.1 shows the composition of the each data point to be used by the WDNN algorithm.

In order to better estimate the classes of each signal, and for removing the effects of the data recorded without the presence of a taste stimulus, we remove the first and the last $2s$ of each EEG signal record and we divided each signal in short windows of $1s$ with $0.5s$ of overlap of each window. With this technique we guarantee that the features from one short window to another won't change a lot (50% overlapping) and that the differences in the signal are represented by the model. The predicted classification chooses the most common classification for each short window of an EEG signal. Figure 7.2 shows the representation of the short windows division of the signal.
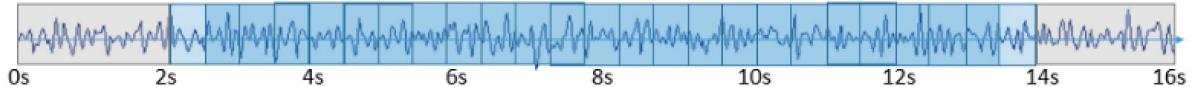
Figure 7.2: Processing of the EEG signal in short windows with overlapping and elimination of first and last $2s$.

---

[1]http://www.mathworks.com/matlabcentral/fileexchange/30119-complete-higuchi-fractal-dimension-algorithm

7.2   Model Setup

The setup of the model was done in Python, version 3, using the *sklearn.neighbors.KNeighborsClassifier* class[2]. The classifier allows the implementation of KNN and WDKNN algorithms. For the implementation of the WDNN, the k-neighbors parameter of the *KNeighborsClassifier* class is set to 1. The Euclidean distance is used to calculate the distance between two points as:

$$dist(P_i, P_j) = \sqrt{\sum_{k=1}^{f}(P_j k - P_i k)^2} \tag{7.1}$$

And the class of closest point of the training set is used as the classified class.

For each point, the model determine the predicted class and compares with the expected one. When the expected class is different from the predicted one, the point is added to the classifier as training set.

7.3   Results

The WDNN model was tested using the EEG signals collected for 3 different subjects with 4 different taste stimulus (chocolate, salt, red wine and white wine) in two different ways:

1. The data of one subject is used to predict the stimulus of a different subject

2. A percentage of the data from each subject is used as training and the rest of the data as test

   Each test was performed 10 times and considering the 3 different situations:

- the class result is equal to the subject's personal evaluation of the stimulus (1 to 5)

- the class result is equal to good or bad, using the personal evaluation of the stimulus but considering 1, 2 bad and 3, 4, 5 good. In this situation, the good emotion is represented by the number 1 and the bad by the number 2

- the class result is equal to *chocolate, salt, red_wine or white_wine* where the classes are represented by 1, 2, 3, 4 respectively.

Table 7.1 shows the results of situation 1 with the classification respecting the subject's questionnaires. Table 7.2 shows the results of situation 1 with the classification respecting the subject's questionnaires but being simplified for good and bad signals.
Table 7.3 shows the results of situation 1 with the classification according to the taste stimulus (chocolate, salt, red wine or white wine).
Table 7.4 shows the result of situation 1 when we try to classify the experiments per subject (subject 1, subject 2 or subject 3).

As can be seen in the tables, the model does not correctly classify the data when we try to determine the emotions retracted in the experiments of the type of stimulus. How ever, the classification when we try to determine which subject has each experiment is better (getting to 7/11 correct results), how ever still not acceptable. This indicates the model is not a good approach for classifying the emotions, using different subjects for test. More than that, it also indicates that the classification of the data is personal related. In order to confirm if the data is strongly personal related and if the model is able to correctly classify the emotions using a sample of the data from each experiment for each subject, the experiments were repeated, without the incremental learning and using a sample of the data for training and a sample of the data for testing.

---

[2]http://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html

Classification of Emotion (1-5)																			
Exp. 1		Exp. 2		Exp. 3		Exp. 4		Exp. 5		Exp. 6		Exp. 7		Exp. 8		Exp. 9		Exp. 10	
P	E	P	E	P	E	P	E	P	E	P	E	P	E	P	E	P	E	P	E
1	4	5	3	1	3	4	3	1	3	5	3	5	1	5	4	4	1	4	3
4	4	3	4	1	4	4	1	3	4	5	4	1	4	5	4	1	3	3	4
4	5	4	5	4	5	1	5	1	5	3	4	1	3	4	5	3	5	3	1
1	4	3	3	3	5	4	5	3	5	4	5	1	5	5	5	3	4	4	5
4	5	4	5	5	4	1	5	1	4	5	1	3	1	5	1	1	3	4	4
1	1	3	5	5	4	3	5	4	1	4	5	4	1	4	1	4	5	4	5
1	3	5	4	4	1	5	4	5	4	4	5	3	5	1	4	3	1	4	1
1	3	4	1	4	1	3	1	1	5	1	3	1	4	5	3	5	1	1	5
3	5	3	1	5	5	4	3	1	3	3	1	5	4	4	5	5	5	1	5
3	5	5	4	1	3	3	1	3	5	4	1	1	3	1	3	3	4	1	3
5	1	4	1	1	5	5	4	3	1	5	5	1	5	3	1	3	5	3	1

Table 7.1: Classification of emotions from (1-5) using EEG signals and the subject's questionnaires

Classification of Good/Bad Emotion (1-2)																			
Exp. 1		Exp. 2		Exp. 3		Exp. 4		Exp. 5		Exp. 6		Exp. 7		Exp. 8		Exp. 9		Exp. 10	
P	E	P	E	P	E	P	E	P	E	P	E	P	E	P	E	P	E	P	E
1	2	1	2	1	2	1	1	1	2	1	1	1	1	2	2	1	2	1	1
2	1	2	2	1	2	1	1	2	1	1	1	1	1	2	1	2	1	1	1
1	2	2	1	2	1	1	1	1	1	1	1	1	2	2	1	1	1	1	2
2	1	2	1	2	1	1	2	2	1	1	2	1	2	1	1	1	2	1	2
2	1	2	1	1	2	1	1	1	1	1	1	1	2	2	1	2	1	1	1
1	1	1	2	1	1	1	1	1	1	1	2	2	1	1	1	1	2	1	1
2	1	2	1	2	1	1	1	1	1	1	1	2	1	1	2	1	1	1	1
1	1	1	1	2	1	1	1	2	1	1	1	2	1	1	1	2	1	1	1
2	1	2	1	2	1	1	2	1	2	1	2	1	1	2	1	2	1	1	2
1	2	1	1	1	1	1	2	1	2	1	1	1	1	2	1	1	1	1	1
1	1	2	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1

Table 7.2: Classification of good and bad emotions (1 or 2) using EEG signals and the subject's questionnaires

The results of table 7.5 that the use of data samples increases the accuracy of the classifier. The result is even better when we try to classify the subject of each experiment. These results makes the theory that the EEG signals, recorded by the Muse device, present a very personal related data when used to detect emotions.

Classification of Food (1-4)																			
Exp. 1		Exp. 2		Exp. 3		Exp. 4		Exp. 5		Exp. 6		Exp. 7		Exp. 8		Exp. 9		Exp. 10	
P	E	P	E	P	E	P	E	P	E	P	E	P	E	P	E	P	E	P	E
4	4	1	2	4	2	3	3	1	3	1	3	1	4	2	1	1	2	3	1
4	3	1	1	4	2	3	4	3	4	3	2	1	2	2	4	1	4	3	4
3	1	1	3	4	1	3	2	3	1	3	4	1	2	1	4	4	4	1	4
3	2	3	4	2	1	3	4	1	2	3	1	1	3	4	4	4	2	1	3
1	2	3	1	2	1	3	1	1	2	3	1	2	4	4	2	4	1	3	2
2	4	1	4	1	3	1	1	2	4	1	2	2	2	2	2	2	1	4	4
4	3	4	2	1	3	1	3	1	4	1	4	2	4	2	1	1	2	1	3
3	1	2	3	1	3	3	1	1	2	1	3	2	1	1	3	2	4	3	1
1	2	1	2	1	4	1	4	2	3	1	3	2	3	2	3	1	3	1	2
3	1	2	4	2	4	3	2	1	1	1	4	3	1	1	1	1	3	3	1
1	3	1	3	1	2	3	2	1	3	1	2	1	3	1	3	1	3	3	2

Table 7.3: Classification of Food stimulus (chocolate, red wine, salt or white wine) as indexes 1-4

Classification of Subject (1-3)																			
Exp. 1		Exp. 2		Exp. 3		Exp. 4		Exp. 5		Exp. 6		Exp. 7		Exp. 8		Exp. 9		Exp. 10	
P	E	P	E	P	E	P	E	P	E	P	E	P	E	P	E	P	E	P	E
1	3	2	2	1	3	2	1	3	2	2	3	1	3	3	1	3	1	2	2
1	1	2	1	3	3	1	1	2	1	2	3	1	1	3	3	1	1	2	2
3	3	1	2	3	3	1	2	3	3	3	3	1	1	1	2	1	2	2	2
1	2	1	1	1	2	2	2	3	3	2	1	1	3	1	2	2	2	2	3
1	3	2	3	1	1	2	2	1	2	1	1	1	2	1	1	1	1	2	3
1	1	3	1	2	2	1	1	1	1	1	1	1	1	2	2	2	2	2	1
1	2	1	1	2	2	1	3	2	2	2	2	1	2	3	3	1	1	3	3
3	3	3	3	1	2	1	3	2	2	1	2	2	3	1	1	3	3	2	1
2	2	2	2	1	3	3	3	1	1	1	1	2	2	1	1	1	2	2	1
1	1	3	3	1	1	1	1	3	1	2	2	2	2	3	3	3	3	1	1
3	2	3	3	1	1	1	3	3	3	2	3	3	3	3	2	2	3	3	3

Table 7.4: Classification of subject for each experiment (subjects 1-3)

Classification of Emotion using sample of data				
	Emotions 1-5	Emotions 1-2	Experiment type from 1-4	Subjects
Avg Accuracy	0,824	0,874	0,727	0,94

Table 7.5: Classification of the data using 30% of each data entrance as training set.

Chapter 8

Model 2: Relative Wavelet Energy and Probabilistic Neural Network as an analysis of EEG signals

In this chapter, the EEG signals are analyzed by using Wavelet transform, and then, a PNN was implemented to classify the EEG signals in two classes: pleasant emotion and unpleasant emotion.

The present model consists on three main modules: feature extractor to extract a vector representing the EEG signals, feature selection for selecting the most important features to represent the EEG signals and feature classifier which gives the classes based on the extracted features.

8.1   Data preparation

The main objective of this work was to analyze the frequencies generated in the brain by certain food stimulus. For this purpose, we first analyze the video recorded altogether with the STFT spectrum, for each experiment, for each participant, as explained in section 2.5. In such a way that we could detect the approximated instant when the stimuli was given and the subsequent 3 seconds. We extracted only these signals for all the experiments. For instance, for the chocolate experiment of the participant 10, we analyzed the video recorded and we find out that the stimuli was given in the second 5.3, afterwards we analyzed the STFT of this signal and we discovered that some frequencies were altered when the stimuli was given, so we select the data from second 5.3 to second 8.3 to analyze its wavelet transform.
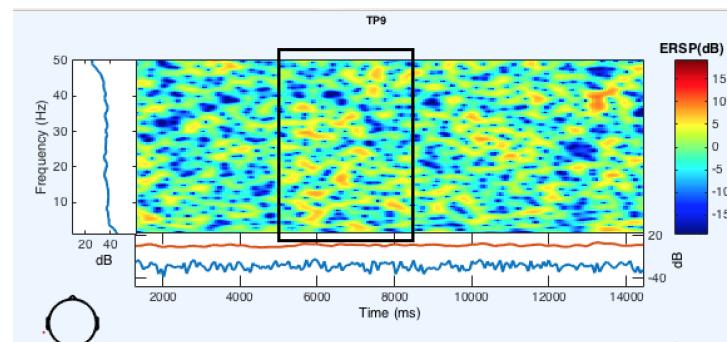

Figure 8.1: STFT applied to the EEG signal (TP9 channel) of participant 10 for the chocolate experiment.

8.2   Feature extraction and feature selection

8.2.1   Analysis of EEG signals with discrete wavelet transform

The main purpose of the wavelet analysis is to decompose the signals into different frequency bands depending on the number of decomposition levels, as previously explained in section 2.5.

The selection of the appropriate wavelet as well as the number of decomposition levels is crucial for analyzing signal using DWT. The number of decomposition levels is chosen in accordance with the dominant frequency of the components. Due to the EEG signals rarely have frequencies above 30Hz, we have decided to chose 5 decomposition levels. Also, with 5 decomposition levels the different brain waves can be detected in the levels.

All the signals were sampled at 220Hz by MUSE, and we aimed to obtain their DWT coefficients. Since the signals are sampled at 220Hz, the highest frequency component that exists in these signals is 110Hz by following the Nyquist theorem.

Supposing we have a signal with 880 samples at 220Hz. At first level, the signal is passed through the low-pass filter $h[n]$, and the high-pass filter $g[n]$, and then the outputs are subsampled by two, as explained in section 2.5. The high-pass filter output is the first level of DWT coefficients. At this level we have 440 samples that represents the frequency band at the [55 - 100]Hz range. The low-pass filter output, which also has 440 samples, represents the frequency band of [0- 55]Hz, that will be further decomposed by passing the signal through the same h[n] and g[n]. The figure 8.2. explains this process, while the table 8.1 summarizes the frequency bands obtained at each level of decomposition with a sampling frequency of 220Hz and a Wavelet db4.
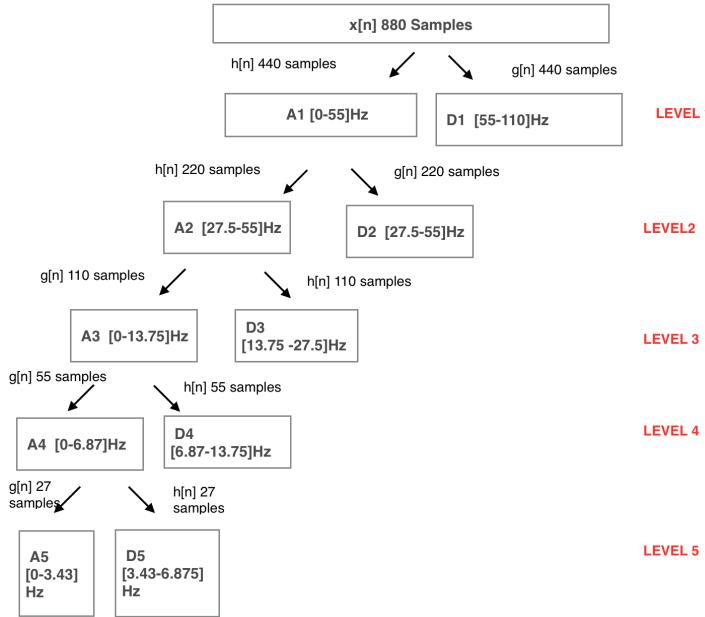

Figure 8.2: Example of subband decomposition of DWT

Decomposed Coefficients	Frequency bands [Hz]	Decomposition level
D1	55-110	1
D2	27.5-55	2
D3	13.75-27.5	3
D4	6.87-13.75	4
D5	3.43-6.85	5
A5	0-3.43	5

Table 8.1: Table of frequencies at each decomposition level

The wavelet coefficients (detail coefficients) and the scaling coefficients (approximation coefficients) were computed by using the "Wavelets" Package in R [1].

The figure 8.3 shows the Details (D1, D2, D3, D4, D5) coefficients and the Approximation or scaling (A5) coefficients and the original EEG signal of the experiments carried out with the participant 10.
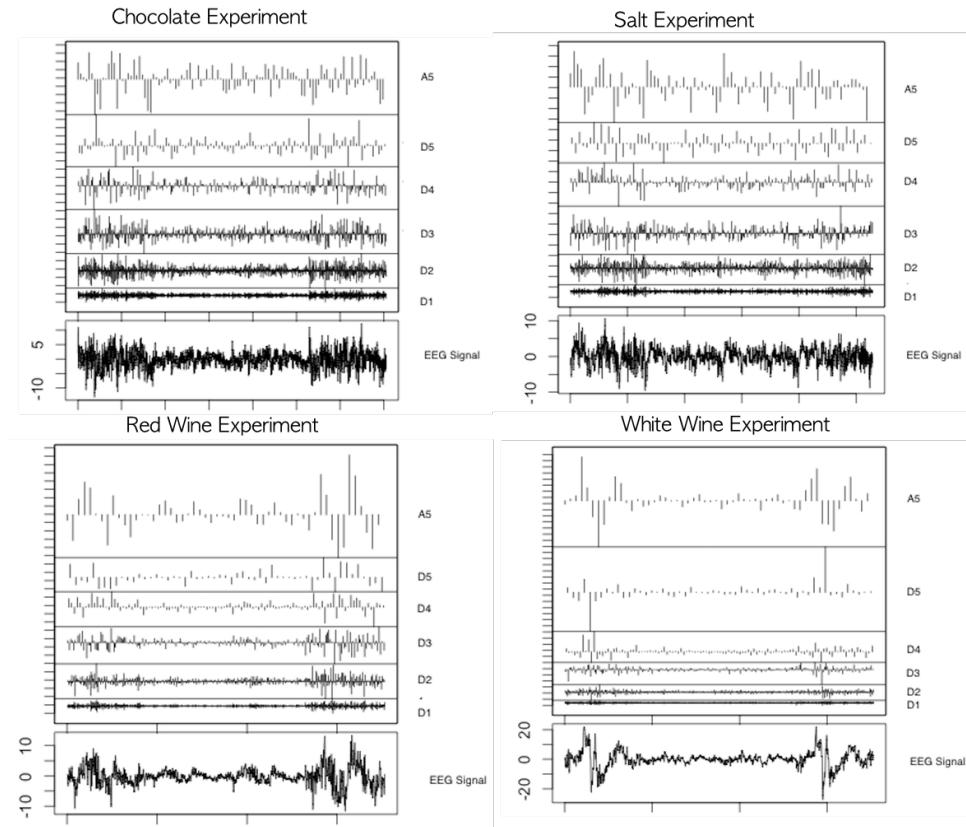

Figure 8.3: Approximation and Details for the experiments of the participant 10

8.2.2 Feature extraction based on the relative wavelet energy

Due to we are interested on the power of the signals in different frequency bands when a stimuli is given, the relative wavelet energies were used as feature extraction for each EEG signal, as proposed by Guo et al. in[27]. As the Wavelet Transform is an orthonormal basis, the concept of energy applies in the same manner as the notions derived from the Fourier theory.

---

[1] http://cran.r-project.org/web/packages/wavelets/wavelets.pdf

The energy of the detail coefficients at different decomposition levels (from 1 to N) is the energy o the wavelet coefficients $d_{j,k}$, and the energy of the approximation decomposition at level N is the energy $a_{N,k}$. Therefore, the detail energy at each decomposition level is defined as:

$$E_j = \sum_k |d_j, k|^2 \tag{8.1}$$

where $j = 1, ..., N$

And the approximation energy at the decomposition level N is given by:

$$E_N = \sum_k |a_k|^2 \tag{8.2}$$

The total energy of the signal after the wavelet decomposition is obtained by:

$$E_{total} = \sum_j^N E_j + E_N \tag{8.3}$$

Thus, the relative wavelet energy for the detail coefficients is defined, as:

$$\rho_j = \frac{E_j}{E_{total}} \tag{8.4}$$

for $j = 1, ..., N$

And the relative wavelet energy for the approximation coefficient is given by:

$$\gamma_N = \frac{E_N}{E_{total}} \tag{8.5}$$

The table 8.2 shows the average Relative Wavelet Energy (RWE) for each of the experiments performed with the participant 10. We can see that as expected in the first level of decomposition which covers the frequency band [55 - 110]Hz the RWE is very small compared with the rest.

RWE	Chocolate Signal	Salt Signal	Red EEG Signal	White wine Signal	Frequency band [Hz]
ρ_1	0.062	0.057	0.045	0.023	55-110
ρ_2	0.233	0.224	0.169	0.104	27.5-55
ρ_3	0.235	0.234	0.171	0.133	13.75-27.5
ρ_4	0.148	0.138	0.086	0.094	6.87-13.75
ρ_5	0.098	0.108	0.108	0.205	3.43-6.85
γ_5	0.224	0.257	0.420	0.441	0-3.43

Table 8.2: Average of RWE during the different experiments of the participant 10

8.2.3 Feature selection

Choosing the features to characterize the signals is crucial for classifying them. Thus, we have analyzed in detail the coefficients given by each experiment for each subject in order to find the most appropriate features that will help us to correctly classify the signals.

As stated by the authors in [28], the positive emotions and the positive judgment are associated with the electrodes FP1 and TP9 respectively meanwhile the negative emotions and negative judgment are associated with the electrodes FP2 and TP10. Having said that, we aimed to see if it held in our experiments [2].

---

[2]http://www.brainm.com/software/muse/MOOD%20MANUAL.pdf

First, we attempt to see any pattern between the RWE at each decomposition level. In figure 8.4, the left side shows the chocolate experiment realized to the participants 10, 11 and 12, where each line in the plot represents a particular channel, the right side shows the results when the channels FP1-TP9 (green line) and FP2-TP10 (red line) were added, we did the same for the experiment with salt, figure 8.5, and with red wine, figure 8.6
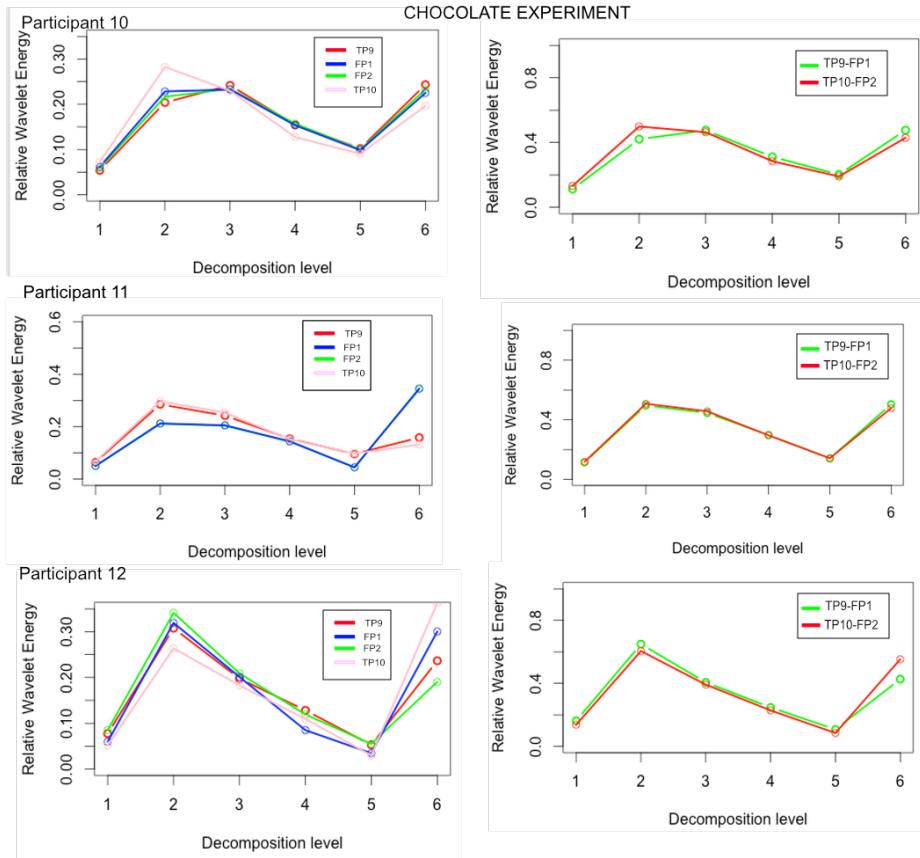

Figure 8.4: Distribution of the channels trough the decomposition levels for the Chocolate experiment in participants 10,11 and 12. Left side shows the distribution of the four channels whereas the right side shows the distribution of the channels TP9-FP1(green line) and FP2-TP10 (red line)
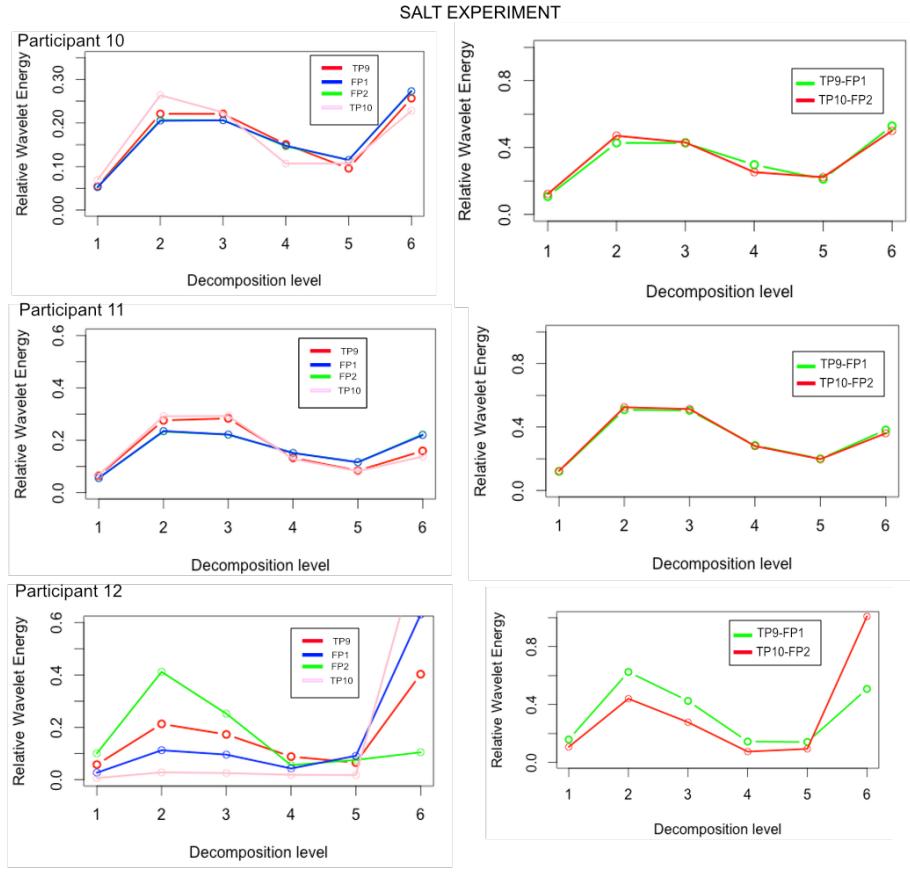
Figure 8.5: Distribution of the channels trough the decomposition levels for the Salt experiment in participants 10,11 and 12. Left side shows the distribution of the four channels whereas the right side shows the distribution of the channels TP9-FP1(green line) and FP2-TP10 (red line)
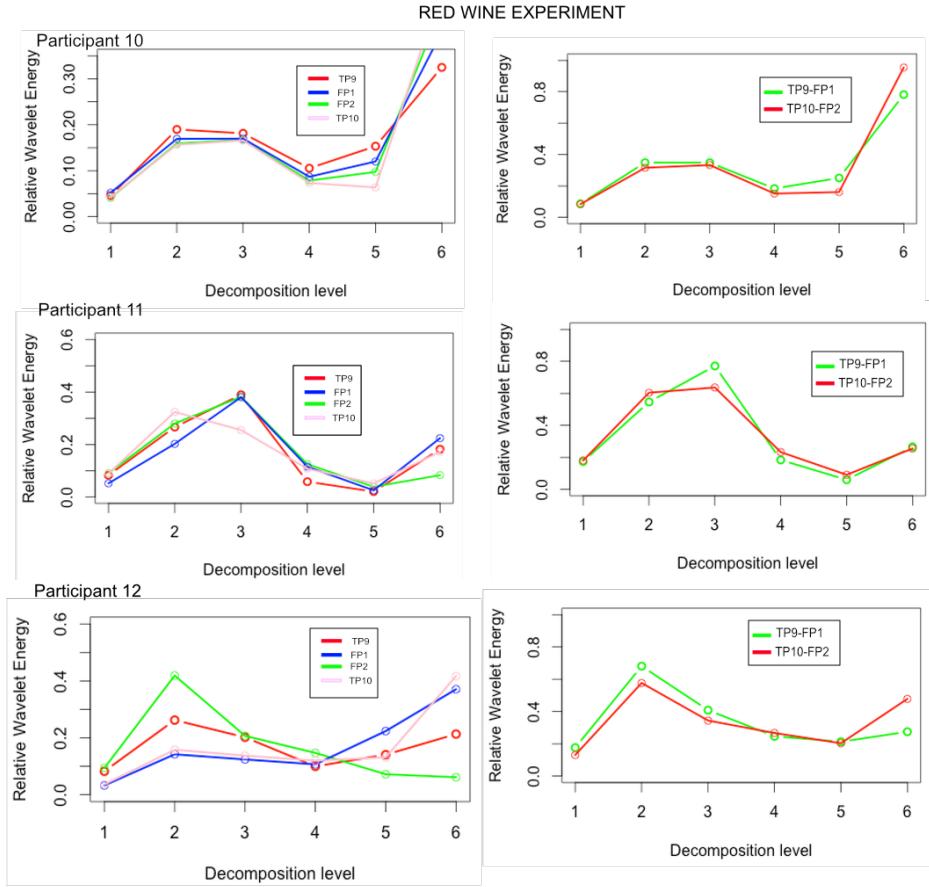
Figure 8.6: Distribution of the channels trough the decomposition levels for the Red wine experiment in participants 10,11 and 12. Left side shows the distribution of the four channels whereas the right side shows the distribution of the channels TP9-FP1(green line) and FP2-TP10 (red line)

As we can observe there is not any specific pattern that we can follow, as each participant's waves behave in a different manner for each experiment.

Therefore, our second attempt was to find the distributions of the decomposition levels trough the channels, first with the four channels and then by adding channels TP9-FP1 and TP10-FP2.
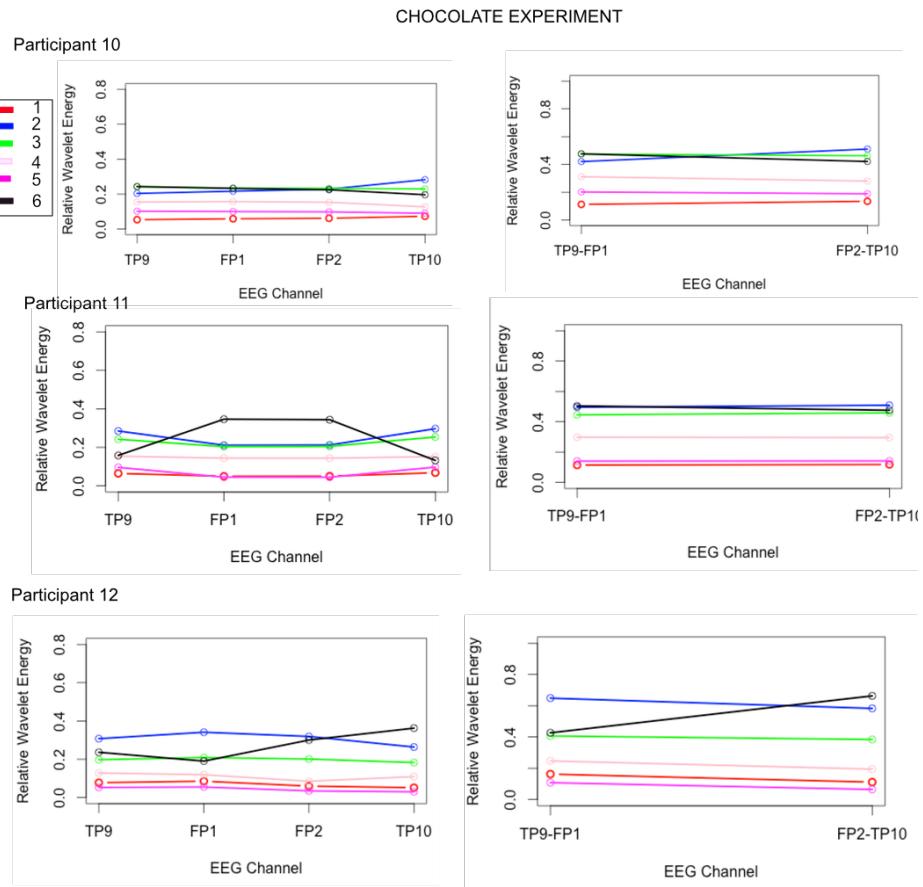
Figure 8.7: Distribution of the decomposition levels through the channels for the Chocolate experiment in participants 10,11 and 12. Left side shows the distribution of the decomposition levels over the four channels whereas the right side shows the distribution of them through the added channels.
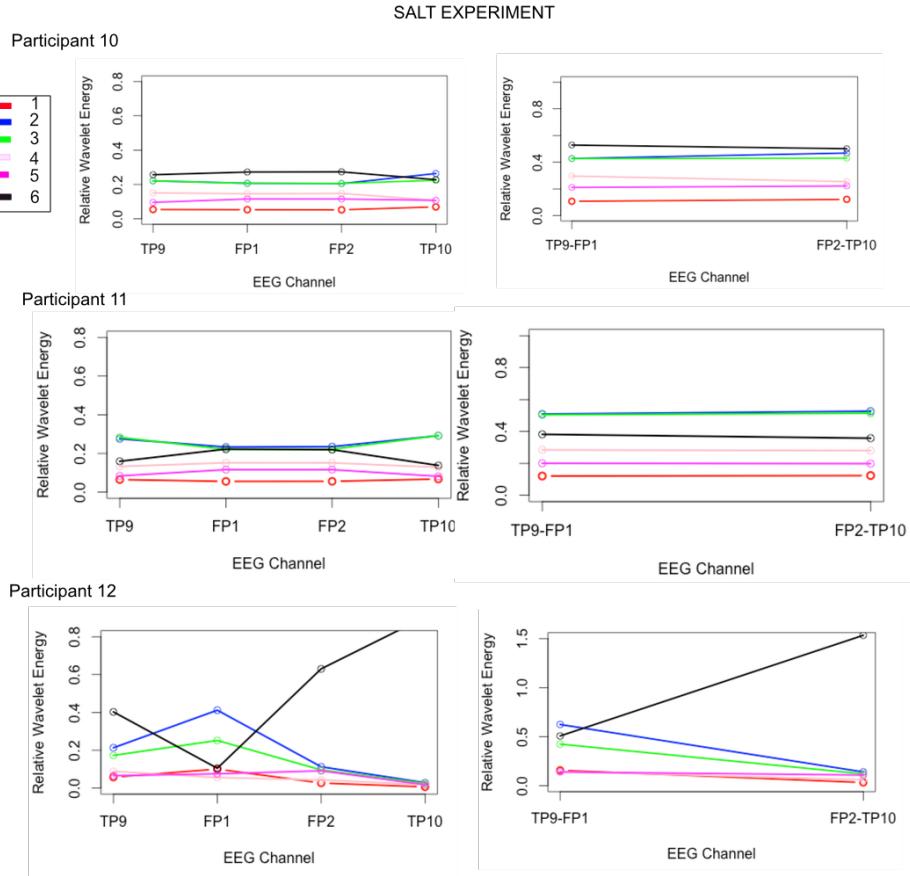
Figure 8.8: Distribution of the decomposition levels through the channels for the Salt experiment in participants 10,11 and 12. Left side shows the distribution of the decomposition levels over the four channels whereas the right side shows the distribution of them through the added channels.
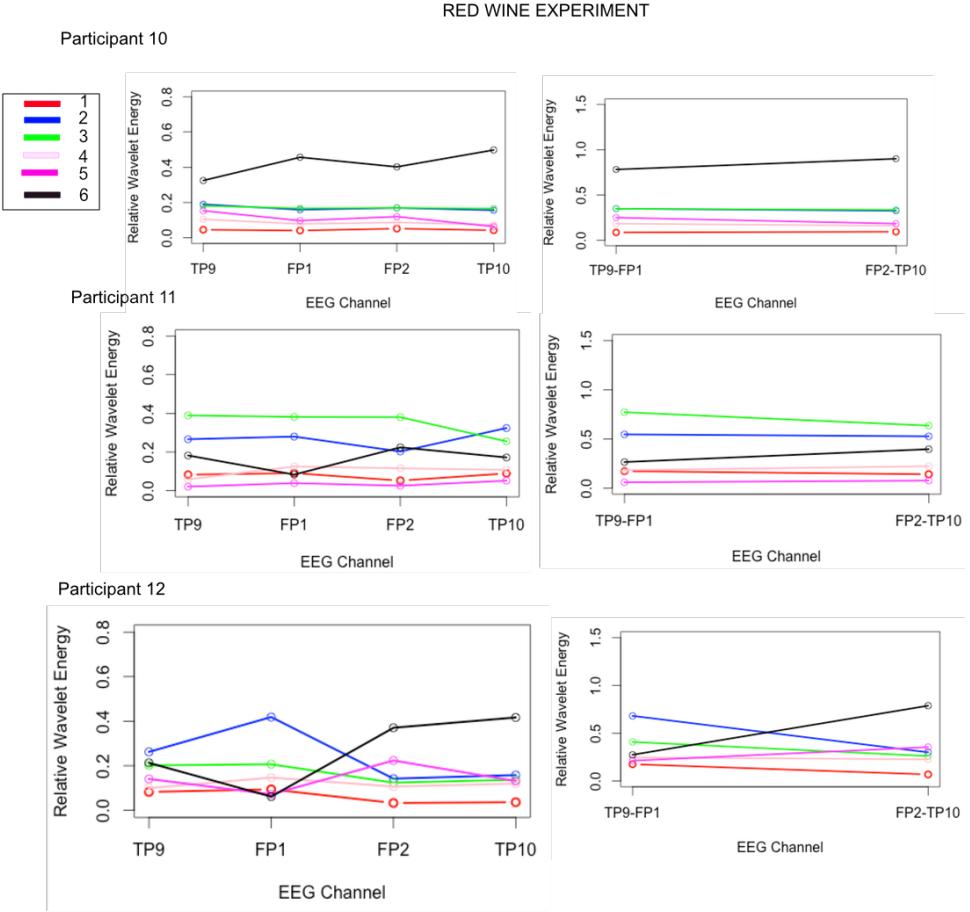
Figure 8.9: Distribution of the decomposition levels through the channels for the Red wine experiment in participants 10,11 and 12. Left side shows the distribution of the decomposition levels over the four channels whereas the right side shows the distribution of them through the added channels.

In figure 8.7, the left side shows the distribution of the decomposition levels over the EEG channels (TP9, FP1, FP2 and TP10) whereas the right side shows the distribution of the decomposition levels over the added EEG channels TP9-FP1 and FP2-P10. Unfortunately, this analysis only supports the previous one, the experiments are self-dependent and the stimuli reacts in different manners for each subject.

Due to in the first analysis the distribution of the waves tend to be more similar,we decided to generate the featured vectors by taking the RWE levels (1-6) for each channel (TP9, FP1, FP2,TP10), producing a vector with 24 parameters for each different experiment.

8.3 Feature classifier: Probabilistic Neural Network (PNN)

For classifying the emotions risen in each experiment according to the stimuli, we used as a classifier the Probabilistic Neural Network (PNN) as proposed in [17]. The PNN was implemented in R by using the "PNN" package [3]. The training set was the featured vectors for each experiment of the participant 11, and the testing sets were featured vectors for the participants 10 and 12, due to the propose of this work is to know if we can generalize the emotions perceived in each stimuli for any person. We binarized the classes in such a way that if the subject graded the experiment bigger than 3 is positive (1) and lower of

---

[3]http://cran.r-project.org/web/packages/pnn/pnn.pdf

it, it is negative(0).

8.3.1   Results

The table 8.3, summarizes the results obtained, when the class was binarized.

Predicted result	Expected Result
1	1
1	0
1	1
1	1
1	1
1	0
1	1
1	1

Table 8.3: Results of the PNN, Predicted Vs. Expected

Since all the emotions were classified as only one, it means that we cannot distinguish between them, we did a further experiment by using the "original" scale from 1 to 5, and the results can be observed in table 8.4

Predicted result	Expected Result
3	4
3	1
3	5
4	4
3	3
3	1
3	5
4	4

Table 8.4: Results of the PNN, Predicted Vs. Expected

As we expected from the analysis of the EEG signals with DWT, with the results obtained we cannot clearly conclude anything, besides that the sensations or emotions provoked by the stimuli are self-dependent.

Chapter 9

Interpretation of the Results

After the development and individual evaluation of each model, this chapter will compare the models and propose a common explanation for the results obtained by them.

With the results obtained by the WDNN model 7 and Probabilistic Neural Networks 8 it is possible to conclude that the Muse device cannot be used to identify task stimuli in different subjects.

Obtaining similar results with two completely distinct approaches gave us the confidence to say that the expected task cannot be fullfilled in a satisfying way with the presented equipment. The next immediate step is to determine the reason for this. Several theories can be developed regarding the reasons for the obtained results and we discuss about some of them below.

- The difference found in the brain of different subjects result in very "personal" EEG signals where each data signal is more correlated to the subject producing the data than to the food stimuli. Therefore, EEG signals could be compared to a physiological signature such as fingerprints. It might be useful in the classification of neural diseases.

- The device used for the collection of the signal: Muse, inserts a considerable amount of noise in the data that cannot be removed by our pre-processing. With this theory we propose two different assumptions:

  1. the noise of the data is related to the subject and therefore, a subject that inserts more noise in a stimulus signal will also insert more noise in another different stimulus.

  2. The position of the electrodes in the Muse device are not appropriate for experiments involving taste stimuli. With this assumption we consider that the electrodes positioned in the forehead of the subject are highly exposed to noise during the record of the EEG signals, especially when taste stimuli are used and the subject considerably moves its head.

  3. By analyzing in detail the EEG signals with WDT, we confirmed that, between the three participants, there were no pattern found. Thus, the next step is to run more experiments with at least 15 subjects to confirm this preliminary version.

  4. In the PNN, we found that when the class is binarized as only pleasant and unpleasant, we got an accuracy of 75%, however this is of no relevance since everything was classified as only one class. However, when we respect the original values for the class, the accuracy is 37.5%. These bad results could be derived by two possible situations, either there is no way in which we can classify the emotions provoked by stimuli with the MUSE device or the participants were too few and we do need more participants for training and for testing.

Chapter 10

Conclusions and further work

The human brain is a complex system which exhibits spatiotemporal dynamics. In this work, we attempt to associate the brain activity with emotions evoked by different stimuli. However, it was not possible to classify emotions stimulated with "food" stimuli, with the MUSE headset. This low-cost EEG headset contains most of the sensors in only one line located altogether in the forehead making it more sensible to EMG noise, which was usually generated when the participants tasted the different flavors. During the cleaning process, by trying to remove the EMG noise we possibly lost some important information.

During the development of this Case Study project we had the opportunity to face a complete data mining process, applied to a non-trivial type of data: EEG signals. The EEG signals compose a type of time-series dataset, which make its analysis more complicated than a normal stationary dataset. During this work, we have tested and learned a lot of techniques in order to extract features, interpret and classify EEG signals such as Fourier analysis, STFT, DWT, AR, WDNN, SVM and PNN. From this, we can conclude that the best technique to extract features from the signals is the Wavelet Transform, since the design of this method fits perfectly with the behavior of the biological signals, specially with EEG, EMG and ECG signals. For this study, we only used supervised algorithms for classifying the signals, thus as a future work, we propose to implement an unsupervised algorithm for clustering the signals such as SOM (Self Organizing Maps) or also known as unsupervised neural networks.

As a further work of this project, we suggest the test of the models developed here with a different type of low cost EEG equipment. A test like this would possibly determine the cause of the obtained results, as proposed in section 9.

Finally, based on the data collected and on the distinct analytic models that we have presented, we have to conclude that it is not possible to get reliable and robust classification for the identification of taste-based emotions with the MUSE device. According to our observations, these outcome might be similar for other low-cost EEG devices.

Bibliography

[1] M. Teplan, "Fundamentals of eeg measurement."

[2] S. Sanei and J. A. Chambers, *EEG signal processing.* John Wiley & Sons, 2013.

[3] S. Cheemalapati, M. Gubanov, M. Del Vale, and A. Pyayt, "A real-time classification algorithm for emotion detection using portable eeg," in *Information Reuse and Integration (IRI), 2013 IEEE 14th International Conference on.* IEEE, 2013, pp. 720–723.

[4] M. Murugappan, M. Rizon, R. Nagarajan, S. Yaacob, D. Hazry, and I. Zunaidi, "Time-frequency analysis of eeg signals for human emotion detection," in *4th Kuala Lumpur International Conference on Biomedical Engineering 2008.* Springer, 2008, pp. 262–265.

[5] D. Nie, X.-W. Wang, L.-C. Shi, and B.-L. Lu, "Eeg-based emotion recognition during watching movies," in *Neural Engineering (NER), 2011 5th International IEEE/EMBS Conference on.* IEEE, 2011, pp. 667–670.

[6] D. O. Bos, "Eeg-based emotion recognition," *The Influence of Visual and Auditory Stimuli*, pp. 1–17, 2006.

[7] M. Murugappan, M. Rizon, R. Nagarajan, S. Yaacob, I. Zunaidi, and D. Hazry, "Eeg feature extraction for classifying emotions using fcm and fkm," *International journal of Computers and Communications*, vol. 1, no. 2, pp. 21–25, 2007.

[8] E. Parvinnia, M. Sabeti, M. Z. Jahromi, and R. Boostani, "Classification of eeg signals using adaptive weighted distance nearest neighbor algorithm," *Journal of King Saud University-Computer and Information Sciences*, vol. 26, no. 1, pp. 1–6, 2014.

[9] G. Repovs, "Dealing with noise in eeg recording and data analysis." *Informatica Medica Slovenica*, vol. 15, no. 1, pp. 18–25, 2010.

[10] E. C. Ifeachor and B. W. Jervis, *Digital signal processing: a practical approach.* Pearson Education, 2002.

[11] G. Gómez-Herrero, W. De Clercq, H. Anwar, O. Kara, K. Egiazarian, S. Van Huffel, and W. Van Paesschen, "Automatic removal of ocular artifacts in the eeg without an eog reference channel," in *Signal Processing Symposium, 2006. NORSIG 2006. Proceedings of the 7th Nordic.* IEEE, 2006, pp. 130–133.

[12] W. De Clercq, A. Vergult, B. Vanrumste, W. Van Paesschen, and S. Van Huffel, "Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram," *Biomedical Engineering, IEEE Transactions on*, vol. 53, no. 12, pp. 2583–2587, 2006.

[13] S.-F. Wang, Y.-H. Lee, Y.-J. Shiah, and M.-S. Young, "Time-frequency analysis of eegs recorded during meditation," in *Robot, Vision and Signal Processing (RVSP), 2011 First International Conference on*. IEEE, 2011, pp. 73–76.

[14] C. Fadzal, W. Mansor, L. Khuan, and A. Zabidi, "Short-time fourier transform analysis of eeg signal from writing," in *Signal Processing and its Applications (CSPA), 2012 IEEE 8th International Colloquium on*. IEEE, 2012, pp. 525–527.

[15] S. Valipour, A. Shaligram, and G. Kulkarni, "Spectral analysis of eeg signal for detection of alpha rhythm with open and closed eyes," *International Journal of Engineering and Innovative Technology (IJEIT)*, vol. 3, no. 6, pp. 1–4, 2013.

[16] F. Yger and A. Rakotomamonjy, "Wavelet kernel learning," *Pattern Recognition*, vol. 44, no. 10, pp. 2614–2629, 2011.

[17] E. D. Übeyli, "Probabilistic neural networks combined with wavelet coefficients for analysis of electroencephalogram signals," *Expert Systems*, vol. 26, no. 2, pp. 147–159, 2009.

[18] P. Stoica and R. L. Moses, *Spectral analysis of signals*. Pearson/Prentice Hall Upper Saddle River, NJ, 2005.

[19] S. De Waele and P. M. Broersen, "The burg algorithm for segments," *Signal Processing, IEEE Transactions on*, vol. 48, no. 10, pp. 2876–2880, 2000.

[20] F. Cervantes-De la Torre, J. González-Trejo, C. Real-Ramírez, and L. Hoyos-Reyes, "Fractal dimension algorithms and their application to time series associated with natural phenomena," in *Journal of Physics: Conference Series*, vol. 475, no. 1. IOP Publishing, 2013, p. 012002.

[21] E. Fix and J. L. Hodges Jr, "Discriminatory analysis-nonparametric discrimination: consistency properties," DTIC Document, Tech. Rep., 1951.

[22] J. Gou, L. Du, Y. Zhang, and T. Xiong, "A new distance-weighted k-nearest neighbor classifier," *J. Inf. Comput. Sci*, vol. 9, pp. 1429–1436, 2012.

[23] MUSE, "Muse: The brain sensing headband tech spec sheet."

[24] R. W. Homan, J. Herman, and P. Purdy, "Cerebral location of international 10–20 system electrode placement," *Electroencephalography and clinical neurophysiology*, vol. 66, no. 4, pp. 376–382, 1987.

[25] C. Park, D. Looney, and D. Mandic, "Estimating human response to taste using eeg," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*. IEEE, 2011, pp. 6331–6334.

[26] S. Rousmans, O. Robin, A. Dittmar, and E. Vernet-Maury, "Autonomic nervous system responses associated with primary tastes," *Chemical senses*, vol. 25, no. 6, pp. 709–718, 2000.

[27] L. Guo, D. Rivero, J. A. Seoane, and A. Pazos, "Classification of eeg signals using relative wavelet energy and artificial neural networks," in *Proceedings of the first ACM/SIGEVO Summit on Genetic and Evolutionary Computation*. ACM, 2009, pp. 177–184.

[28] T. Collura, R. Zalaquett, C. Bonnstetter, and S. Chatters, "Toward an operational model of decision making, emotional regulation, and mental health impact." *Advances in mind-body medicine*, vol. 28, no. 4, pp. 18–33, 2013.