SEPARATION OF STIMULUS-SPECIFIC PATTERNS IN ELECTROENCEPHALOGRAPHY DATA USING QUASI-SUPERVISED LEARNING

A Thesis Submitted to
the Graduate School of Engineering and Sciences of
İzmir Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

in Electronics and Communication Engineering

by
Başak Esin KÖKTÜRK

December 2011
İZMİR

We approve the thesis of **Başak Esin KÖKTÜRK**

Assoc. Prof. Dr. Bilge KARAÇALI
Supervisor

Prof. Dr. Murat ÖZGÖREN
Committee Member

Assist. Prof. Dr. Mustafa Aziz ALTINKAYA
Committee Member

5 December 2011

Prof. Dr. Ferit Acar SAVACI
Head of the Department of
Electrical and Electronics Engineering

Prof. Dr. R. Tuğrul SENGER
Dean of the Graduate School of
Engineering and Sciences

ACKNOWLEDGEMENTS

I would like to express my gratitude to my advisor Assoc. Prof. Dr. Bilge Karaçalı not only for his guidance but also for his support, trust and recommendations and encouragement throughout my thesis study.

I also appreciate to Prof. Dr. Murat Özgören for his guidance and recommendations.

I would also like to express my gratitude to my committee members, Assist. Prof. Dr. Mustafa Aziz Altınkaya, Assoc. Prof. Dr. Mehmet Engin and Assist. Prof. Dr. Zübeyir Ünlü for their guidance and contributions.

I am grateful to my lab mates Devrim Önder and Tunca Doğan for their patient and helps to me. They always deal with my thesis and offer suggestions.

I thank to all my friends especially to Kağan Kılıçaslan and Dilek Çetin for their unfailing encouragement, neverending friendship and support.

I thank my parents their endless encouragement and loving support during my whole life.

My warmest thanks go to Onur Parlak on whose shoulder I can cry.

ABSTRACT

SEPARATION OF STIMULUS-SPECIFIC PATTERNS IN ELECTROENCEPHALOGRAPHY DATA USING QUASI-SUPERVISED LEARNING

In this study separation of the electroencephalography data recorded under different visual stimuli is investigated using the quasi-supervised learning algorithm. The quasi-supervised learning algorithm estimates the posterior probabilities associated with the different stimuli, thus identifying the EEG data samples that are exclusively specific to their respective stimuli directly and automatically from the data. The data used in this study contains 32 channels EEG recording under six different visual stimuli in random successive order. In our study, we have first constructed EEG profiles to represent instantaneous brain activity from the EEG data by various combinations of independent component analysis and the wavelet transform following data pre-processing. Then, we have applied the binary and M-ary quasi-supervised learning to identify condition-specific EEG profiles in different comparison scenarios. The results reveal that the quasi-supervised learning algorithm is successful in capturing the distinction between the samples. In addition, feature extraction using independent component analysis increased the performance of the quasi-supervised learning and the wavelet decomposition revealed the different frequency bands of the features, making more explicit the separation of the samples. The best results we obtained by combining the wavelet decomposition and the independent component analysis before the quasi-supervised learning algorithm.

ÖZET

ELECTROENSEFALOGRAM VERİLERİNDE UYARANA BAĞLI ÖRÜNTÜLERİN YARIGÜDÜMLÜ ÖĞRENME İLE AYRIŞTIRILMASI

Bu çalışmada, değişik görüntü uyaranları altında kaydedilmiş elektroensefalografi verilerinin yarıgüdümlü öğrenme yöntemi ile ayrıştırılması ele alınmıştır. Yarı-güdümlü öğrenme yöntemi değişik uyaranlara bağımlı olasılıkları tahmin etmekte böylelikle de EEG veri örneklerinin otomatik olarak ait oldukları sınıfları belirlemektedir. Bu çalışmada ardışık olarak rastgele sıralanmış altı farklı görüntü uyaranı altında çekilmiş 32 kanallı EEG verisi kullanılmıştır. Çalışmamızda öncelikle anlık beyin aktivitelerine karşılık gelen EEG verileri bağımsız bileşen analizi ve dalgacık dönüşümünün değişik kombinasyonları ile ön işlemden geçirilmiştir. Ardından ikili ve çoklu yarı-güdümlü öğrenme yöntemi uygulanarak karşılaştırma senaryoları ile koşula özgü EEG profilleri tanımlanmıştır. Elde ettiğimiz sonuçlar göstermektedir ki, yarı güdümlü öğrenme algoritması örnekler arasındaki ayırımı yakalamakta ve anormalliklerin tesbitinde başarılıdır. Bağımsız bileşen analizinin yarı-güdümlü öğrenme yönteminin performansını arttırmada başarılı olmasının yanı sıra dalgacık dönüşümü ile değişik frekans bantlarının belirlenmesi örnekler arası ayrımı belirginleştirir. En iyi sonuçlar yarı-güdümlü öğrenme yönteminden önce bağımsız bileşen analizi ve dalgacık dönüşümünün kombine edilerek uygulanması ile elde edilir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Table **Page**

CHAPTER 1

INTRODUCTION

Brain is the center of the nervous system in all vertebrate animals and most of the invertebrate animals. It has a very complex structure and it is the most important organ. It contains 100 billion nerve cells and they produce electrical activity. This electrical activity is important to understand the brain function and the associated signals can be recorded for monitoring by neuroimaging methods such as electroencephalography (EEG) which records electrical activity, functional magnetic resonance (fMRI) which measures changes in cerebral blood oxygenation levels [1], magnetoencephalography (MEG) which measures the magnetic fields [2] and so on. Brain imaging methods provide opportunity to study the brain function.

The electrical activity of the brain was known in the early nineteenth century, recorded first in 1928 by Hans Berger. This recording is the primitive electroencephalography.

In recent years, development of EEG technology increased its ability to read brain activity data from the head simultaneously. It has some disadvantages in the spatial resolution but the temporal solution of the EEG is superior, since EEG is the record of time series electrical activity signals that are captured by many electrodes. EEG is fast and it can record complex activity in the brain after a stimulus onset. This feature makes it popular for brain-computer interface applications as well as in medical research on various diseases such as epilepsy, Parkinson and Alzheimer's diseases [3-5].

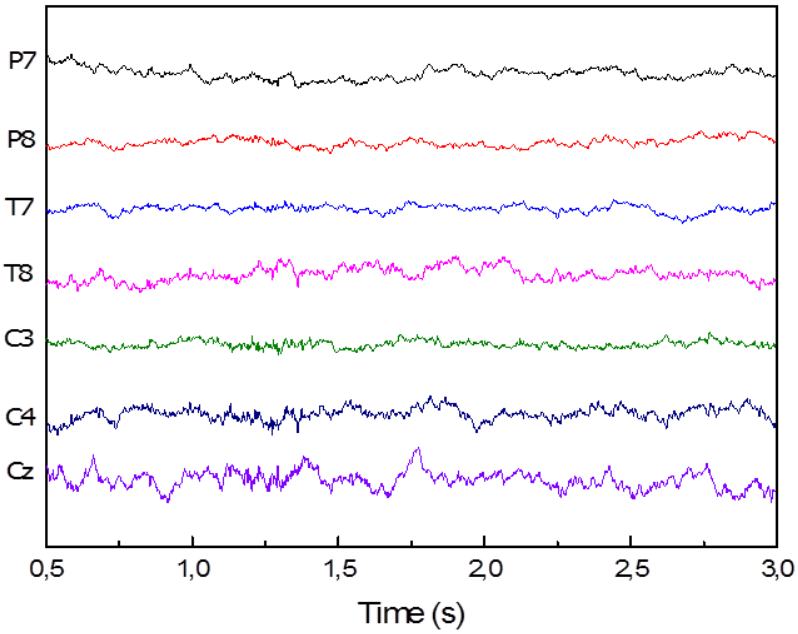

Figure 1. Conventional EEG representation. Measured voltages (in μV) are shown against time (in seconds) for seven electrodes (P7, P8, T7, T8, C3, C4, Cz)

According to the Web of Science Database (data collected in 24.10.2011), the numbers of publications with EEG-related topics were as follows:

Table 1. Numbers of publications with EEG-related topics

Keyword	**Hits/Publications**
EEG	54.279
Brain activity pattern	16.736
EEG Classification	3.091
Brain-Computer Interface	2.703

The numbers show that; understanding brain function, especially using EEG, is a very popular research area. EEG data classification and the extraction of brain activity patterns address detection of diseases such as alcoholism, schizophrenia and the Alzheimer's Disease. For instance, in alcoholic patients, the distinctions between

stimuli-specific brain activity patterns or samples of attention task problems are difficult. They exhibit a higher overlap between stimuli or tasks than normal subjects [6]. EEG research is important not only in detection of diseases but also in brain-computer interface applications. In physically disabled subjects, the brain function is intact but they are unable to carry out physical tasks using muscular activity. Brain-computer interface systems aim to provide them the ability to give commands to a computer directly through EEG. Since the brain functions are complex and contain many background information, classification of EEG data is challenging.

There are many classification methods that are used in brain-computer interface applications [7]. In EEG classification problems, generally supervised learning algorithms are used and they bring their own set of challenges. In supervised learning algorithms, before the algorithm is applied to the data, many preprocessing steps are needed. For example; in a stimuli-specific brain activity pattern classification problem, an expert must determine the stimuli-specific activity patterns' characteristic regions and manually extract the successful trials from the data as a training dataset for the classification algorithm. This preprocessing entails manual subtraction of data and is subject to interpretation differences by different experts.

In this study, we have applied an analytical method to understand the EEG behavior and extract the information about distinctions between samples according to a visual stimulus. This analytical method, Quasi-Supervised Learning Algorithm [6] , estimates the posterior probabilities of each sample according to a stimulus. It's a new formulation for pattern recognition problems and we used this method in conjunction with the independent component analysis (ICA) [8] and the wavelet transform (WT) [9]. Independent component analysis and wavelet transform are used commonly in EEG classification problems to enhance different feature characteristics [10]. We used the independent component analysis and the wavelet transform in a similar fashion to construct profile vectors on instantaneous EEG data, and employed the quasi-supervised learning algorithm to identify the profiles that are specific to different visual stimuli.

The remainder of this thesis is organized as follows. Chapter 2 provides background information on electroencephalography, the independent component analysis and the wavelet transformation. Chapter 3 provides the original quasi-supervised learning algorithm along with the M-ary quasi-supervised learning algorithm. In Chapter 4, the data used in this thesis is described along with the experiment setup. The results of the experiments can be found in Chapter 5. Chapter 6

provides a concluding summary of the study and closes with potential directions for future research.

CHAPTER 2

BACKGROUND

In this chapter, background information on electroencephalography along with independent component analysis and the wavelet transform commonly used in EEG data analysis is represented. The first section introduces the electrical activity in the brain, the history of the electroencephalography, recording techniques of the electroencephalography and the event-related potentials. In the second section, independent component analysis algorithm is clarified, followed by the wavelet transformation and wavelet decomposition in the third section.

2.1. Electroencephalogram

The word ‘electroencephalogram’ (EEG) originally denoted the graphs that show the signals obtained by currents that flow during excited states of the synapses [11]. In other words, EEG refers to the recording of the electrical activity among the neurons in the brain. Over the years, the EEG has come to be used to refer to the recorded signal itself.

2.1.1. The Brain Structure

The brain consists of about 10^{10} nerve cells. Nerve cells respond to sensory stimuli and transmit information. In the human brain, each nerve cell is connected approximately to 10.000 other nerve cells [12]. Communication among these nerve cells or neurons is provided through electrical activity between dendrites and axons. Dendrites are specialized for receiving inputs from other neurons. Via the axon, impulses are sent to other neurons’ dendrites. There is a special interface among two neurons called a synapse. The electrical activity and communication among neurons occurs in the synapses. When neurons are activated, synaptic currents are produced.

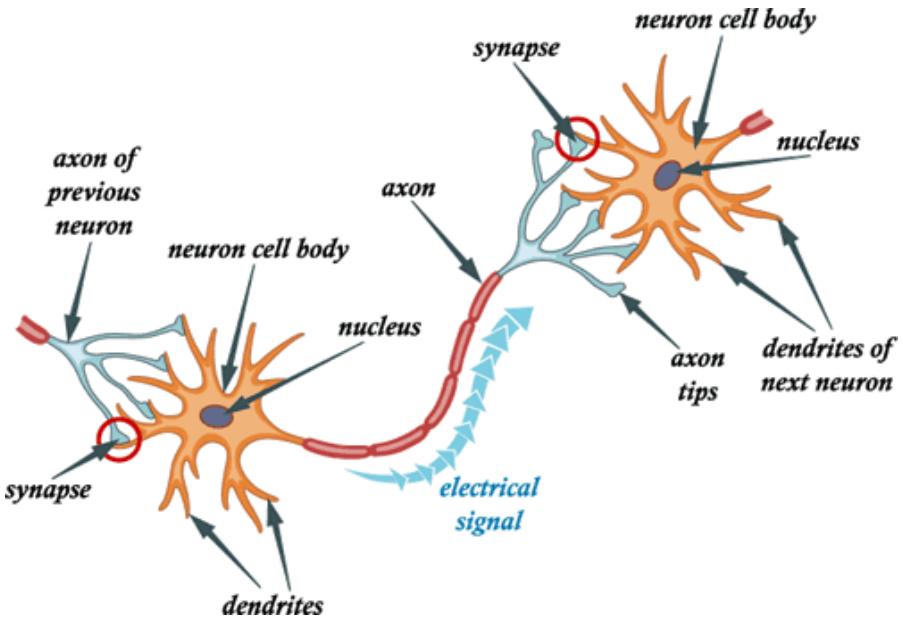

Figure 2. Electrical Signal Production[1]

The information is transmitted between neurons via action potentials (APs). APs are caused by transfer of ions across the membrane along the axon. APs can be created in different types of stimuli; such as chemical, light, pressure, touch etc. To trigger an AP, a stimulus must be applied higher than the activation threshold. When stimuli strength exceeds the threshold, an AP appears and travels along the nerve.

The formation process of an AP involves several steps (see in Fig-3) [12]

1- The spike of an AP is caused by the opening of Na^+ channels.
2- Up to 20 mV of membrane potential the Na^+ ions flow into the cell membrane. This process is called membrane depolarization.
3- When the membrane potential exceeds the +30 mV, Na^+ channels close and K^+ channels open.
4- The membrane begins to repolarize back towards the rest potential due to the open K^+ channels.
5- Repolarization process continues until -90 mV. This level is called as hyperpolarization. In hyperpolarization, generation of a new action potential is prevented.

[1] Available on : http://www.culverco.com/sse/images/body/1-4ci.gif

6- After hyperpolarization, Na+/K+ ions comes to equilibrium and the membrane potential returns to the resting state -70 mV.

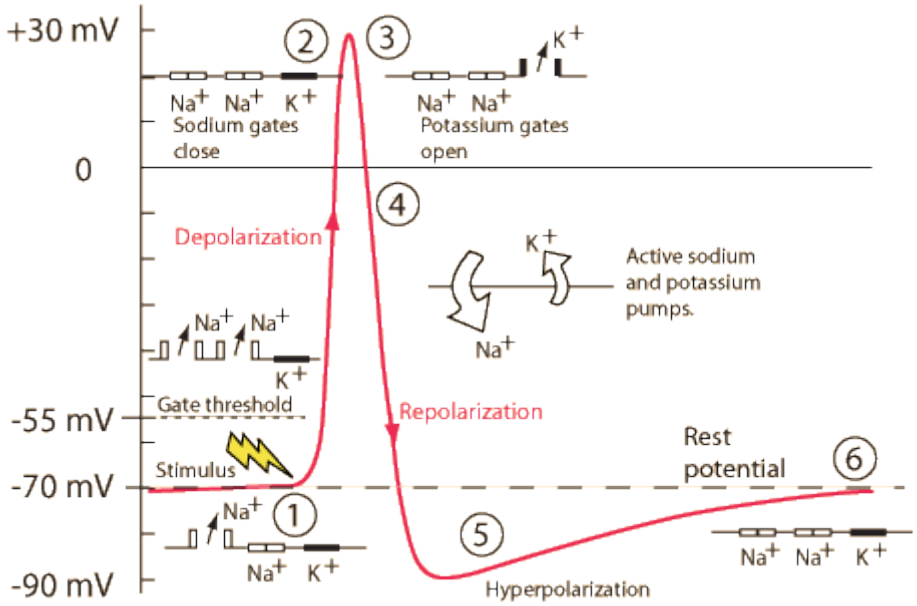

Figure 3. Change of the membrane potential after an stimuli onset
(Source: Charand 2011)

One neuron can generate a small amount of electrical activity and cannot be measured from the scalp. When a large group of neurons activate simultaneously, however, they can generate enough electrical activity to be detected by electrodes positioned on the scalp. These activities generate the EEG signal.

2.1.1.1. Brain Rhythms

EEG signal is a combination of the rhythmic activities in the brain with different characteristic frequency bands. The limits between these frequency bands, however, are not strict. They can change with age or due to the effects of drugs [14]. The signal has a very broad frequency range but the main clinical and physiologic activities are observed between 0.5 Hz – 30 Hz. Some frequency–based categories of the EEG signal are described in Table 2.

Table 2. EEG frequency bands
(Source: Lofhede 2009)

Name	Frequency Limits	Location	Properties
δ (delta)	0.5-4 Hz	Widespread	Occurs in infants and during deep sleep or anesthesia.
θ (theta)	4-8 Hz	Mainly in parietal and temporal lobes	Most prominent in small children and during drowsiness or sleep
α (alpha)	8-13 Hz	Rear half of the head	Occur during awake and resting state, high amplitude when eyes closed.
β (beta)	Above 13 Hz	Most common in frontal and central regions	Often divided in two sub-bands, of which the higher frequencies appear during tension and intense activation of the CNS and the lower are attenuated during mental activity.

Delta (δ) waves occur in the range between 0.5 Hz – 4 Hz and can be seen in infants during deep sleep or under anesthesia. Delta waves can be confused easily with artefacts caused by muscular activity. Theta waves lie between 4 Hz – 8 Hz and they occur in the parietal and temporal lobes of the brain. It is dominated in small children, during drowsiness or sleep and also seen in meditation. Theta waves play an important role in childhood, if the theta waves are dominant in the awake adult, there is likely an abnormality caused by various pathological problems [12]. The principal resting rhythm of the brain is the alpha rhythm. It lies between 8 Hz – 13 Hz and is common in wakeful, resting adults. The EEG power in the alpha waves range is related to cognitive performance and brain maturity [7]. The alpha waves exhibit high amplitudes when the eyes are closed; when the subject opens the eyes or receive an alert; the amplitude of the alpha waves diminishes. Beta waves are the electrical activity of the brain at a frequency than 13 Hz and are common in the frontal and the central regions of the scalp. Generally, they are seen in the waking and active thinking periods. The beta waves are predominant while solving a problem or during an attention requiring task. When a person is in panic, the amplitude of the beta waves increase.

2.1.2. History of the Electroencephalogram

A German neurologist, Hans Berger (1873-1941), became interested in the electrical activity in the brain when he was studying temperature changes in the dog cortex [15]. He used radio equipment to amplify the brain electrical activity and recorded it on graph paper in 1924. His first report in 1929 included the alpha rhythm which is the major component of the electroencephalography signal.

During the 1930's, the first EEG recording of sleep was also taken by Berger. The first biological amplifier was designed by Jan Friedrich Toennies and in 1932, a differential amplifier was produced by the Rockefeller foundation. At the same time frame, another research group in Berlin led by Korn Müller, provided more precise EEG signal recording using this amplifier. They succeed in multi-channel recording. Multichannel recording using a large number of electrodes allows covering a wider brain region. The first epileptic spikes were observed by Fisher and Löwenbach in 1934 [12].

2.1.3. Usage of the Electroencephalogram

Understanding the brain is difficult because of its complexity and that of the processes that are controlled by it. The cheapest and the most painless method of monitoring the brain activity is EEG. This monitoring capability creates a visualization opportunity to help disabled individuals. This area is investigated by the growing brain-computer interfacing field.

EEG is also used in the evaluation of the brain disorders. Most commonly, it has been used to show the type and the location of the activity in the brain during critical events such as a seizure, or in confusion, under coma, tumors and long-term difficulties with thinking and memory. It is also an effective method for early detection of the Alzheimer's Disease [16, 17]. Since 1930s, EEG is the main technique for detecting epilepsy [18].

2.1.4. Recording of the Electroencephalogram

In the clinical practice, several channels of EEG are recorded simultaneously from various locations on the scalp to analyze activities of the different brain regions [19]. To record the EEG signal, a conductive gel is applied on the scalp followed by the placement of the electrodes. Generally, to measure a voltage signal, at least two electrodes have to be used: one of them as the active electrode and the other as the ground. But EEG signals are measured with differential amplifiers where time differences between two electrodes are amplified and the ground electrode is held separate. High quality amplifiers must be used since the brain potentials are very small and the interference caused by the electrical equipment in the vicinity can disrupt the measured brain signals.

The electrodes are often placed according to the 10-20 international electrode placement system with 21 electrodes (illustrated in Fig-4) recommended by The International Federation of Societies for Electroencephalography and Clinical Neurophysiology [19]. In this placement system, the "10-20" term defines the electrode locations by dividing the head into 10% and %20 intervals from the nasion to the inion. First letter of the electrode name indicates the position of the electrode. The letter F is used for the frontal lobe, C for the central lobe, P for the parietal lobe, O for the occipital and T for the temporal lobe. For the left hemisphere, odd numbers are used while the even numbers are used for the right hemisphere. These numbers increase with increased distance from the central lobe and Z denotes the midline electrodes [11].

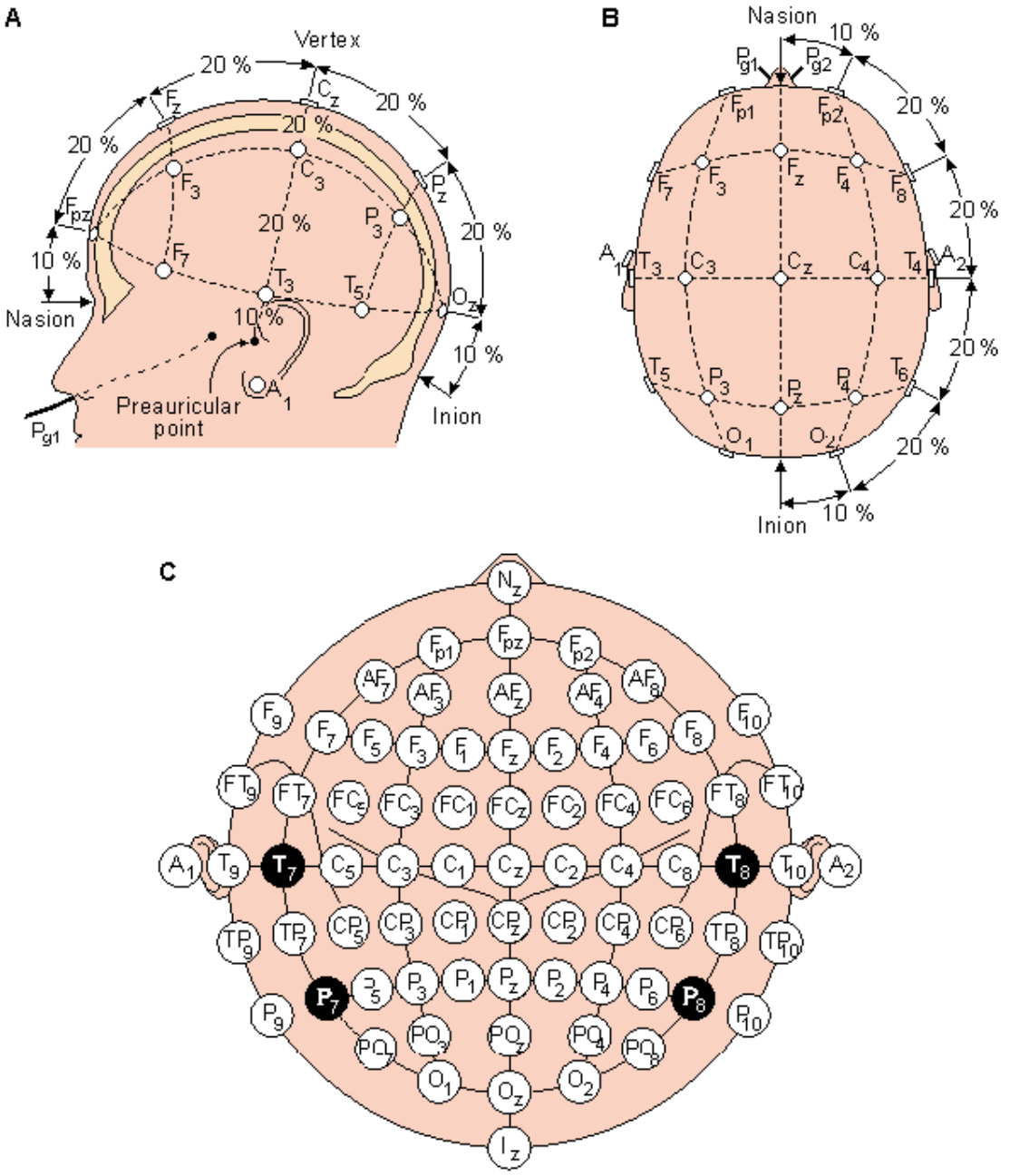

Figure 4. A - B. The international 10-20 electrode placement system C. Extension of standard positions over 70 electrodes [2]

[2] Available on:
http://www.bci2000.org/wiki/index.php/User_Tutorial:EEG_Measurement_Setup

2.1.5. Event Related Potentials

Event-related potentials (ERPs) are the brain's response under any internal or external stimuli (visual, auditory, etc.) or thought. The first ERP recording from an awake individual was taken in 1935. The first study on ERP waveforms was published in 1962 by Galambos et al [9]. After 1964, the modern era of the ERP studies began with Walter et al. who found the first cognitive ERP component that explained the contingent negative variations component of ERP [15]. This proved that consistent patterns could be obtained in spite of the large background data and noise. Another important discovery was the finding of the P300 component by Sutton et al. in 1965 [16]. According to their results, the unpredictable stimuli create large amplitude responses but the predictable stimuli responses were smaller. The P300 signal can be affected by various factors such as the attention level, task difficulty, subject age, stimuli occurrence probability and so on.

ERPs have been used in the literature to understand the brain "language". Under identical stimuli, the same brain responses are expected. But when a stimulus is applied, the brain not only focuses the stimulus but it thinks and works in the background, causing differing brain responses. To make matters worse, the ERPs' amplitudes are typically much smaller than the background EEG data. Furthermore, the ERP changes according to the changing stimuli, and for different stimuli different ERP signals are generated [2].

The ERPs can be present in either P300 or N100 and N400. The digits refer to time in miliseconds after the stimulus onset and the letters P and N indicate the signal amplitude; positive or negative. After 100ms following the stimulus onset, the ERP signal begins to form and makes a peak at around 300ms, called the P300. In Figure 5, ERP is shown after a stimulus onset.

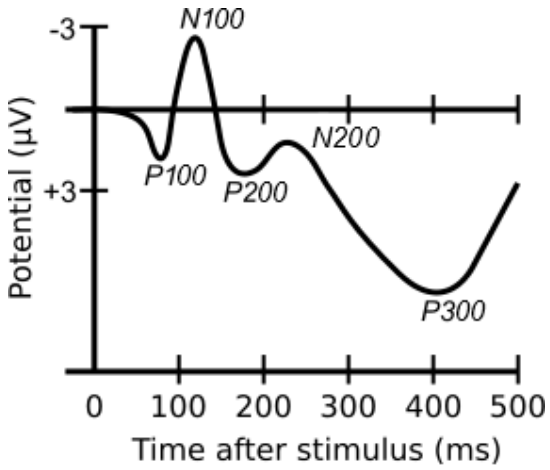

Figure 5. The ERP signal after stimulus onset[3]

2.2. Independent Component Analysis

One of the main problems in multivariate data is finding a suitable representation of the random vectors. This problem can be addressed by a variety of transformations like the principal component analysis, factor analysis and the projection pursuit. Independent Component Analysis (ICA) is an alternative method that finds a refined linear representation of non-Gaussian data. The components identified by the independent component analysis are required to be as statistically independent as possible.

The technical challenge behind independent component analysis can be illustrated by the cocktail party problem. Consider two people speaking simultaneously in a room and two microphones recording the sound signals from different locations. The recordings can be defined as the time signals $x_1(t)$ and $x_2(t)$; and each of these recordings correspond to a weighted mixture of the two speakers. These mixed speech signals can be expressed as

$$x_1(t) = a_{11}s_1(t)+a_{12}s_2(t) \quad (2.1)$$

$$x_2(t) = a_{21}s_1(t) + a_{22}s_2(t) \quad (2.2)$$

[3] Adopted from : http://en.wikipedia.org/wiki/File:ComponentsofERP.svg

where $a_{11}, a_{12}, a_{21}, a_{22}$ denote the mixture rate parameters according to distance. The goal of the independent component analysis task is to recover the source signals $s_1(t)$ and $s_2(t)$, the speech signals of speakers; using only the recorded signals $x_1(t)$ and $x_2(t)$.

EEG recordings resemble the cocktail party problem as the data consists of electric potentials recorded from different locations on the scalp corresponding to weighted mixtures of unknown sources. The number of electrodes, however, is smaller than the number of potential sources (neurons), complicating the analysis. The electric signals propagate in the brain, making the voltage recordings at the electrodes mixed signals of the electrical activity in the brain. Independent component analysis can be used to find independent source signals from the full record [8]. It can be also used to eliminate the time correlation between the different EEG channels and for feature extraction [21].

Consider n linear mixtures of n independent components, where the observed mixtures are denoted by $x_1(t), x_2(t), \dots, x_n(t)$. Then, the j^{th} observation can be expressed as

$$x_j(t) = a_{j1}s_1(t) + a_{j2}s_2(t) + \cdots + a_{jn}s_n(t) \quad \text{for} \quad j = 1, \dots, n \tag{2.3}$$

For simplicity, we take each time instance as a sample and drop the time symbol t. In this way, we can treat each mixture x_j and each independent component s_j as random variables; so the Equation 2.3 can be expressed using a matrix notation.

Let $\mathbf{x}$ denote the random vector whose elements are the mixtures $x_1, x_2, \dots, x_n$ and $\mathbf{s}$ denote the random vector of the sources $s_1, s_2, \dots, s_n$. Let also the matrix $\mathbf{A}$ be the mixture matrix with elements a_{ij}. Using the matrix notation, the Equation 2.3 can be expressed as

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{2.4}$$

The ICA problem represents the task of recovering the source signals $\boldsymbol{s}$ give to observation vectors $\boldsymbol{x}$. To this end, ICA assumes the components s_j to be statistically independent, and tries to find source signals that appear as independent from each other as possible.

Theoretically, two random variables y_1 and y_2 are independent if the information about y_1 does not give any information about y_2, and vice versa. This can be expressed by

$$p(y_1, y_2) = p_1(y_1)p_2(y_2) \tag{2.5}$$

where $p(y_1, y_2)$ denote the joint probability distribution function (pdf) of y_1 and y_2, and $p(y_1)$ and $p(y_2)$ denote the respective marginal distribution functions.

There are different ways of finding independent components, such as maximization of non-Gaussanity, minimization of mutual information, as well as maximum likelihood estimation. Since the sum of independent random variables tends towards a Gaussian distribution by the Central Limit Theorem, independent components can be identified as those that are maximally non-Gaussian. There are several approaches to measure non-Gaussianity. Kurtosis is the fourth order moment μ_4 of a distribution and defined as

$$kurt(y) = E\{y^4\} - 3(E\{y^2\})^2 \tag{2.6}$$

Kurtosis can be described as the degree of peakedness of a distribution [22]. For a Gaussian distribution y, $E\{y^4\} = 3(E\{y^2\})^2$ so the kurtosis of a Gaussian distribution becomes zero. Thus, for measure the non-Gaussianity, kurtosis must be different than zero. It can be positive or negative; if kurtosis is positive, the random variables are super-Gaussian. Otherwise, the random variables are called as sub-Gaussian. Kurtosis can be very sensitive to outliers. Sometimes it can be depend on only a few observations of in the tails of a distribution [23]. Kurtosis is thus not a robust measure of the non-Gaussianity.

Negentropy can also be used to measure non-Gaussianity, and it is defined via the entropy of a random variable. The Shannon entropy H of a random variable Y is defined as

$$H(Y) = -\sum_i P(Y = a_i)\, logP(Y = a_i) \tag{2.7}$$

where a_i are the possible values of Y [24]. The definition of the entropy can be generalized for continuous-valued random variables or vectors in the form of a differential entropy. The differential entropy $H(\boldsymbol{y})$ of a random vector $\boldsymbol{y}$ with density $f(\boldsymbol{y})$ is

$$H(\boldsymbol{y}) = -\int f(\boldsymbol{y}) log f(\boldsymbol{y}) d\boldsymbol{y} \tag{2.8}$$

Negentropy is known as the negative entropy or syntropy. That negentropy is defined as

$$J(\boldsymbol{y}) = H(\boldsymbol{y}_{gauss}) - H(\boldsymbol{y}) \tag{2.9}$$

Gaussian variable has the largest entropy among all random variables of equal variance is among the fundamental results of information theory [8]. As a result, non-Gaussianity can be measured using negentropy, since negentropy is zero for a Gaussian variable and non-negative for non-Gaussian variables.

The main disadvantage of negentropy is the computational cost, because the estimation of negentropy requires estimating the underlying probability density function.

As an alternative to maximization of non-Gaussianity, the ICA problem can be addressed by minimizing the mutual information between the recovered sources. Mutual information is a natural measure of the dependency between random variables. It is always non-negative, and when the random variables are independent, it is zero.

The mutual information I between the random variables y_i where $i = 1,2, \dots, m$ is defined using the (differential) entropy as follows:

$$I(y_1, y_2, \dots, y_m) = \sum_{i=1}^{m} H(y_i) - H(y) \tag{2.10}$$

Since mutual information is the natural information theoretic measure of the independency of a random variable from another, the independent sources can be determined by an unknown transformation minimizing the mutual information.

Finally, the maximum likelihood strategy to the ICA problem is closely connected to infomax principle. Denoting by $W = (w_1, w_2, \dots, w_n)^T$ the separation matrix that recovers the independent components, the log-likelihood takes the form

$$L = \sum_{t=1}^{T} \sum_{i=1}^{n} \log f_i\left(w_i{}^T x(t)\right) + T log|detW| \tag{2.11}$$

where f_i are the density function of source signal s_i, and $x(t),\ t = 1,2, \dots, T$, are the observations. [24].

In a close parallel, the infomax principle is based on maximizing the output entropy. Consider a neural network with input $\boldsymbol{x}$ whose output form is $\Phi_i(w_i{}^T x)$ where the Φ_i is a non-linear scalar function and w_i is the weight vector. The infomax principle maximizes the output entropy ,

$$L = H(\Phi_1(w_1{}^T x), \dots, \Phi_n(w_n{}^T x)) \tag{2.12}$$

when the non-linear scalar function of $w_i{}^T x$ is chosen as the cumulative distribution functions matched with the densities f_i , i.e. $\Phi_i(.) = f_i(.)$.

While there exists several algorithms that address the ICA task, one of the most well-known is the FastICA algorithm. The FastICA algorithm is based on the maximization of a contrast function. It finds a direction vector $\boldsymbol{w}$, which maximizes non-Gaussianity with projection $w^T x$. Let the function $g(.)$ be the derivative of a non-quadratic nonlinearity. Valid choices include $g_1(t) = \tanh(a_1 t)$ and $g_2(t) = t * \exp(-t^2/2)$ where the corresponding non-quadratic functions are $G_1(t) = \frac{1}{a_1}\log(\cosh(a_1 t))$ and $G_2(t) = -\exp\left(-\frac{t^2}{2}\right)$ with $1 \le a_1 \le 2$, though generally taken to be 1.

The FastICA algorithm follows this steps [8]:

1. Choose an initial weight vector $\boldsymbol{w}$.
2. Let $w^+ = E\{\boldsymbol{x} g(\boldsymbol{w^T x})\} - E\{g'(\boldsymbol{w^T x})\}\,\boldsymbol{w}$
3. Let $w = \frac{w^+}{||w^+||}$
4. If not converged, go to step 2.

Convergence is achieved when the direction of old and new values of $\boldsymbol{w}$ are same.

2.3. Wavelet Transform

The wavelet transform decomposes a signal onto a set of basis functions that are called wavelets. Wavelets can be used in different applications such as signal analysis [25], image processing [26, 27], and data compression [28]. Wavelets are mathematical functions that decompose time data into different frequency bands while preserving the time information. This provides advantages beside the traditional Fourier Transform. Since the Fourier Transform tends to fail to localize signal discontinuities or spikes in time. Wavelets can be used to decompose a signal at various resolutions; this process is named as *multiresolution signal decomposition* [9]. For instance, an EEG signal can be considered as a superposition of different components at different scales in different times [29].

The wavelet transform W of a signal $f(t)$ is defined as the correlation between the signal and a wavelet function Ψ as

$$W_{a,b} = \int_{-\infty}^{\infty} f(t) \frac{1}{\sqrt{|a|}} \Psi^* \left(\frac{t-b}{a} \right) dt \tag{2.13}$$

where * denotes the complex conjugation, a is the scaling variable and b represent the translation variable of the wavelet. These two variables are associated with the frequency and time localization of the wavelet function.

The expression above can be rewriten by defining the wavelet function as

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \Psi \left(\frac{t-b}{a} \right) \tag{2.14}$$

providing

$$W_{a,b} = \int_{-\infty}^{\infty} f(t) \Psi^*{}_{a,b}(t) dt \tag{2.15}$$

Computing the continuous wavelet transform the signal for every possible a and b values requires more computational effort compared to the discrete wavelet

transformation. In a discrete wavelet transform, the scale parameter a and the translation parameter b are taken at discrete values based on powers of two defined as

$$a_j = 2^j \ , \ \ b_{j,k} = k * 2^j \ \text{ for all } \ j,k \ \in Z \tag{2.16}$$

Then, Equation 2.14 becomes

$$\Psi_{j,k}(t) = 2^{-\frac{j}{2}}\Psi\left(2^{-j}, t-k\right) \ \text{ for all } \ j,k \in Z \tag{2.17}$$

In this configuration the discrete wavelet transform of a signal can be computed very efficiently using conventional filtering and subsampling operations [30].

In the first step; for a given discrete signal s of length N, two sets of coefficients are produced; approximation coefficients cA_1 and detailed coefficients cD_1. These coefficients are obtained by filtering the signal s by a low-pass filter and a high-pass filter, followed by downsampling with a factor 2. In every level of the decomposition, these filtering and downsampling operations are applied repeating to the approximation coefficients of the previous level (see Figure-6). The wavelet decomposition of an EEG signal using a Daubechies wavelet of order 4 is shown in Figure-8.

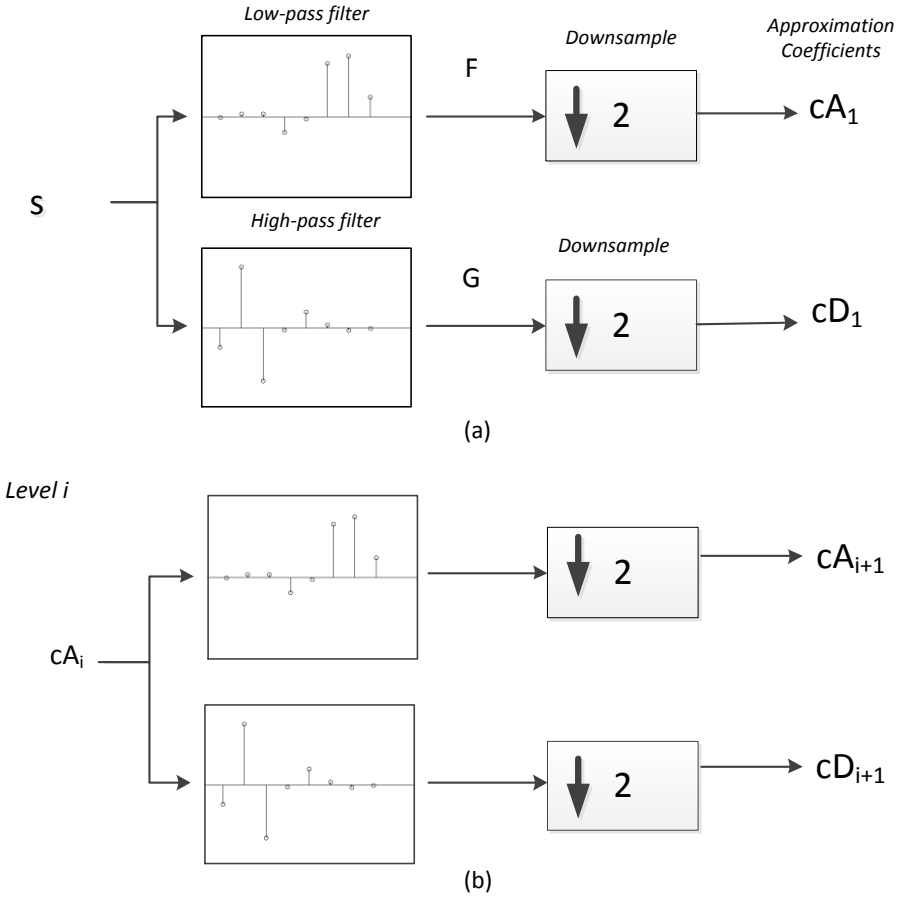

Figure 6. The algorithm of discrete wavelet transform (a) first level of decomposition (b) decomposition at each level

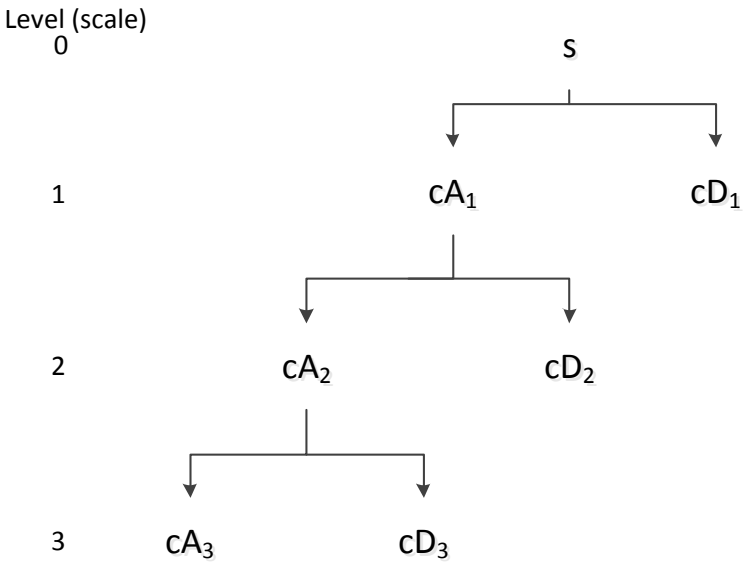

Figure 7. The sub-band coding algorithm of discrete wavelet transform

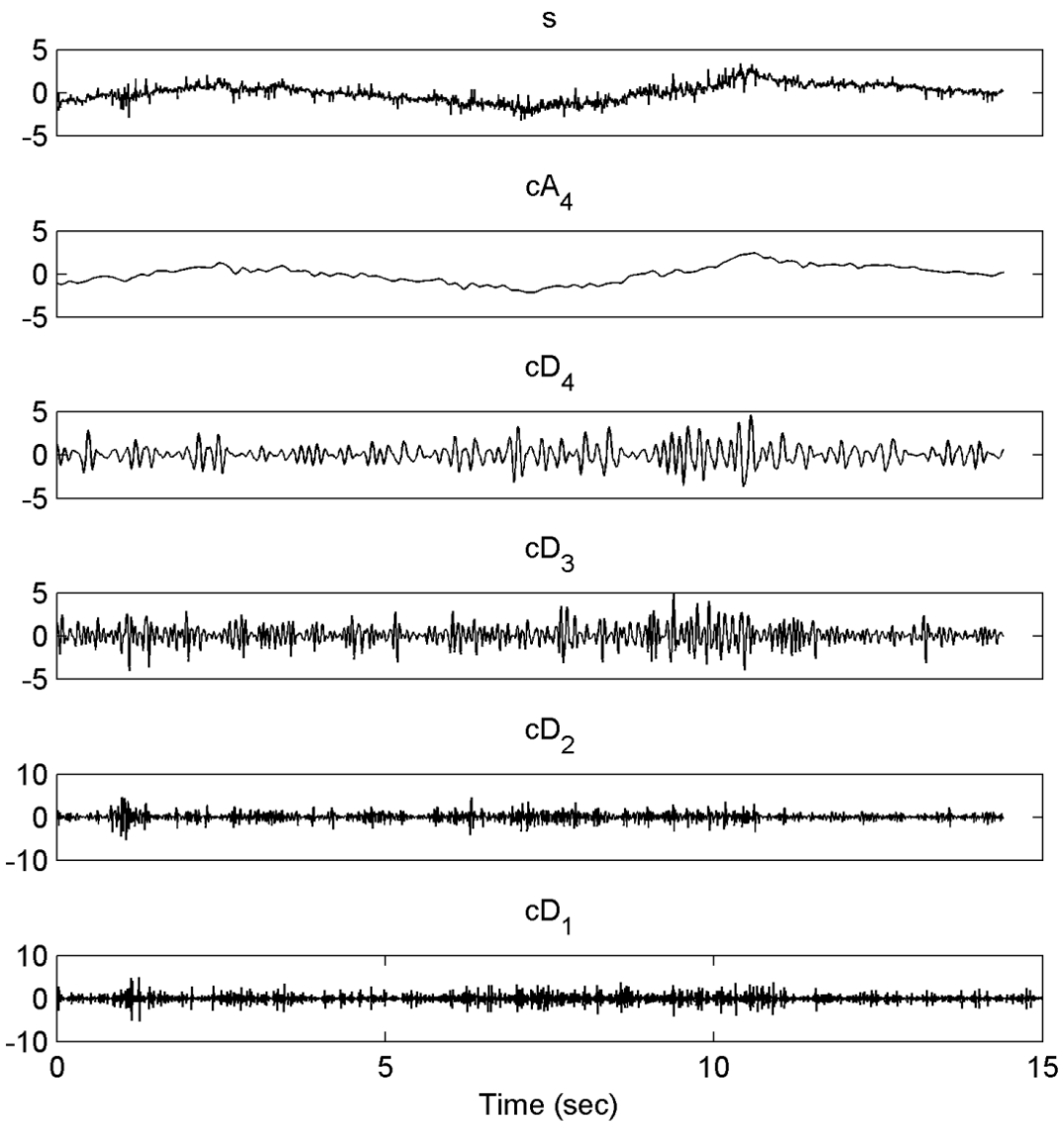

Figure 8. Sub-bands of an EEG channel with 4-tap Daubechies wavelet with 4 level.

CHAPTER 3

QUASI-SUPERVISED LEARNING ALGORITHM

Supervised learning algorithms need a ground truth datasets prepared by an expert containing labels for all samples. This process can be laborious; however, the labels can be varied from expert to expert. The interpretation differences between experts cause the different labeling. The quasi-supervised learning algorithm (QSL) is proposed to avoid this interpretation differences and reduce the workload.

3.1. The Binary Quasi-Supervised Learning Algorithm

Briefly, the quasi-supervised learning algorithm estimates the posterior probabilities of data groups at each sample in the combined dataset [6]. In an abnormality detection application, there are two groups of data; the first one has only normal samples and the other has unlabeled samples. The posterior probabilities of these groups at individual samples can be used as a measure of the resemblance between the samples in the unlabeled dataset and the normal samples. Specifically, the unknown abnormal samples must have low normal dataset posterior probabilities, and using a suitable threshold, abnormal samples can thus be recognized automatically.

3.1.1. Likelihood Ratio Estimation via the Nearest Neighbour Rule

Given a reference set $R = \{x_i, y_i\}$ of points $x_i \in X$ and their respective class labels $y_i \in \{0,1\}$ representing the classes C_0 and C_1, $i = 1,2, \dots, l$, a nearest neighbour classifier is defined by

$$F_R(x) = y_{i_0} \text{with } i_0 = \arg\min_{i=1,2,\dots,l} d(x, x_i) \tag{3.1}$$

for $x \in X$, where $d(.,.)$ denotes the distance metric on X.

Let $\{R_j\}$, $j = 1,2,\dots,M$, be a collection of independent and identically distributed reference sets, consisting of n points from each of the two classes. Define $f_0(x)$ and $f_1(x)$ by

$$f_0(x) = \frac{\sum_{j=1}^{M} \mathbf{1}(F_{R_j}(x)=0)}{M} \quad \text{and} \quad f_1(x) = \frac{\sum_{j=1}^{M} \mathbf{1}(F_{R_j}(x)=1)}{M} \tag{3.2}$$

where $\mathbf{1}(statement)$ is 1 when $statement$ holds and 0 otherwise. It can be shown that for sufficiently large M,

$$f_0(x) \cong \frac{p(x|x\in C_0)}{p(x|x\in C_0)+p(x|x\in C_1)} \quad \text{and} \quad f_1(x) \cong \frac{p(x|x\in C_1)}{p(x|x\in C_0)+p(x|x\in C_1)} \tag{3.3}$$

where $p(x|x \in C_0)$ and $p(x|x \in C_1)$ represent the class conditional probability densities for the classes C_0 and C_1. Note also that by the Bayes rule, $f_0(x)$ and $f_1(x)$ also compute the posterior probabilities $p(C_0|x)$ and $p(C_1|x)$, respectively. Furthermore, the ratio of $f_0(x)$ and $f_1(x)$ provides

$$\frac{f_0(x)}{f_1(x)} = \frac{p(x|x \in C_0)}{p(x|x \in C_1)} \tag{3.4}$$

the likelihood ratio.

In this formulation, the accuracy of the estimated posterior probabilities at a sample depends on M, the number of successive random nearest neighbour classifications. Since the reference set is formed with $2n$ samples with n samples from each class, the total number of distinct reference sets that can be formed is $\binom{l_0}{n}\binom{l_1}{n}$ where l_0 and l_1 denote the number of points in the set $\{x_i\}$ belonging to the respective classes.

3.1.2. Analytical Computation of the Posterior Probabilities

Let $f_0(x)$ represent the fraction of times a point x is assigned to C_0 after a nearest neighbor classification based on all distinct reference sets containing n parts from each class. Consider the distance $d_i = d(x, x_i)$ between x and each sample x_i for

$i = 1,2, \dots, l$ and let $d_{(i)}$ be the ordered sequence of all $\{d_i\}$ with $d_{(1)} \leq d_{(2)} \leq \cdots \leq d_{(l)}$. Similarly let $y_{(i)}$ be the sequence of labels of $x_{(i)}$ producing the $d_{(i)}$. Let also R denote a reference set constructed by n points chosen randomly from each class among the points $\{x_i\}$. Then, $f_0(x)$ coincides with the probabilty $Pr\{y = 0\}$ of assigning x to the class C_0 in a nearest neighbor classification based on the random reference set R,

$$f_0(x) = \Pr\{y = 0\} \tag{3.5}$$

Note that the term $\Pr\{y = 0\}$ can be decomposed conditionally on whether or not the point $x_{(1)}$ is in R as

$$f(x) = \Pr\{x_{(1)} \in R\}\, \mathbf{1}(y_{(1)} = 0) + \Pr\{x_{(1)} \notin R\}\Pr\{y=0|x_{(1)} \in R\} \tag{3.6}$$

Since $\Pr\{y = 0|x_{(1)} \in R\}$ is 1 if $y_{(1)} = 0$, and 0 otherwise. For notational simplicity, we can define E_k as the joint event $x_{(1)}, x_{(2)}, \dots, x_{(k)} \notin R$. Carrying at the same decomposition for $\Pr\{y = 0|E_1\}$ provides

$$\begin{aligned} \Pr\{y = 0|E_1\} = \Pr\{x_{(2)} \in R|E_1\}\, \mathbf{1}(y_{(2)} = 0) + \dots \\ \Pr\{x_{(2)} \notin R|E_1\}\Pr\{y = 0|E_2\} \end{aligned} \tag{3.7}$$

In general,

$$\begin{aligned} \Pr\{y = 0|E_{k-1}\} = \Pr\{x_{(k)} \in R|E_{k-1}\}\, \mathbf{1}(y_{(k)} = 0) + \dots \\ \Pr\{x_{(k)} \notin R|E_{k-1}\}\Pr\{y = 0|E_k\} \end{aligned} \tag{3.8}$$

Furthermore, since R must have at least $2n$ data points, n from each class, the decomposition does not need to be carried out beyond some k_* given by

$$k_* = \max\{k| \sum_{k'=k}^{l} \mathbf{1}(y(k') = 0) \geq n \ and \ \sum_{k'=k}^{l} \mathbf{1}(y(k') = 1) \geq n\} \tag{3.9}$$

as $\Pr\{x_{(k_*)} \in R|E_{k^*-1}\} = 1$ and $\Pr\{x_{(k_*)} \notin R|E_{k_*-1}\} = 0$.

The following algorithm elucidates the computation of $f_L(x)$ when $L \in \{0,1\}$ based on $\{x_i, y_i\}$ where $i = 1,2, \dots, l$.

- Compute $d_i = d(x, x_i)$
- Sort $\{d_i\}, d_{(1)} \leq d_{(2)} \leq \cdots \leq d_{(l)}$ and determine the $\{x_i\}$ and $\{y_i\}$ according to the sorted distances.
- Determine k_*, and set $\Pr\{y = L | E_{k_*-1}\} = \mathbf{1}(y_{(k_*)} = L)$
- Calculate $\Pr\{y = L | E_{k_*}\} = \Pr\{x_{(k+1)} \in R | E_{k_*}\} \mathbf{1}(y_{(k+1)} = L) + \Pr\{x_{(k+1)} \notin R | E_k\} \Pr\{y = L | E_{k+1}\}$ for $k = k_* - 1, k_* - 2, \dots, 1$
- Finally compute

  $f_L(x) = \Pr\{x_{(1)} \in R\} \mathbf{1}(y_{(1)} = L) + \dots + \Pr\{x_{(1)} \notin R\} \Pr\{y = L | E_1\}$

Note that $\Pr\{x_{(k)} \in R | E_{k-1}\}$ can be calculated by

$$\Pr\{x_{(k)} \in R | E_{k-1}\} = 1 - \Pr\{x_{(k)} \notin R | E_{k-1}\} = 1 - \frac{\binom{{l_0}^{k+1}}{n}\binom{{l_1}^{k+1}}{n}}{\binom{{l_0}^{k}}{n}\binom{{l_1}^{k}}{n}} \tag{3.10}$$

where ${l_0}^k$ represents the number of points that belong to C_o in the set $\{x_{(k)}, x_{(k+1)}, \dots, x_{(l)}\}$ and ${l_1}^k$ denotes the number of points in the same set that belong to C_1 .

$${l_0}^k = \sum_{i=k}^{l} \mathbf{1}(y_{(i)} = 0) \text{ and } {l_1}^k = \sum_{i=k}^{l} \mathbf{1}(y_{(i)} = 1) \tag{3.11}$$

Clearly when $y_{(k)} = 0$, ${l_0}^{k+1} = {l_0}^k - 1$ and ${l_1}^{k+1} = {l_1}^k$; and similarly when $y_{(k)} = 1$, ${l_0}^{k+1} = {l_0}^k$ and ${l_1}^{k+1} = {l_1}^k - 1$. This provides

$$\Pr\{x_{(k)} \in R | E_{k-1}\} = \begin{cases} \dfrac{n}{{l_0}^k} & if \quad y_{(k)} = 0 \\ \dfrac{n}{{l_1}^k} & if \quad y_{(k)} = 1 \end{cases} \tag{3.12}$$

Finally note that the posterior probabilities can be calculated for any x_i by carrying out the algorithm above for $x = x_i$ based on collection of parts $(C_0 \cup C_1)\{x_i\}$.

3.1.3. Class Overlap Measures

In abnormality detection and classification problems, class-overlap measures provide information about the separability of the classes. Good classification, in particular, can be constructed with small-class overlaps.

There are several measures of class overlap that can be computed using the estimated posterior probabilities. One of them is $M_{LLR}(x)$ that computes the log-likelihood ratio of two classes at a point $x \in X$, given by

$$M_{LLR}(x) = \log\left(\frac{f_0(x)}{f_1(x)}\right) \tag{3.13}$$

for all x with $f_0(x) \neq 0$ and $f_1(x) \neq 0$. When x is in the region where the two classes overlap the ratio$f_0(x)/f_1(x)$ goes to one and $M_{LLR}(x)$ goes to zero.

Another way of measuring the class-overlap or similarity between the underlying distributions is using an affinity measure. One of the affinity measures between distributions is the Henze-Penroze affinity that computes the integral,

$$\int_x \frac{2p_1(x)p_2(x)}{p_1(x) + p_2(x)}\,dx \tag{3.14}$$

for any given probability distributions $p_1(x)$ and $p_2(x)$. The integral goes to one when $p_1(x) = p_2(x)$ for all x. Based on Equation 3.14 the measure $M_{HP-like}(x)$ can be formulated as

$$M_{HP-like}(x) = f_0(x)f_1(x) \cong \frac{p(x|x \in C_0)p(x|x \in C_1)}{(p(x|x \in C_0) + p(x|x \in C_1))^2} \tag{3.15}$$

When $p(x|x \in C_0) \cong p(x|x \in C_1)$, $f_0(x) \cong f_1(x) \cong 1/2$ and $M_{HP-like}(x)$ approaches. When x clearly belongs to one of C_0 or C_1, then $M_{HP-like}(x)$ goes to zero.

Finally; M_{Diff} can be defined as another measure of the class overlap. On the differences between distributions using $f_0(x)$ and $f_1(x)$ by

$$M_{Diff}(x) = f_0(x) - f_1(x) \cong \frac{p(x|x \in C_0) - p(x|x \in C_1)}{p(x|x \in C_0) + p(x|x \in C_1)} \tag{3.16}$$

Note that this measure is similar to $M_{LLR}(x)$ at the region where the two distributions overlap, $M_{Diff}(x)$ goes to zero. It's advantage against to M_{LLR} is that it can be computed for all x even when $f_0(x)$ or $f_1(x)$ is equal to zero.

3.1.4. Selection of the Optimal Reference Set Size n

The accuracy of the algorithm that estimates the posterior probabilities of the two classes at individual samples is related to the reference set size. Recall that n denotes the number of samples from each class in the reference sets that drive the nearest neighbor classifications. Ideally; the best n provides the minimum class overlap. So, M_{LLR} and M_{Diff} should never be around zero and $M_{HP-like}$ must be zero for all x_i. In addition n must be chosen as small as possible to avoid too flexible nearest neighbor classifications.

In order to choose the optimal reference set size under these considerations, the following cost function must be minimized:

$$E(n) = 4\sum_i M_{HP-like}(x_i) + 2n \tag{3.17}$$

The reason behind the first term is related with the aim of the quasi-supervised learning algorithm to minimize the class overlap. The second term aims to reduce reference set size for a better generalization ability. The scaling factor 4 in front of the first term balances two scenarios; First, when the classes overlap entirely,

$$4\sum_i M_{HP-like}(x_i) = l$$

In turn, when all samples are included in the reference set, $2n = l$.

3.2. M-ary Quasi-Supervised Learning Algorithm

Supervised learning algorithms aim at producing a learning model for classification from a labeled dataset. Several successful techniques are applied to binary classification, but in machine learning, multiclass classification is the more complex and special case of classification with a choice of several class labels for each sample. In the literature, most multiclass supervised learning algorithms are configured as multiple "one against the others" classifications.

The motivation behind the quasi-supervised learning algorithm is abnormality detection by estimating posterior probabilities. Abnormality detection can be thought of as a binary problem because the task is to determine whether a sample belongs to the normal class or not. But the quasi-supervised learning algorithm can be adopted for a multi-class labeling problem easily thanks to the learning algorithm's generality that derives from nearest neighbour classification.

3.2.1. Methodology of M-ary Quasi-Supervised Learning Algorithm

Suppose we have M classes $C_1, C_2, \dots, C_M$, represented in the collection $\{x_i, y_i\}$ with $y_i \in \{1,2, \dots, l\}$, $i = 1,2, \dots, M$. Suppose that a random reference set R has n samples from each class and allows nearest neighbor classification of a sample x into one of $C_1, C_2, \dots, C_M$. Let the functions $f_1(x), f_2(x), \dots, f_n(x)$ represents the probability of assigning x to the possible classes by

$$\begin{aligned} f_1\{x\} &= \Pr\{y = 1\} \\ &\vdots \\ f_M\{x\} &= \Pr\{y = M\} \end{aligned} \tag{3.18}$$

As before, consider the sorted distances $d(i)$ for a point $x_{(i)}$ and their class labels are $y_{(i)}$. These probability represented by $f_i(x)$ can be decomposed conditionally on whether or not the point $x_{(1)}$ is in R, providing

$$f_1(x) = \Pr\{x_{(1)} \in R\}\, \mathbf{1}(y_{(1)} = 1) + \Pr\{x_{(1)} \notin R\}\Pr\{y=1|x_{(2)} \in R\} \tag{3.19}$$

Again, for notational simplicity, define E_k as the joint event $x_{(1)}, x_{(2)}, \dots, x_{(k)} \notin R$. Carrying the same decomposition strategy further to $\Pr\{y = 1|E_1\}$ provides

$$\begin{aligned}\Pr\{y = 1|E_1\} = \Pr\{x_{(2)} \in R|E_1\}\,1(y_{(2)} = 1) + \dots \\ \Pr\{x_{(2)} \notin R|E_1\}\Pr\{y = 1|E_2\}\end{aligned} \tag{3.20}$$

Since R must have at least n data points from each class, the decomposition does not need to be carried out beyond a k_* given by

$$\begin{aligned}k_* = \max\{k| \sum_{k'=k}^{l} \mathbf{1}(y(k') = 1) \geq n \ and \ \sum_{k'=k}^{l} \mathbf{1}(y(k') = 2) \geq n \ , \\ \cdots, \sum_{k'=k}^{l} \mathbf{1}(y(k') = M) \geq n\}\end{aligned} \tag{3.21}$$

Finally, the following modified algorithm computes the probabilities $f_L(x)$ for $L \in 1,2,\dots,M$ for a given x based on the dataset $\{x_i, y_i\}$ where $i = 1,2,\dots,l$, and for a fixed n.

- Compute $d_i = d(x, x_i)$
- Sort $\{d_i\}, d_{(1)} \leq d_{(2)} \leq \cdots \leq d_{(l)}$ and determine the corresponding sequences $\{x_{(i)}\}$ and $\{y_{(i)}\}$
- Identify the k_*, and set $\Pr\{y = L|E_{k_*-1}\} = \mathbf{1}(y_{(k_*)} = L)$
- Calculate $\Pr\{y = L|E_{k_*}\} = \Pr\{x_{(k+1)} \in R|E_k\}\,\mathbf{1}(y_{(k+1)} = L) + \Pr\{x_{(k+1)} \notin R|E_k\}\Pr\{y = L|E_{k+1}\}$ for $k = k_* - 1, k_* - 2, \dots, 1$
- Identify $f_L(x) = \Pr\{x_{(1)} \in R\}\,\mathbf{1}(y_{(1)} = L) + \Pr\{x_{(1)} \notin R\}\Pr\{y = L|E_1\}$

In a general form $\Pr\{x_{(k)} \in R\,|E_{k-1}\}$ can be calculated by

$$\begin{aligned}\Pr\{x_{(k)} \in R\,|E_{k-1}\} = 1 - \Pr\{x_{(k)} \notin R|E_{k-1}\} \\ = 1 - \frac{\binom{l_1^{\,k+1}}{n}\binom{l_2^{\,k+1}}{n}\dots\binom{l_M^{\,k+1}}{n}}{\binom{l_1^{\,k}}{n}\binom{l_2^{\,k}}{n}\dots\binom{l_M^{\,k}}{n}}\end{aligned} \tag{3.22}$$

where ${l_L}^k$ indicates the number of C_L points in the set $x_{(k)}, x_{(k+1)}, \dots, x_{(l)}$ for $L = 1,2, \dots, M$,

$$ {l_L}^k = \sum_{i=k}^{l} \mathbf{1}(y_{(i)} = L) \tag{3.23} $$

Since when $y_{(k)} = L$, ${l_L}^{k+1} = {l_L}^k - 1$ and ${l_i}^{k+1} = {l_i}^k$ for $i \neq L$, we obtain

$$ \Pr\{ x_{(k)} \in R \,|E_{k-1}\} = \begin{cases} \dfrac{n}{{l_1}^k} & if \quad y_{(k)} = 1 \\ \vdots & \\ \dfrac{n}{{l_M}^k} & if \quad y_{(k)} = M \end{cases} \tag{3.24} $$

which can be used in the 4^{th} step of the algorithm.

3.2.2. Evaluating the M-ary Class Overlap

A class overlap measure is important for M-ary quasi-supervised learning algorithm as well, since the classes are desired to be as separable as possible. A generalization of the log-likelihood ratio to measure the class overlap for M-ary quasi-supervised learning ${M_L}^{LLR}$ can be defined by

$$ {M_L}^{LLR}(x) = log\left(\frac{f_L(x)}{\sum_{i \neq L}^{M} f_i(x)}\right) \tag{3.25} $$

for each class label L. Clearly, when $f_L(x) \geq \frac{1}{2}$; ${M_L}^{LLR}$ is positive, and otherwise it is negative. An overlap measure that is independent of a class label can be defined, however, based on the Shannon entropy on the posterior probability computed by $f_1(x), f_2(x), \dots, f_M(x)$. Thus, for a point x, we define $M^{ent}(x)$ by

$$ M^{ent}(x) = -\frac{1}{log_2 M} \sum_{j=1}^{M} f_j(x) log_2 f_j(x) \tag{3.26} $$

3.2.3. Selection of the Optimal Reference Set Size n

As before, the accuracy of the algorithm above is related with the reference set size. In a multiclass problem, a new cost function must be defined on the entropy-based class overlap measure. Such a cost function can be defined as :

$$E(n) = \frac{1}{log_2 M} \sum_i (-f_i * log f_i) + M * n \tag{3.27}$$

Again, this cost function must be minimized to determine the optimum reference set size. First term of the cost function penalizes the overlap between the classes, and the second term penalizes large n.

CHAPTER 4

THE EEG DATASET

The dataset that was used in this thesis was originally used in a brain-computer interface study for disabled subjects [31]. In the EEG experiment, the subjects were facing a laptop screen on which six images were flashed successively in a random order (see in Figure 9). The images were that of a television, a telephone, a lamp, a door, a window and a radio. The images were selected according to an application scenario aiming to allow users control electrical appliances via a brain-computer interface system.

During the experiment, the images were flashed randomly and with only one image at a time. The duration of an image flash was 100ms and for the following 300ms, the screen was blank.

The dataset was recorded at 2048 Hz sampling frequency from a total of 32 electrodes. The electrodes were placed according to the 10-20 international electrode placement protocol shown in Figure 4.

The dataset was provided for download in the MATLAB format, and obtained from the web page of Multi Media Signal Processing Group of École Polytechnique Fédéralede Lausanne [32].

The EEG data was recorded on five disabled and four healthy subjects in total. In this thesis, data from one of the healthy subjects was used. All subjects were Ph.D. students at the laboratory where the experiment was carried out, were males and with on average age of 30 years $\pm$ 2.3 years.

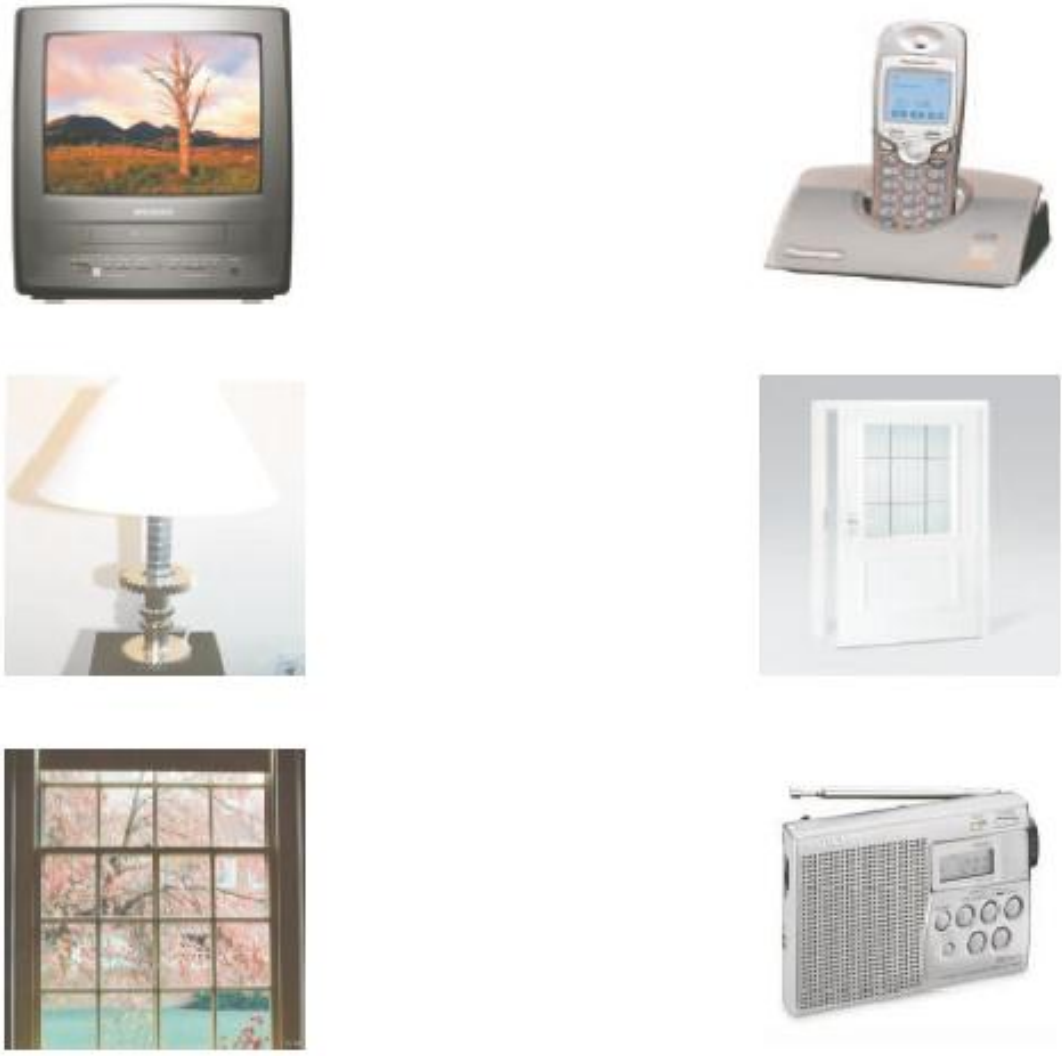

Figure 9. The images that were flashed in the experiment, representing the different visual stimuli

CHAPTER 5

EXPERIMENT SETUP AND RESULTS

In this chapter; we describe the experiment setup and the results obtained in the experiments. First, we detail the pre-processing operations carried out on the EEG data as preparation for the subsequent analyses. Then we elaborate on the results obtained on the data in terms of various comparison scenarios. The EEG recording which we choose was taken from normal subject with 1 minute duration. It has 10 seconds baseline between 15^{th} and 25^{th} seconds.

5.1. EEG Data Pre-processing

As a first step, we have resampled the EEG data from 2048Hz down to 128Hz in order to reduce the computational load. Next, we have removed the affine component by fitting a first order polynomial to all EEG channels to eliminate the effects of streaming time from the analysis. Finally, each channel normalized to unit variance.

5.2. Construction of EEG Profiles

After the affine fitting and the normalization, to increase the separation performance of the quasi-supervised learning EEG profiles were constructed using independent component analysis and the wavelet decomposition. Independent component analysis was applied to data to obtain independent signal constitutions. To derive multi-scale signal components the wavelet decomposition was used that described in Figure 8. Also to get more distinct EEG profiles, independent component analysis and the wavelet decomposition were used together to obtain multi-scale signal components that were statistically independent from each other.

The feature construction is explained in the Figure 11.

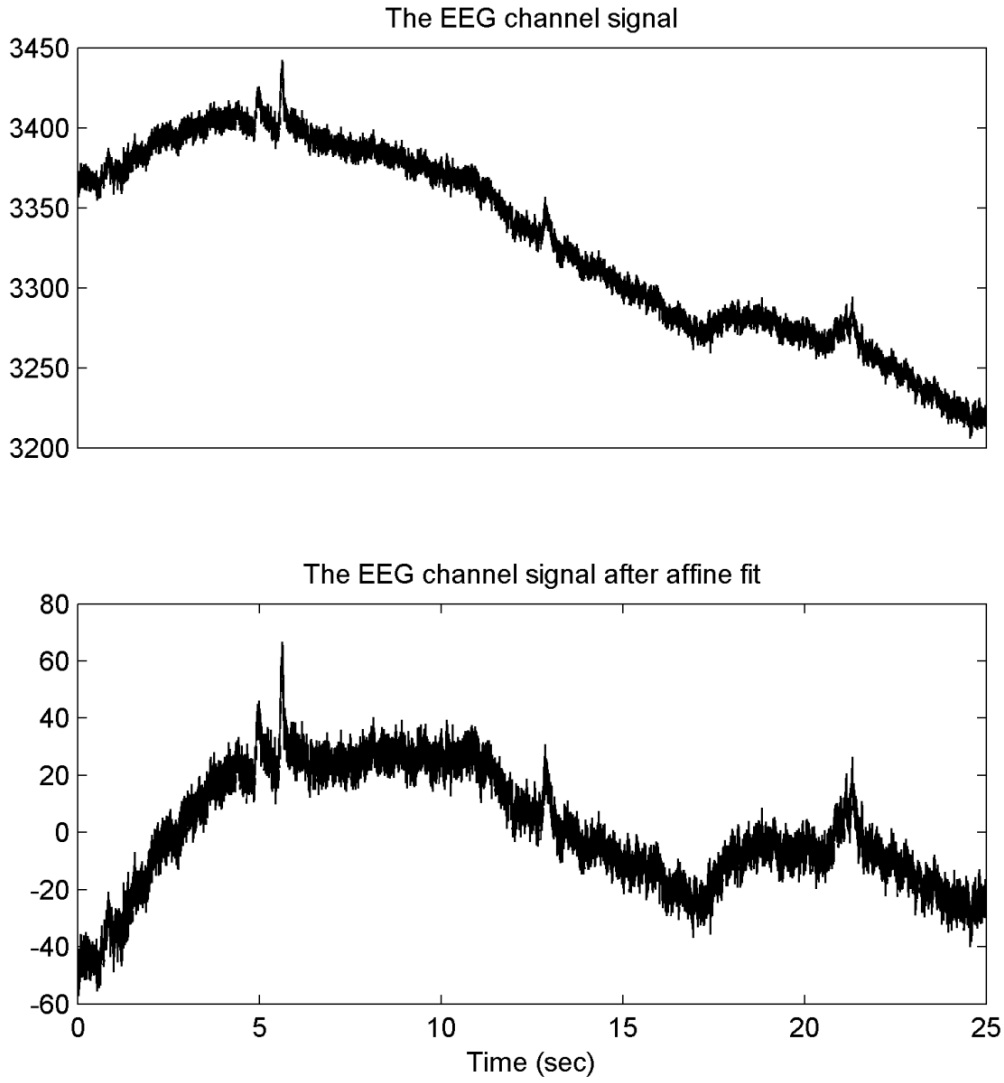

Figure 10. The original EEG channel signal and the signal after affine component removed

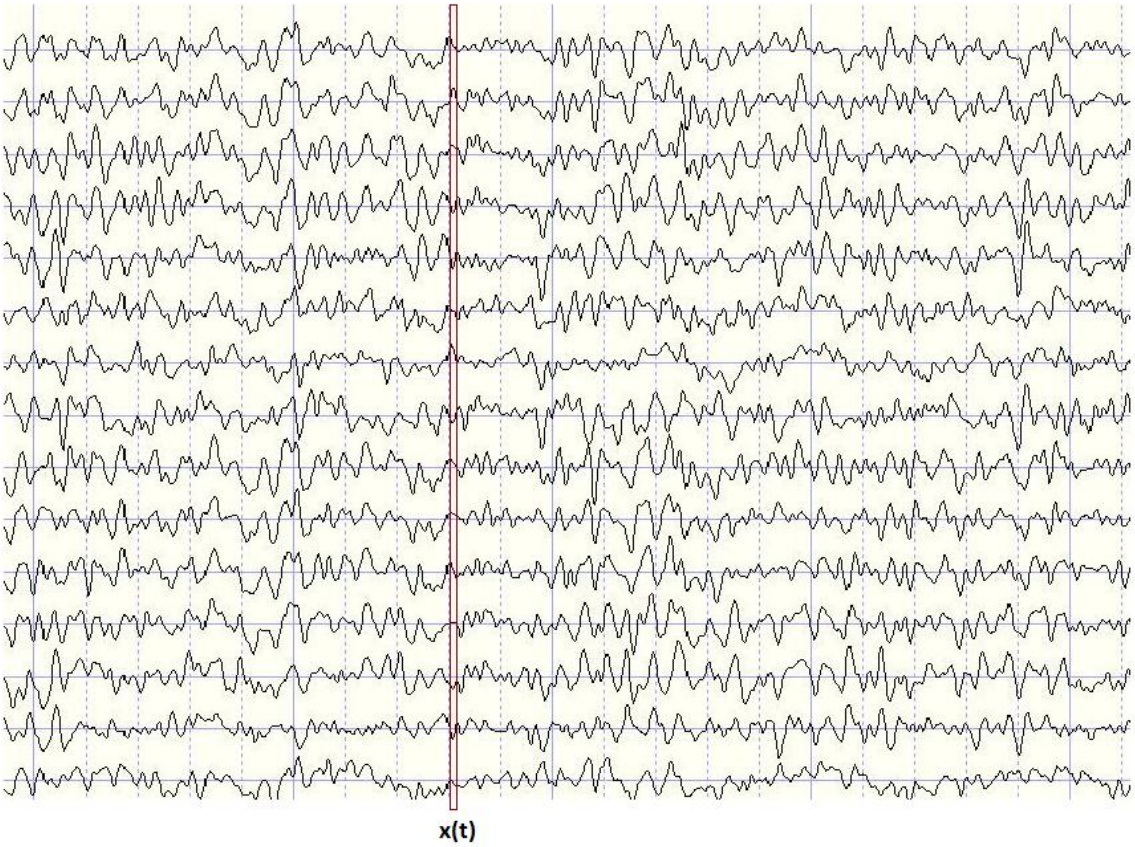

Figure 11. Feature construction of the EEG data

5.3. Visual Stimuli vs Baseline

In this part, we compare the effects of the visual stimuli on the data. To this end, we carried out a binary comparison between the first 15 seconds that contained visual stimuli and the subsequent 10 seconds of EEG recording against a blank screen that constitutes the baseline data. For this 25 second data, we applied the binary quasi-supervised learning algorithm on the EEG profiles obtained from the pre-processed data along with those constructed following the independent component analysis, the wavelet decomposition, and the independent component analysis on the wavelet decomposition signals.

In these experiments, the visual stimuli area (0-15 second) was taken as the first class and the baseline area (15-25 second) was taken as the second class. Note that in a perfect separation case the estimated posterior probability of a sample is one when its label is same with the posterior probability class and zero otherwise. But brain has a complex structure and it does not behave according to the ideal scheme. By using quasi-

supervised learning algorithm toolbox, we obtained the differences and abnormalities between samples (see Figure 12).

We observed some abnormality around the fifth second and thirteenth second. The posterior probability curve made some raises unexpectedly. This can be caused by many reason such as the subject could think different objects from that image, he could have different thoughts. Also this abnormality can be caused by the equipment or the conditions.

As seen from the Figure 12, passing from the visual stimuli area to baseline area is not restrict. The stimuli effect continues in the beginning of the baseline region with decreasing. This transition seen in the Figure 13 more detailed.

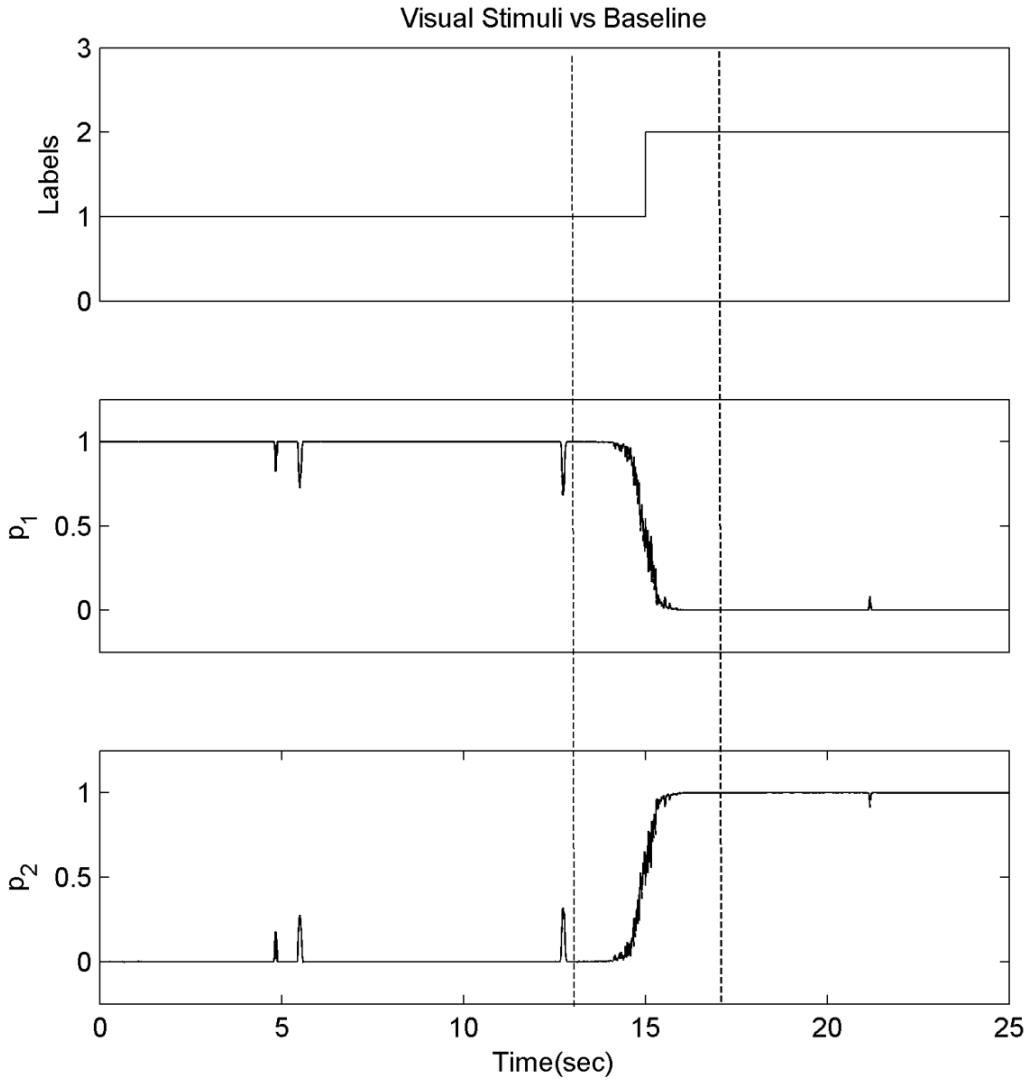

Figure 12. The class labels and the estimated posterior probabilities of a time series EEG data.

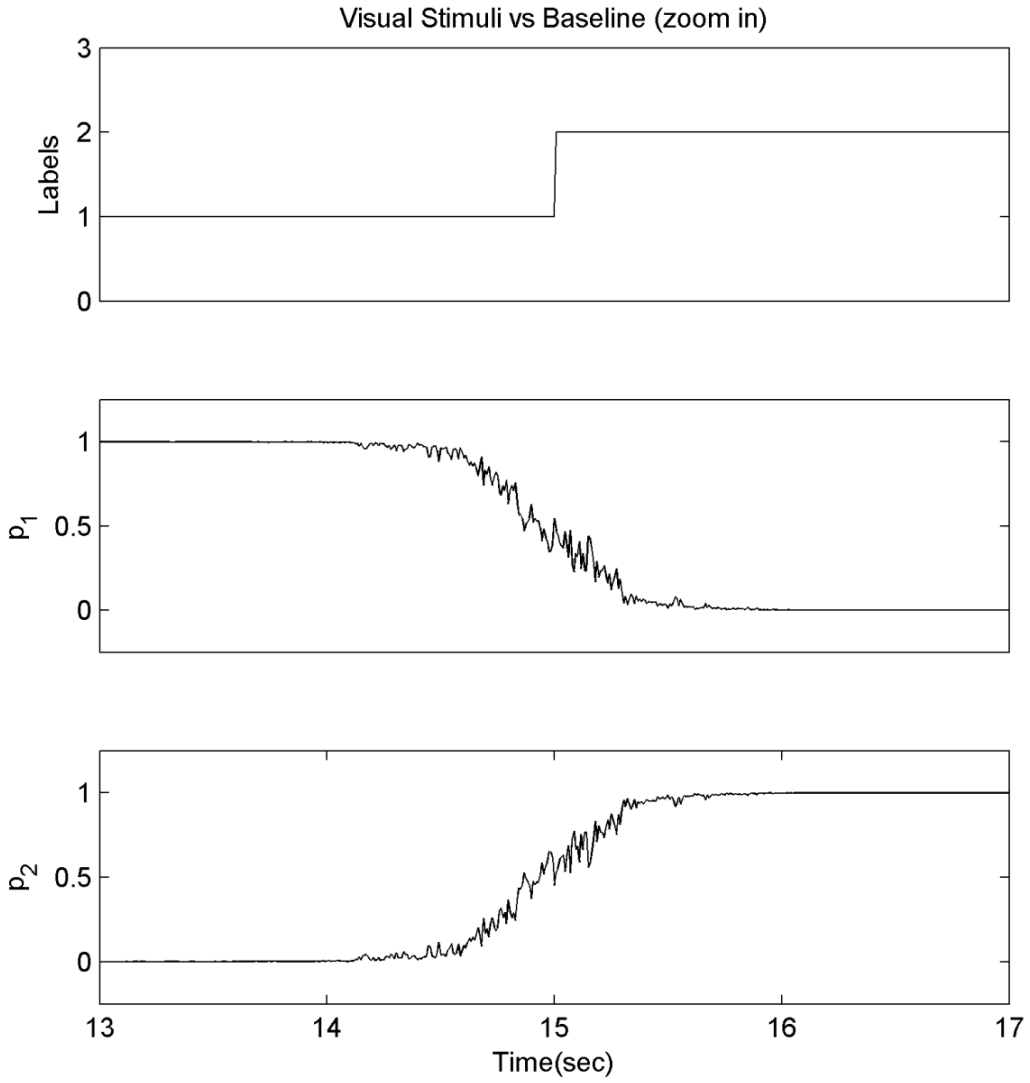

Figure 13. The zoomed in transition region between stimuli area and baseline area.

The independent component analysis extracts the independent components from data and makes features more explicit. To increase the separation performance in our dataset before applying quasi-supervised learning algorithm, we applied independent component analysis to the EEG profiles and reduced the feature number and organized the data in independent as possible from each other.

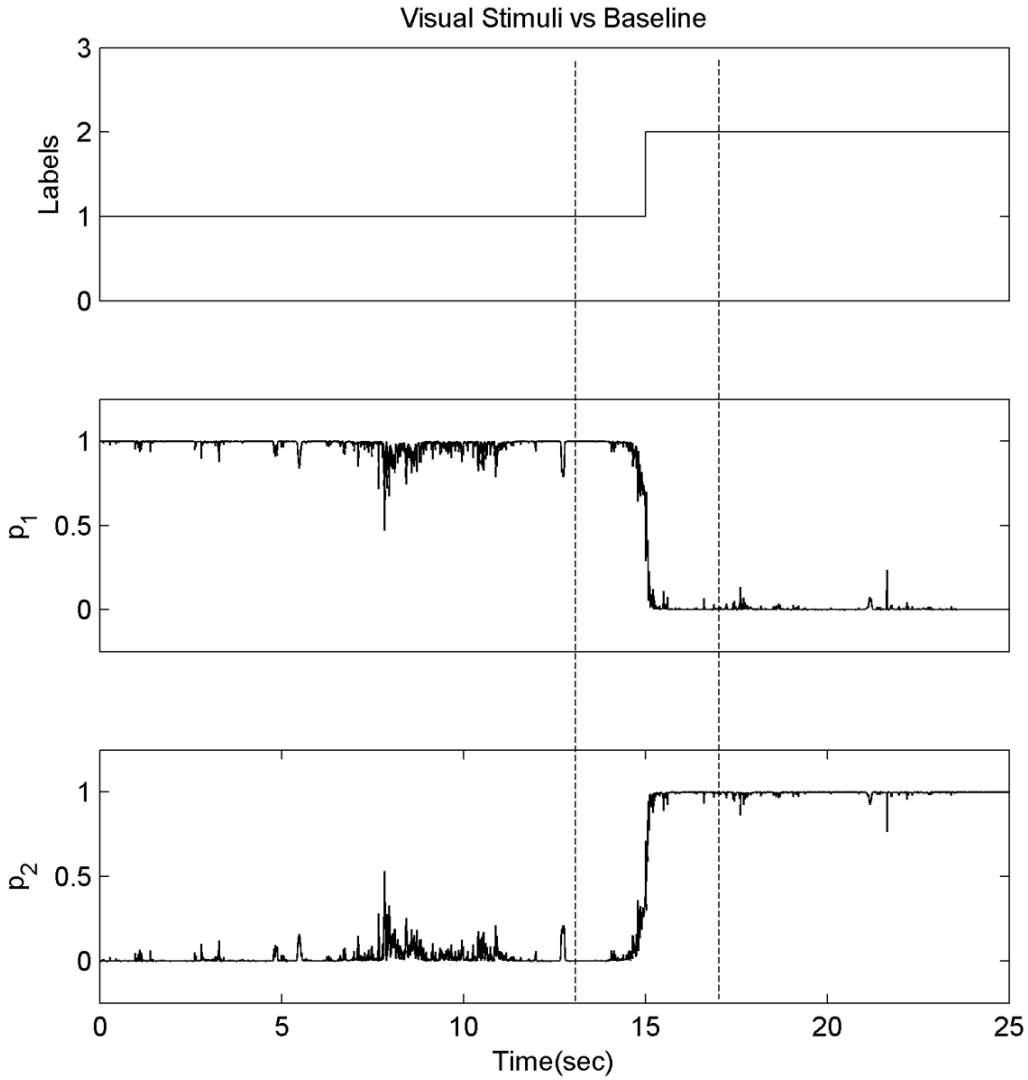

Figure 14. The posterior probability curves of the time series EEG data after independent components analysis was applied.

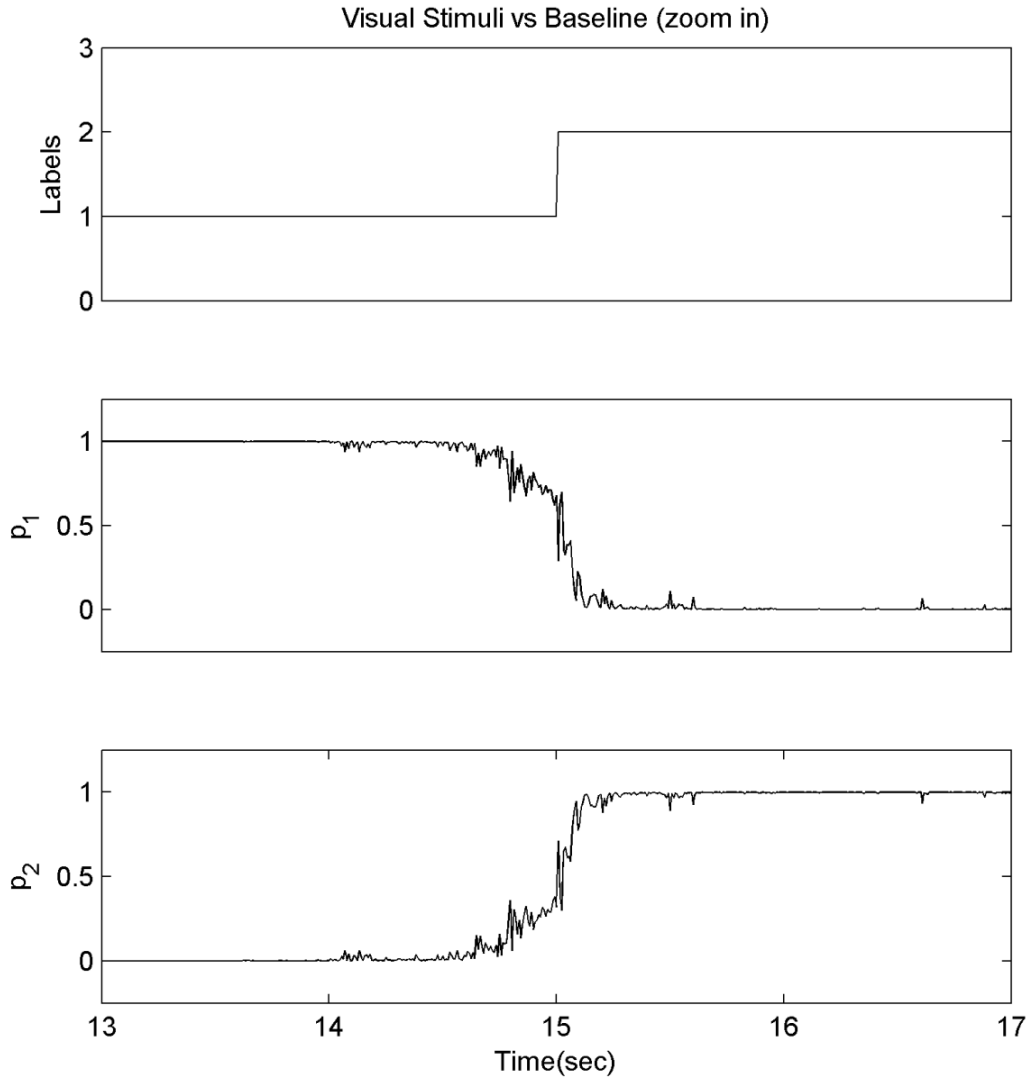

Figure 15. Transition region between stimuli area and baseline area after independent component analysis was applied.

As seen in the previous two sections, the quasi-supervised learning algorithm is successful in finding differences between samples observed under varying stimulus conditions, different feature extraction methods help to see these differences more clearly. Indeed, the quasi-supervised learning algorithm applied to the wavelet decomposition-based EEG profiles produced smoother posterior probability curves, indicating better separation of the two conditions in classification.

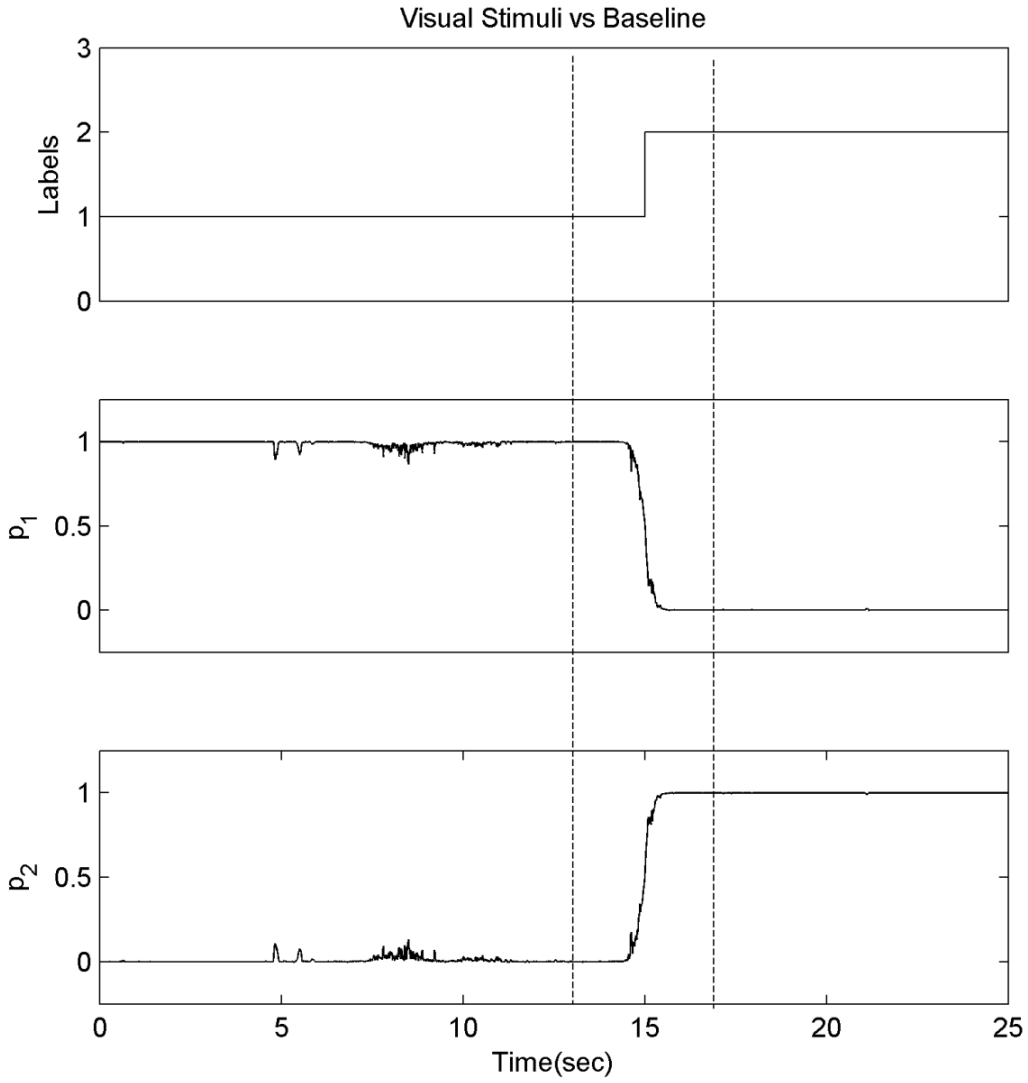

Figure 16. The posterior probability curves of the time series EEG data after wavelet decomposition was applied.

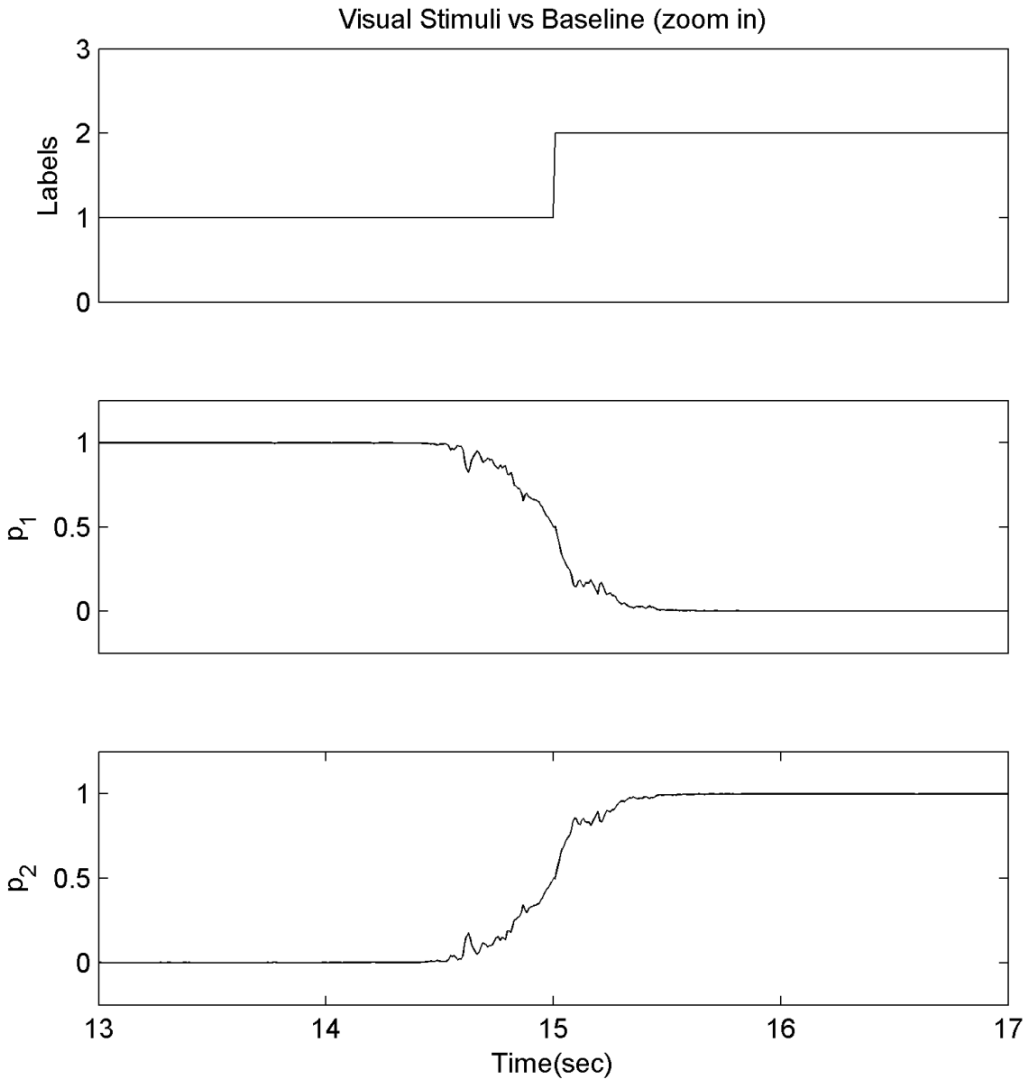

Figure 17. Transition region between stimuli area and baseline area after wavelet decomposition.

Using the wavelet decomposition-based EEG profiles, also increased dimensionality of the learning problem as well as computational cost because of adding many new features. The use of independent component analysis in this step is a feature reduction tool that projects the data into a lower dimensional space without sacrificing the data.

The posterior probabilities curves obtained by this strategy indicate an even sharper transition between the conditions, arguably due to the lower dimensionality.

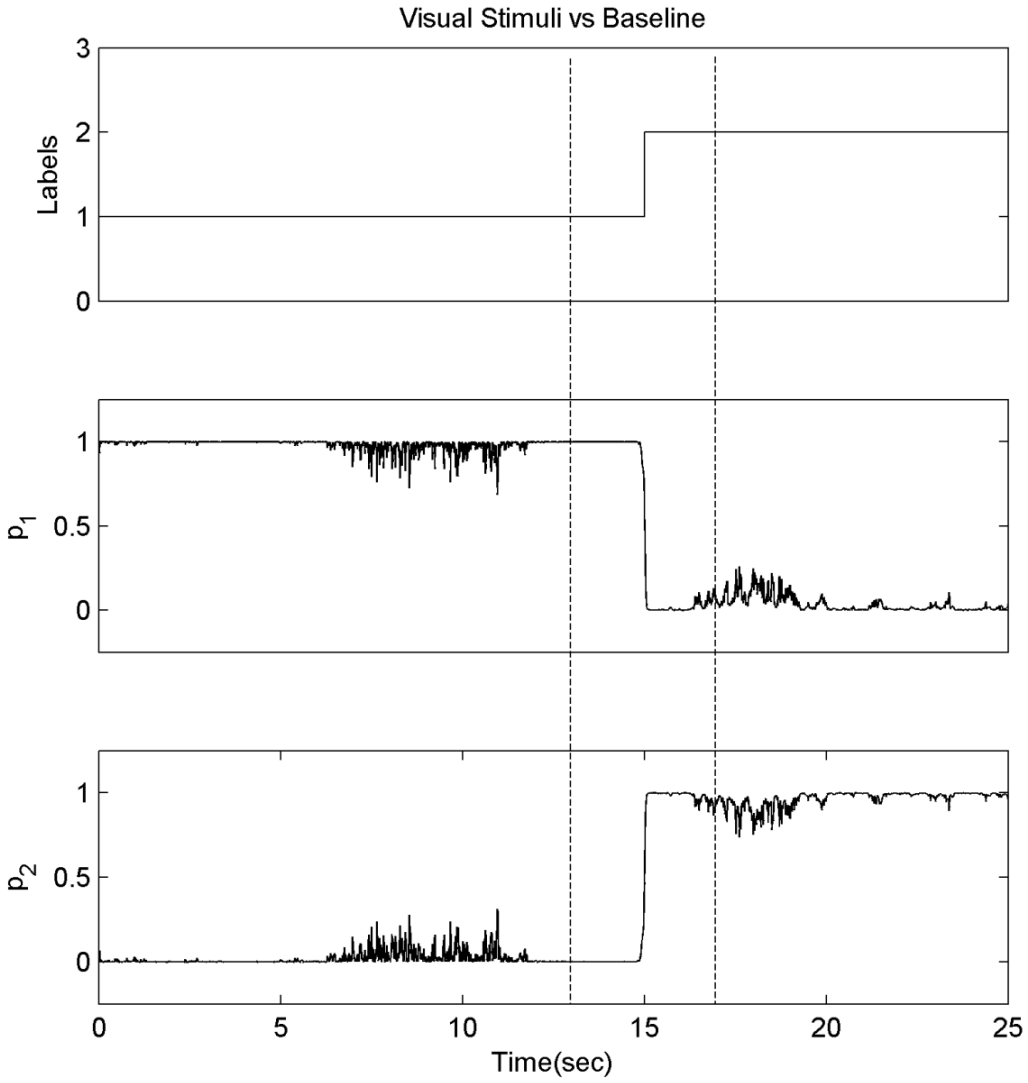

Figure 18. The posterior probability curves of the time series EEG data after wavelet decomposition was applied to independent component identified profiles.

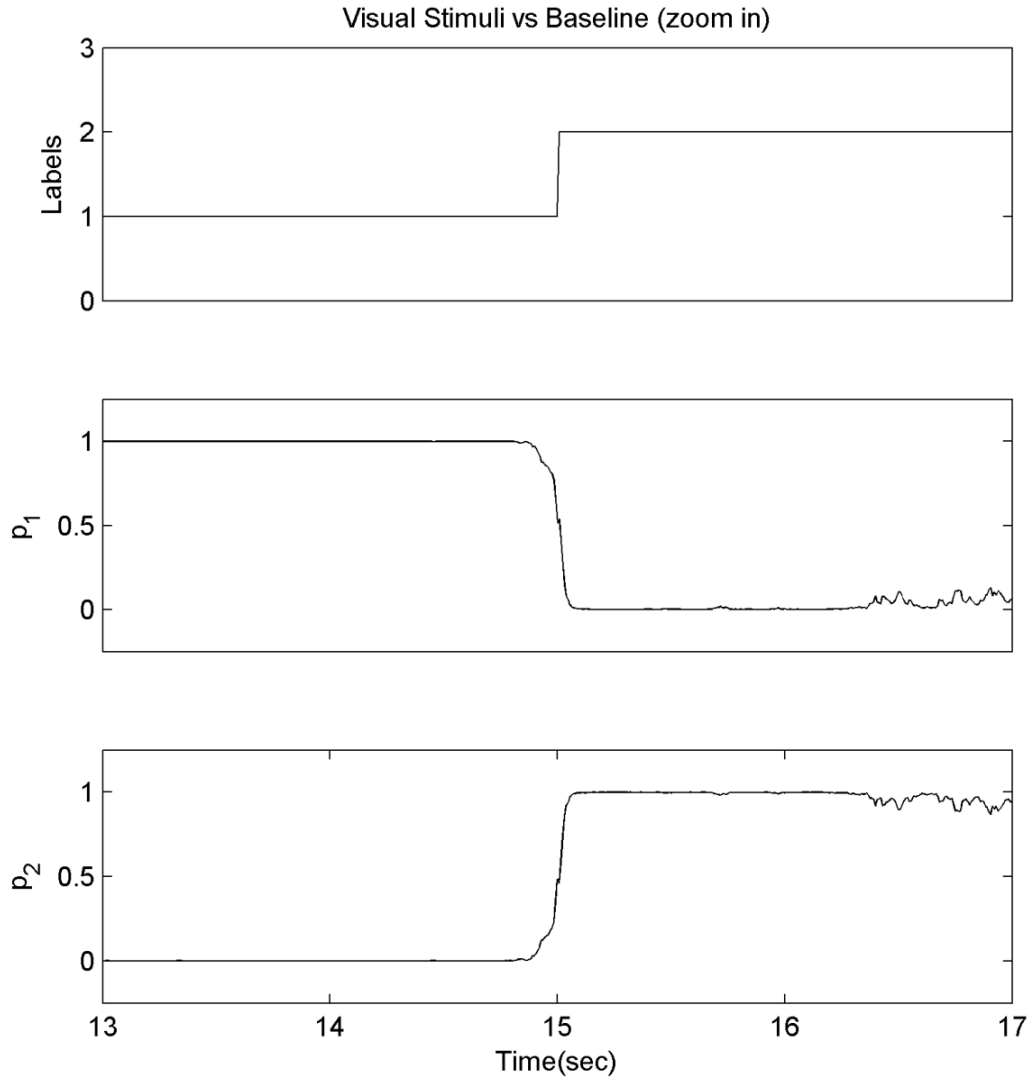

Figure 19. Transition region between stimuli area and baseline area after wavelet decomposition was applied to independent component identified profiles.

5.4 Visual Stimuli vs Blank Screen

In this section, we investigated the differences between the different visual stimulus samples and samples observed during the blank screen. The different visual stimulus areas were labeled with the number of the flashed image. The blank screen areas between the images were labeled as the seventh stimulus. After labeling among the quasi-supervised learning algorithm was actual on the varying EEG profiles.

The optimum reference set size n is chosen by line search with the cost minimization. The optimum reference set size n was determined to be 35 on the assisted EEG profiles (Fig-20). The posterior probabilities of each image class label were estimated using this reference set size (Fig-21). To visualize of the resulting separation, we constructed a confusion matrix and assigned an image label to all samples using maximum a posteriori rule. According to the confusion matrix in Table 2, the separation performance was %31.82. Note that while a confusion matrix can be a performance measure for a classification algorithm, that we only used this representation to make a comparison between the methods that we use in this thesis. Also, the quasi-supervised learning algorithm provides another class overlap measures in the log-likelihood ratio that rises when the samples are distinct. (Fig-22)

Table 3. Confusion Matrix of Visual Stimuli vs Blank Screen Using Quasi-Supervised Learning Algorithm

		Estimated Labels							Total
		1	2	3	4	5	6	7	
Real Labels	1	77	0	0	0	0	0	0	77
	2	0	91	0	0	0	0	0	91
	3	0	0	77	0	0	0	0	77
	4	0	0	0	63	0	0	0	63
	5	0	0	0	0	91	0	0	91
	6	0	0	0	0	0	90	0	90
	7	194	202	215	177	252	269	122	1431
Total		176	293	292	240	343	359	122	1920

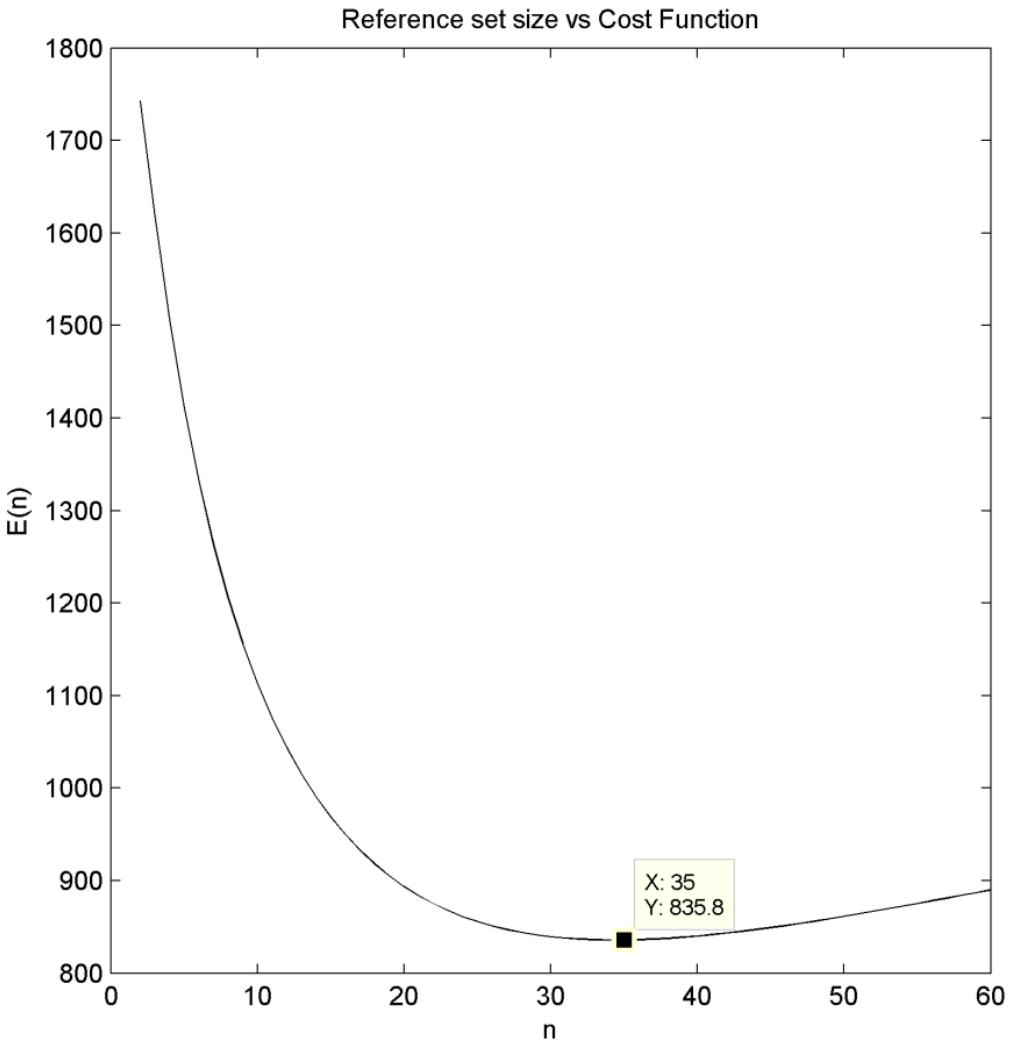

Figure 20. The plot of $\boldsymbol{E(n)}$ for n =1,...,60. The functional $\boldsymbol{E(n)}$ attained its minimum value at $\boldsymbol{n = 35}$.

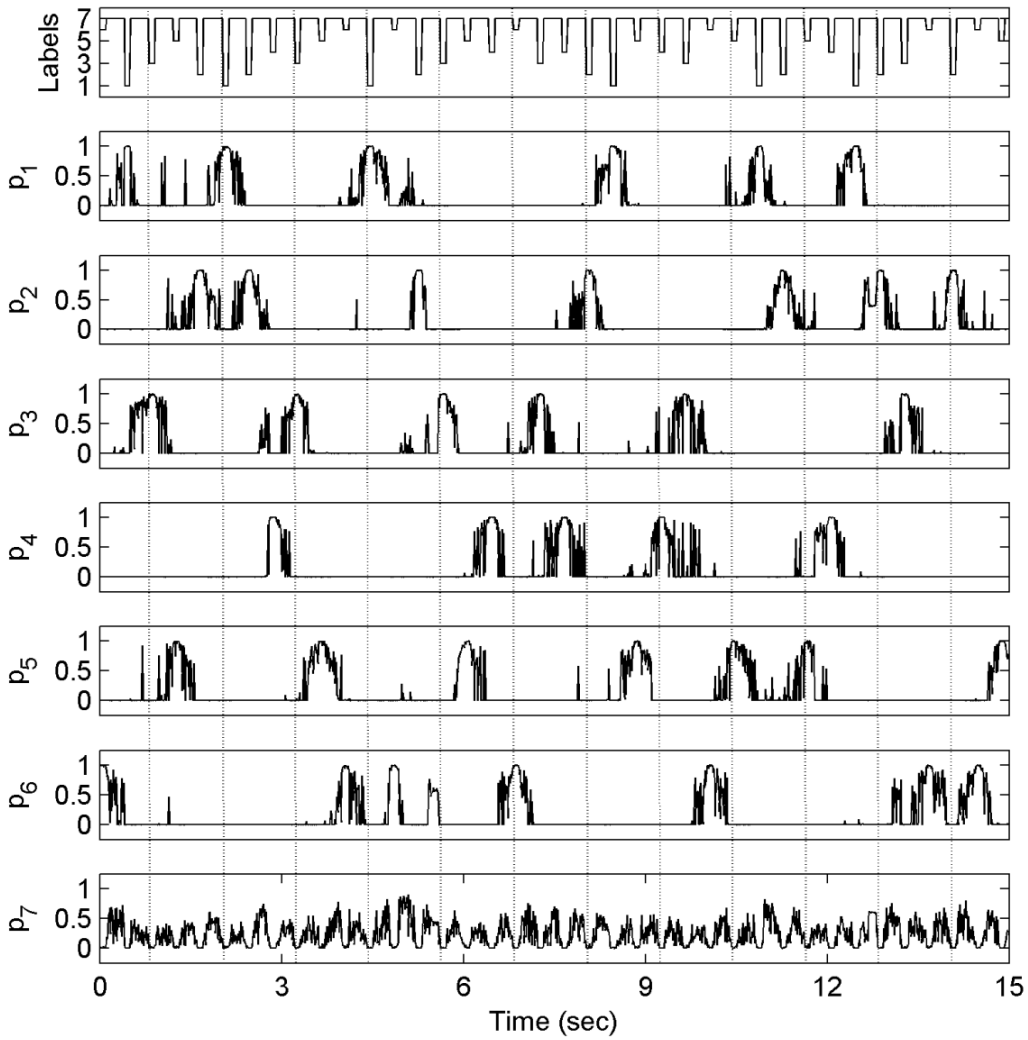

Figure 21. The posterior probability curves obtained with the labeling system which is described in section 5.2.1

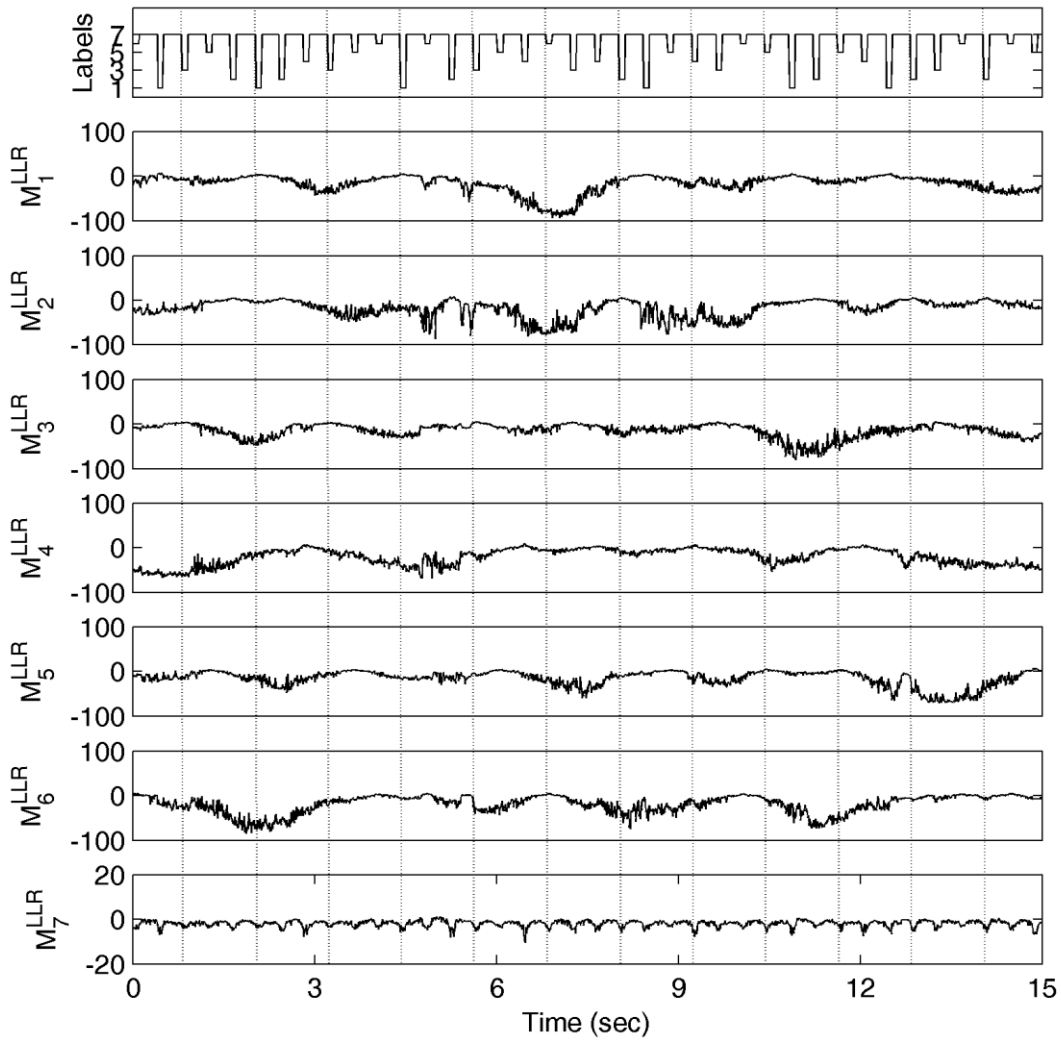

Figure 22. The class overlap measures according to stimuli

When applied to the EEG profiles produced by independent component analysis, the optimum reference set size n *l*ike section one, we applied independent component analysis to this experiment setup and search out the details between samples. We applied the independent component analysis and then determined the optimum reference set size n as 35 (Fig-23). After determining the optimum n we estimated the posterior probabilities of each sample according to the image classes (Fig-24), calculate the confusion matrix (Table 3) and the log-likelihood ratios (Fig-25). The algorithm performance was measured as %33.23. We can say that independent component analysis make clear the features.

Table 4. Confusion Matrix of Visual Stimuli vs Blank Screen Using Quasi-Supervised Learning Algorithm after Independent Component Analysis

Real Labels	Estimated Labels								Total
		1	2	3	4	5	6	7	
	1	77	0	0	0	0	0	0	77
	2	0	91	0	0	0	0	0	91
	3	0	0	77	0	0	0	0	77
	4	0	0	0	63	0	0	0	63
	5	0	0	0	0	91	0	0	91
	6	0	0	0	0	0	90	0	90
	7	235	186	217	167	236	241	149	1431
Total		312	277	294	230	327	331	149	1920

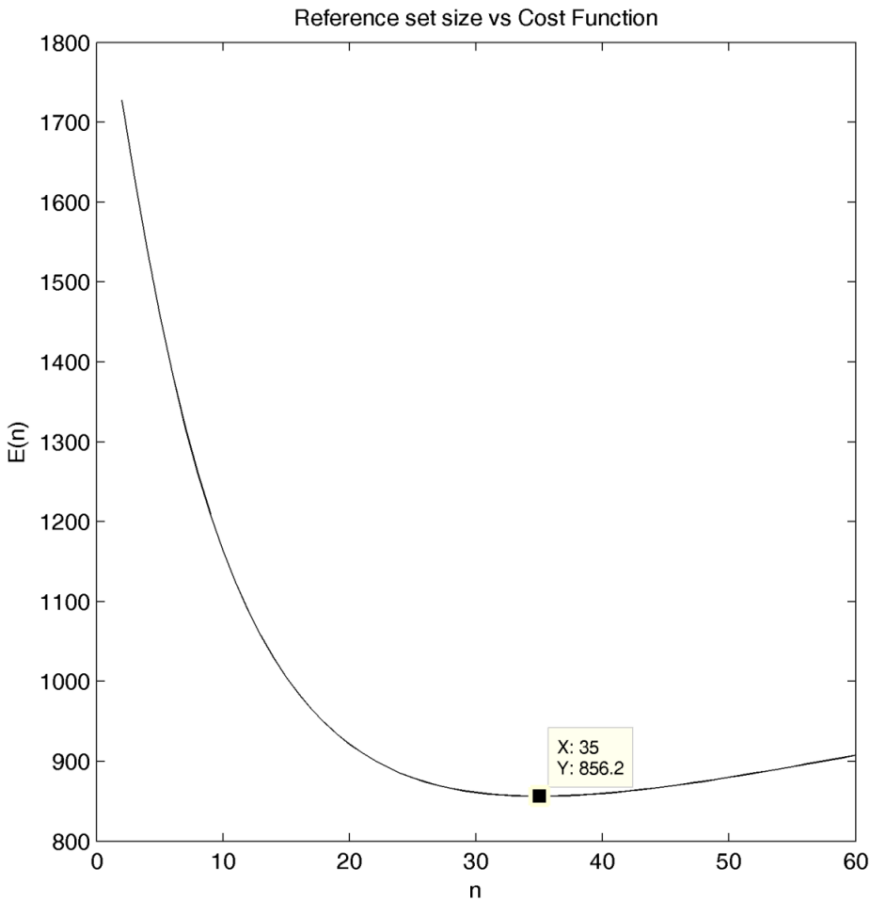

Figure 23. The plot of $\boldsymbol{E(n)}$ for n =1,...,60 after independent components are identified. The functional $\boldsymbol{E(n)}$ attained its minimum value at $\boldsymbol{n = 35}$.

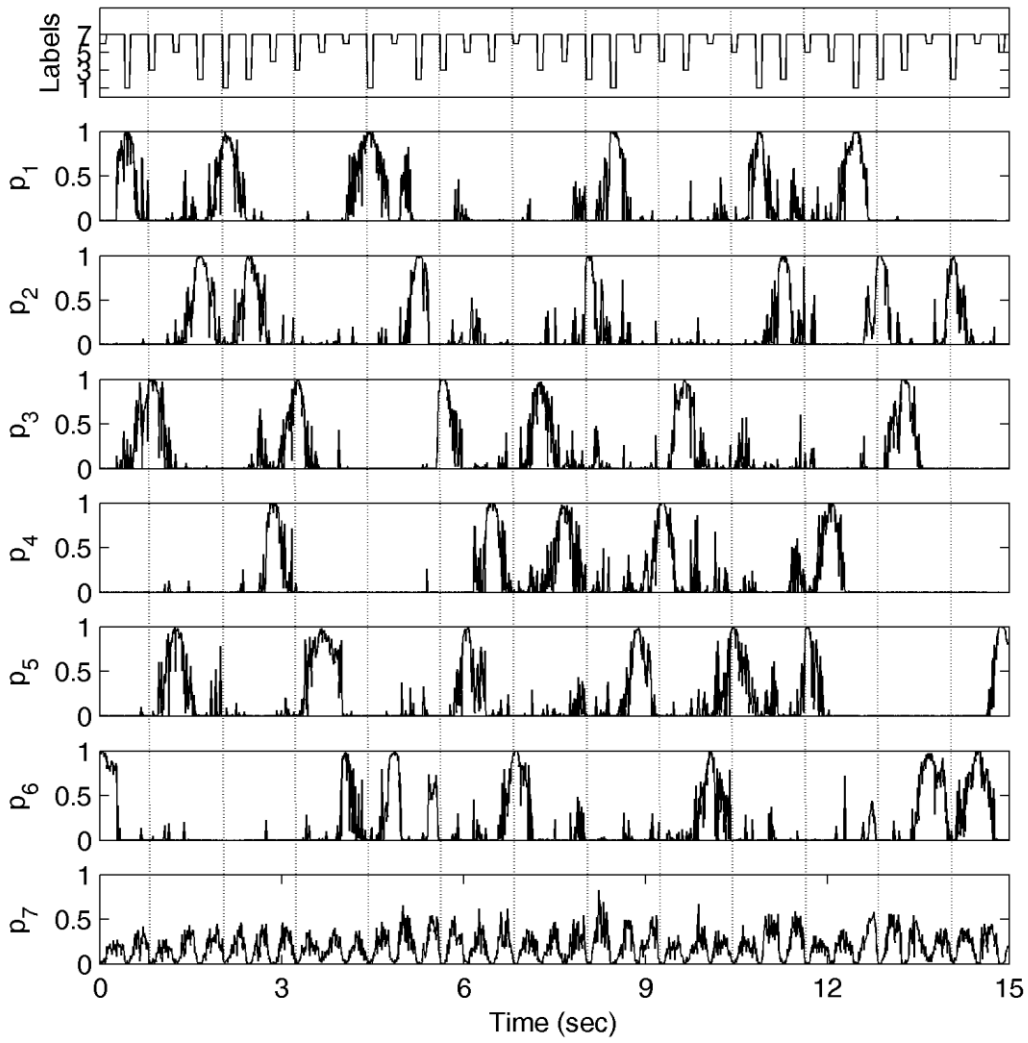

Figure 24. The posterior probability curves after independent component analysis was applied to data with the labeling system which was described in section 5.3

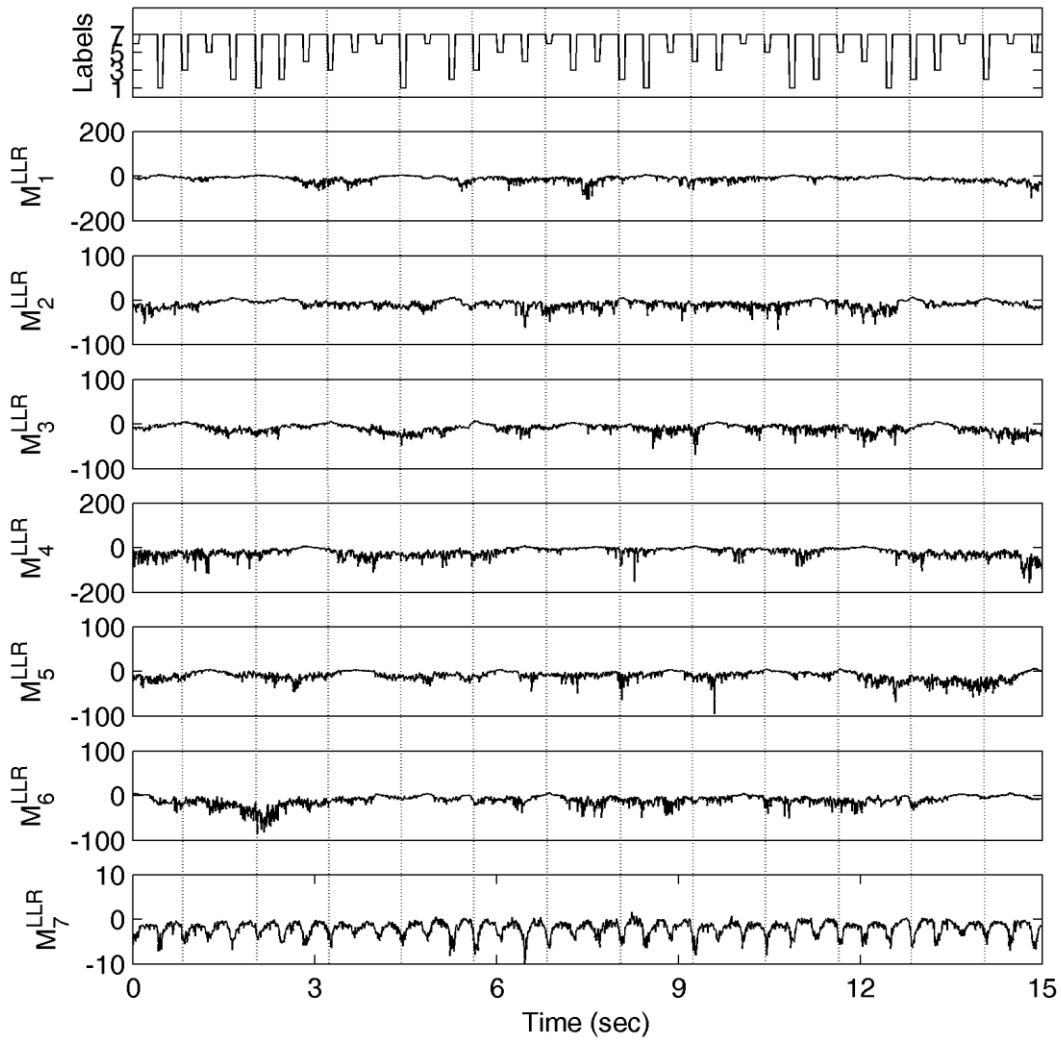

Figure 25. The class overlap measures according to stimuli after independent component analysis was applied to data.

In the previous part, we applied the wavelet decomposition before using quasi-supervised learning algorithm. The results showed that wavelet decomposition creates more distinct areas between the different classes. Because of this we decided to apply same method in this experiment. Again, the optimum reference set size was calculated and the optimum n was chosen as 27 (Fig-26). The algorithm performance is calculated as % 32.76. There were some peaks like noise in posterior probability curves in the previous sections but wavelet decomposition reduced this effect and more smooth curves were created.

Table 5. Confusion Matrix of Visual Stimuli vs Blank Screen Using Quasi-Supervised Learning Algorithm after Wavelet Decomposition

Real Labels	Estimated Labels								Total
		1	2	3	4	5	6	7	
	1	77	0	0	0	0	0	0	77
	2	0	91	0	0	0	0	0	91
	3	0	0	77	0	0	0	0	77
	4	0	0	0	63	0	0	0	63
	5	0	0	0	0	91	0	0	91
	6	0	0	0	0	0	90	0	90
	7	180	247	188	139	252	285	140	1431
Total		257	338	265	202	343	375	140	1920

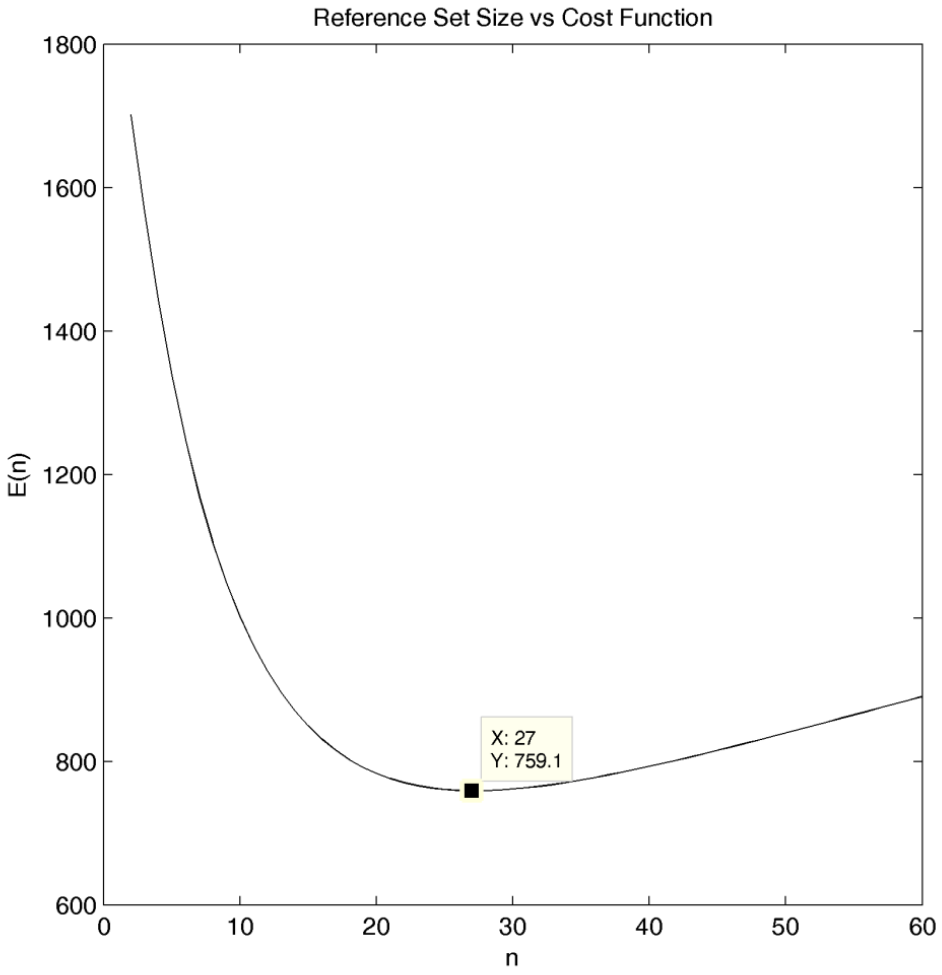

Figure 26. The plot of $\boldsymbol{E(n)}$ for n =1,...,60 after the wavelet decomposition is applied to data. The functional $\boldsymbol{E(n)}$ attained its minimum value at $\boldsymbol{n = 27}$.

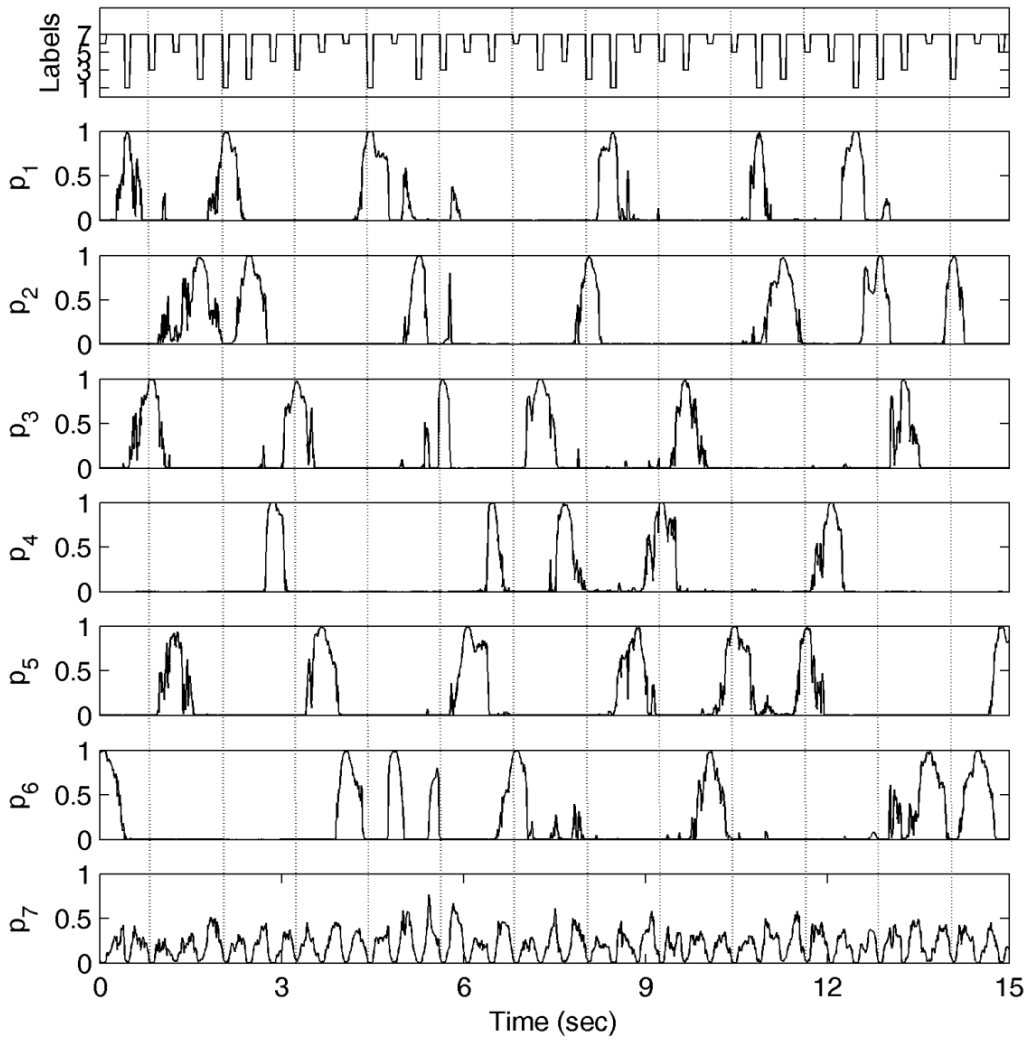

Figure 27. The posterior probability curves after wavelet decomposition was applied to data with the labeling system which is described in section 5.4

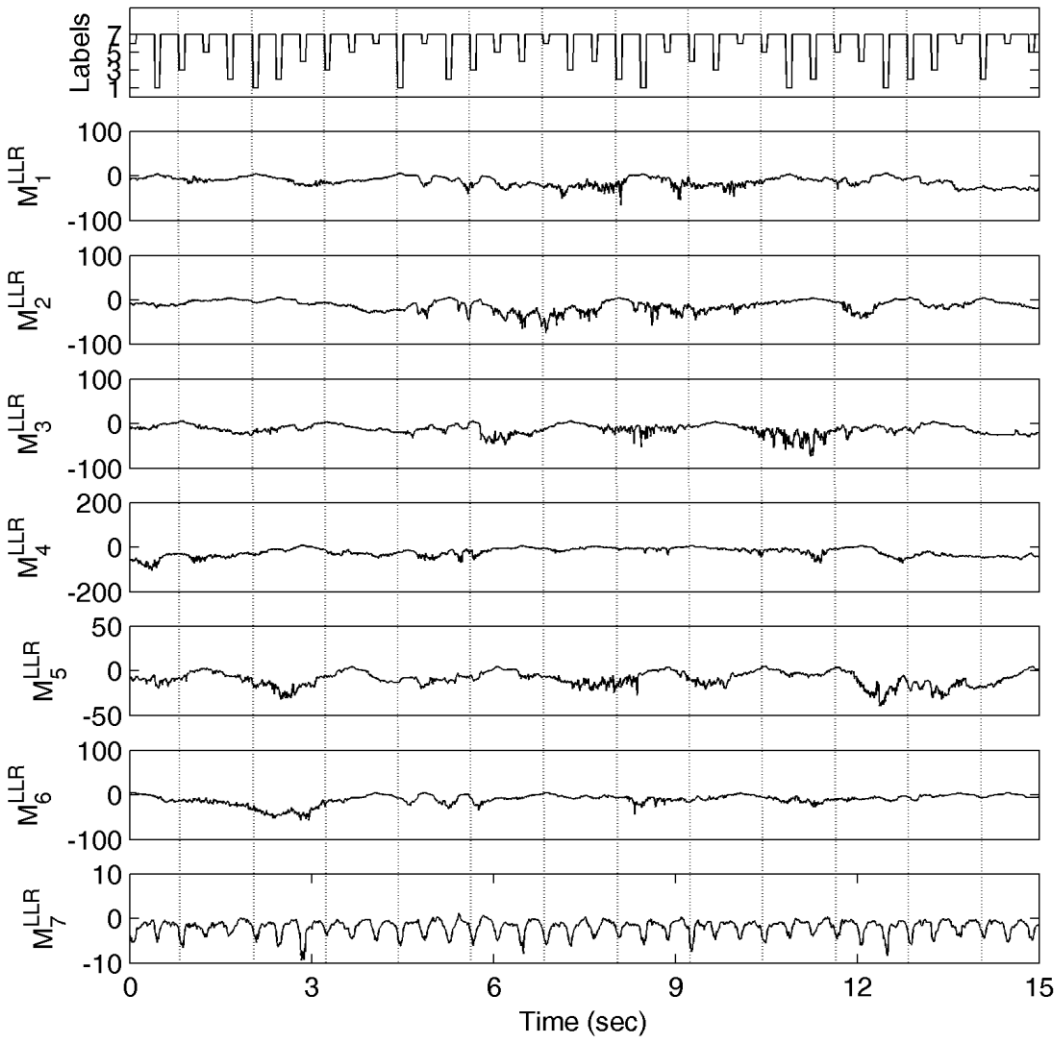

Figure 28. The class overlap measures according to stimuli after wavelet decomposition was applied to data.

As we described in the section 5.1, we applied independent component analysis to data which expanded using wavelet decomposition. After the feature reduction by independent component analysis we applied the quasi-supervised learning algorithm and the algorithm performance increased to %59.64 when the reference set size n chosen as 42.

Table 6. Confusion Matrix of Visual Stimuli vs Blank Screen Using Quasi-Supervised Learning Algorithm after Wavelet Decomposition

		Estimated Labels							Total
		1	2	3	4	5	6	7	
Real Labels	1	77	0	0	0	0	0	0	77
	2	0	91	0	0	0	0	0	91
	3	0	0	77	0	0	0	0	77
	4	0	0	0	63	0	0	0	63
	5	0	0	0	0	91	0	0	91
	6	0	0	0	0	0	90	0	90
	7	137	101	131	60	128	218	656	1431
Total		214	192	208	123	219	308	656	1920

All results show that, the visual stimuli classification did not change. Always same number of sample is classified as with image class labels but the blank screen classification changed. As seen from the posterior probability curves the posterior probability of the seventh class (when the screen is blank) never went to one. It moved around the 0.5. This show us the samples which are labeled with their own stimuli label can be used in the classification algorithms as train data.

From posterior probability curves we also see that, a visual stimuli effect was not only in the time interval which was on the screen (100ms), it continued when the screen went to blank. So we decided to change our labeling system and applied all methods the new labeled data.

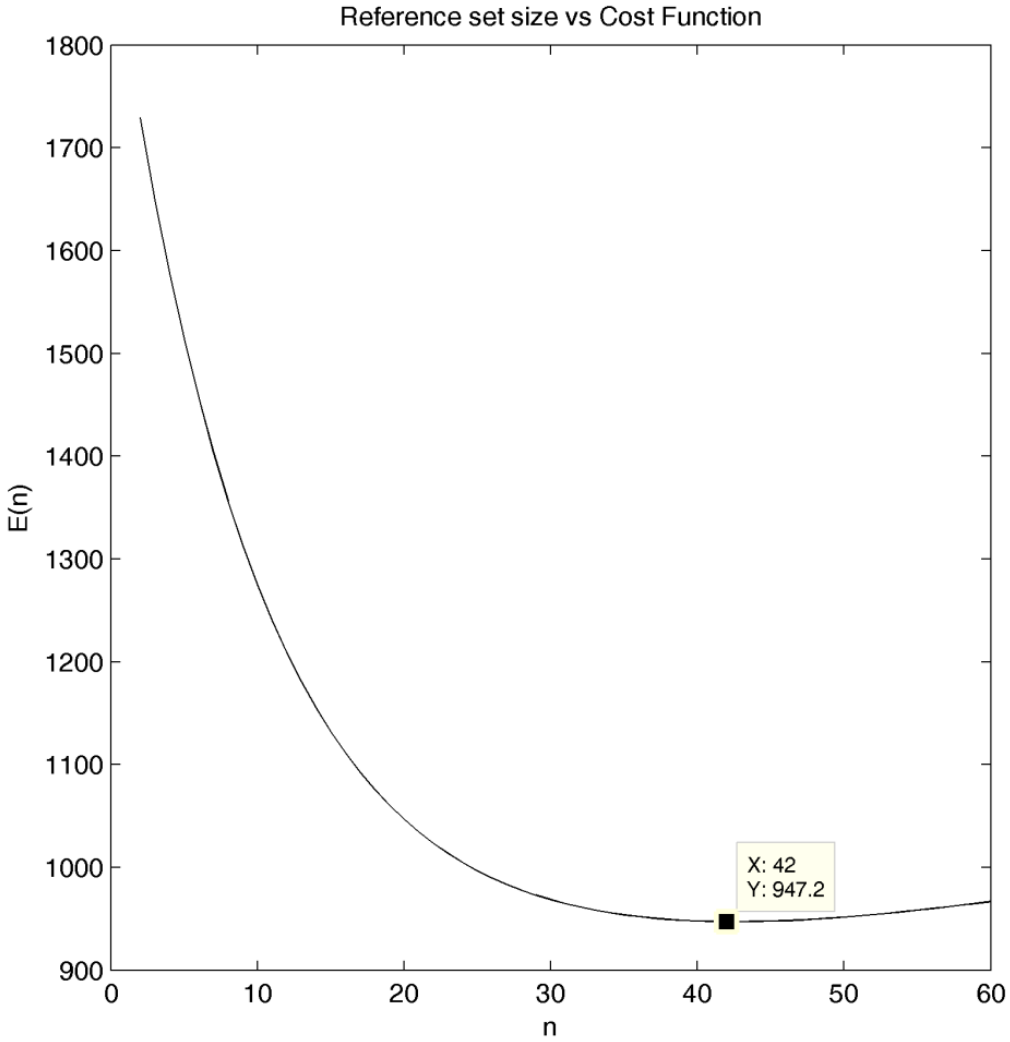

Figure 29. The plot of $\boldsymbol{E(n)}$ for n =1,...,60 after independent component analysis was applied to wavelet decomposition-based EEG profiles. The functional $\boldsymbol{E(n)}$ attained its minimum value at $\boldsymbol{n = 42}$.

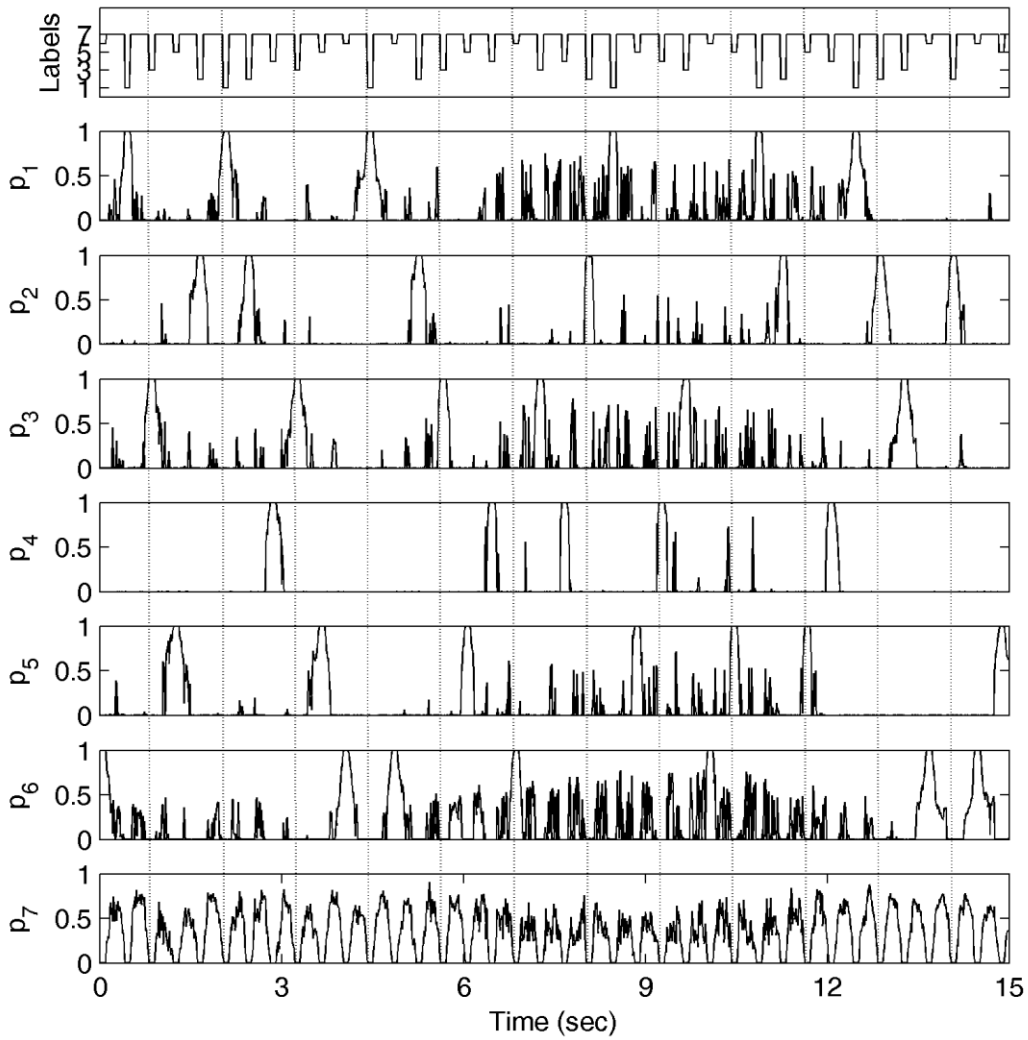

Figure 30. The posterior probability curves wavelet decomposition and independent component analysis was applied to data with the labeling system which is described in section 5.4

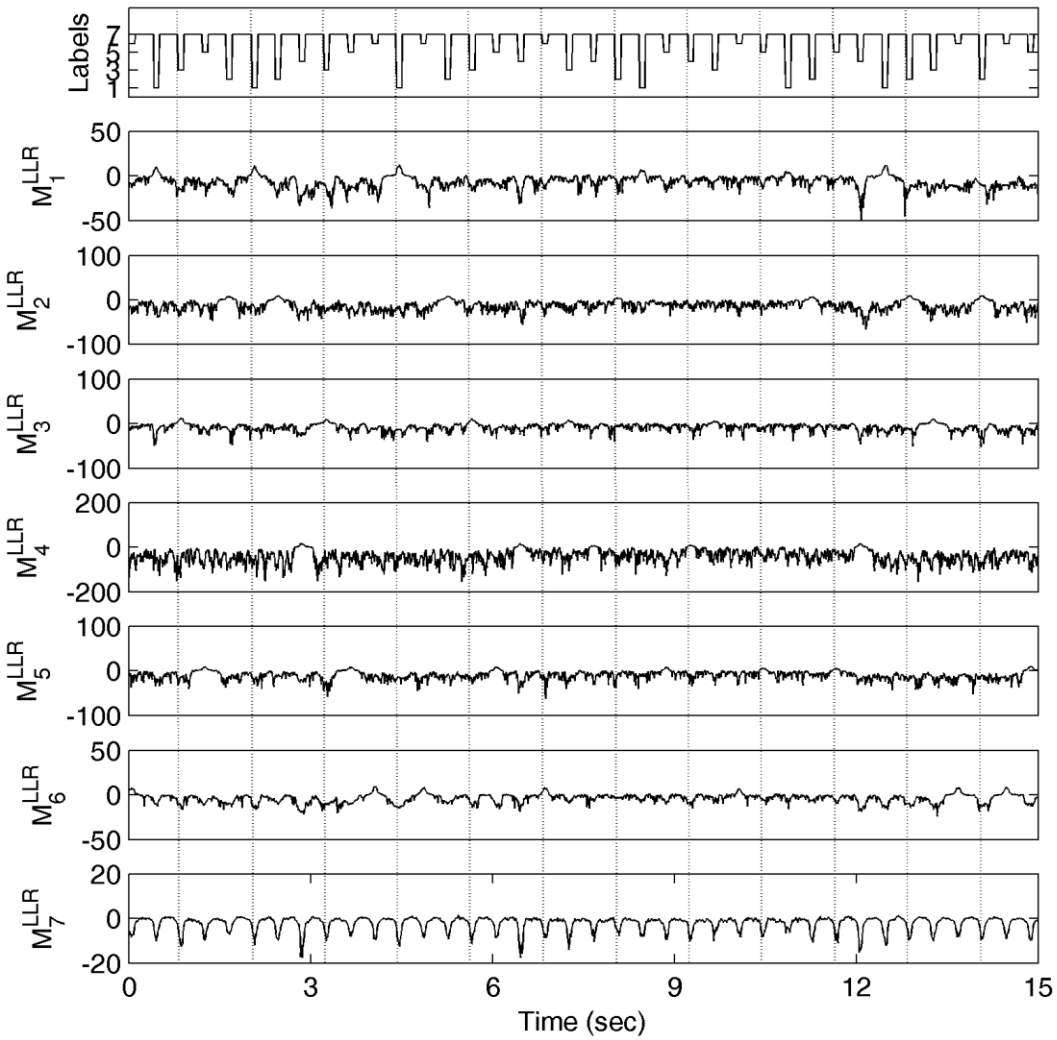

Figure 31. The class overlap measures according to stimuli after wavelet decomposition and independent component analysis was applied to data.

5.5. Differences of Visual Stimuli

The results in the section 5.3 show that, the visual stimuli effect continues after the stimulus offset and it proceeds until the other stimulus was onset. That suggests that a better labeling system which covers the stimulus time with the blank screen interval following the stimulus offset. In other words, in the new experiments, we labeled the 400ms after the stimulus onset with the stimuli image index. After labeling we run the same algorithms in section 5.4.

Under this new labeling setting, the M-ary quasi-supervised learning algorithm on the original EEG profiles obtained a separation performance of %95.31. In this case the optimum reference set size was determined be 49.

Table 7. Confusion Matrix of Differences Between Visual Stimuli Using Quasi-Supervised Learning Algorithm

		Estimated Labels						Total
		1	2	3	4	5	6	
Real Labels	1	293	1	0	4	3	4	305
	2	9	343	2	3	1	2	360
	3	5	0	289	4	9	1	308
	4	0	0	6	249	2	0	257
	5	2	1	3	4	319	3	332
	6	1	8	4	5	3	337	358
Total		310	353	304	269	337	347	1920

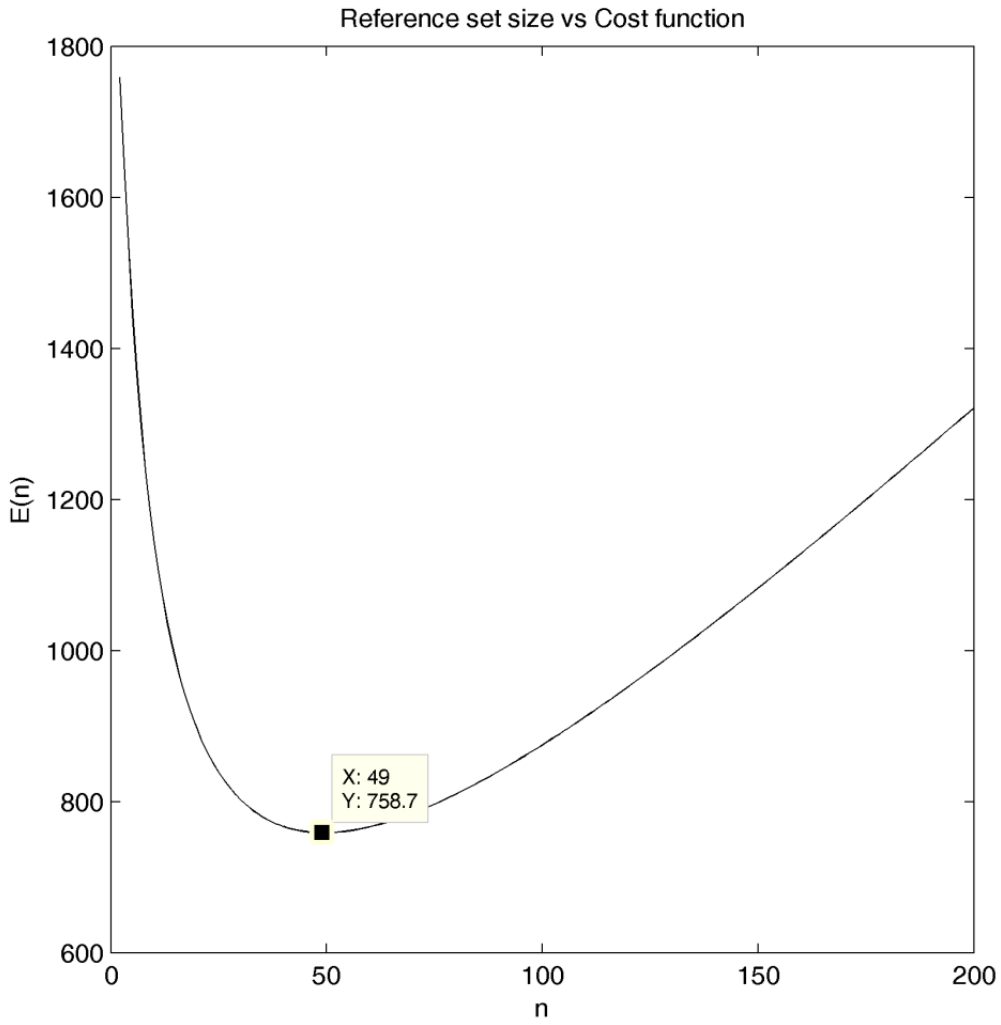

Figure 32. The plot of $\boldsymbol{E(n)}$ for n =1,...,200. The functional $\boldsymbol{E(n)}$ attained its minimum value at $\boldsymbol{n = 49}$.

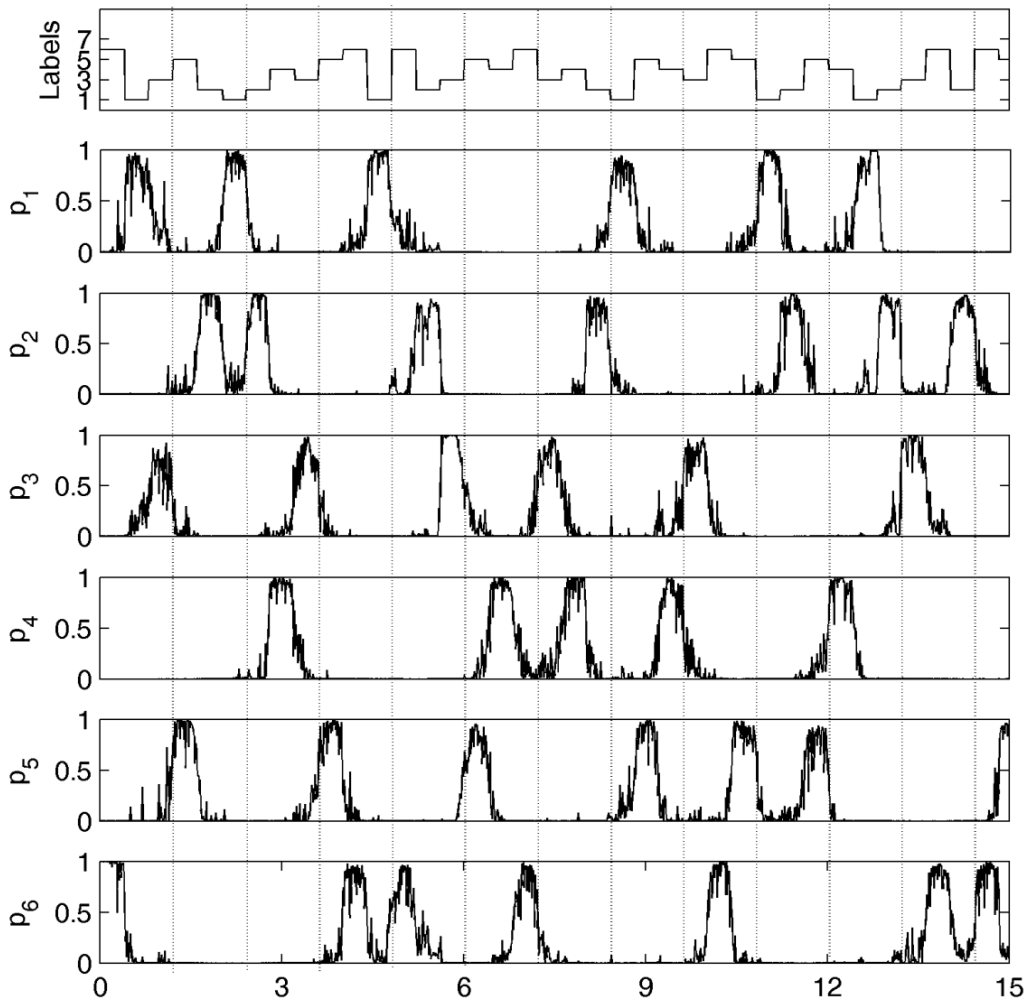

Figure 33. The posterior probability curves with the labeling system which is described in section 5.5

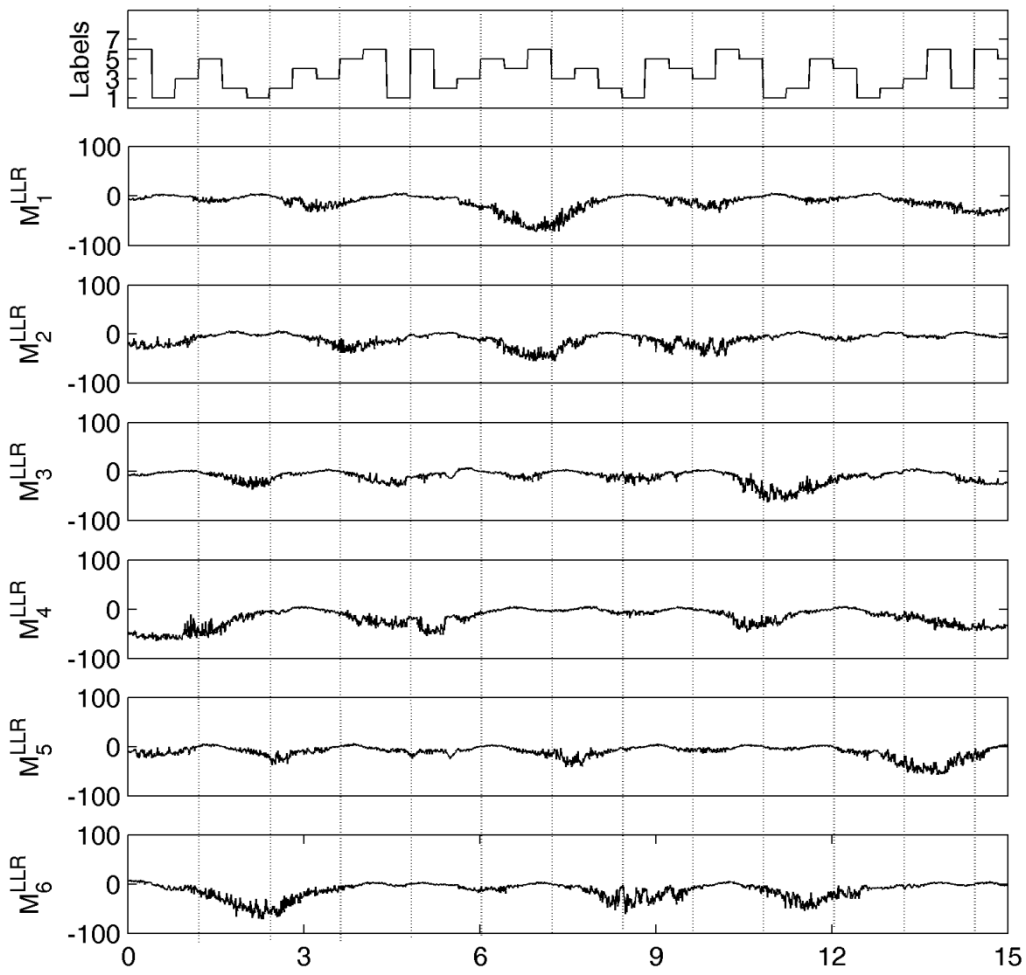

Figure 34. The class overlap measures according to stimuli

The separation of the stimulus using the independent component analysis profiles. In this experiments the optimum reference set size was 56 , and the separation performance was %97.81. This shows that independent component analysis makes more explicit the features of the samples, and provides more qualified input data to supervised learning algorithms.

Table 8. Confusion Matrix of Differences between Visual Stimuli after Independent Component Analysis Using Quasi-Supervised Learning Algorithm

		Estimated Labels						Total
		1	2	3	4	5	6	
Real Labels	1	298	1	0	0	2	4	305
	2	3	355	0	2	0	0	360
	3	1	0	300	3	3	1	308
	4	0	0	2	254	1	0	257
	5	0	2	0	4	325	1	332
	6	3	3	0	1	5	346	358
Total		305	361	302	264	336	352	1920

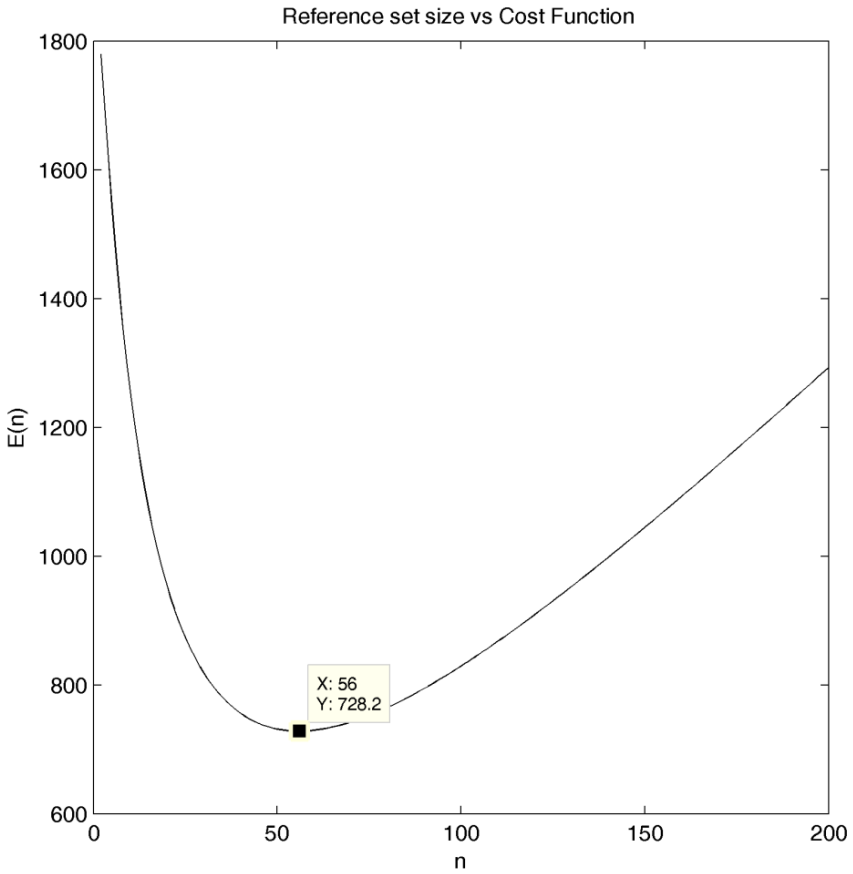

Figure 35. The plot of $\boldsymbol{E(n)}$ for n =1,...,200 after independent component analysis applied to data. The functional $\boldsymbol{E(n)}$ attained its minimum value at $\boldsymbol{n = 56}$.

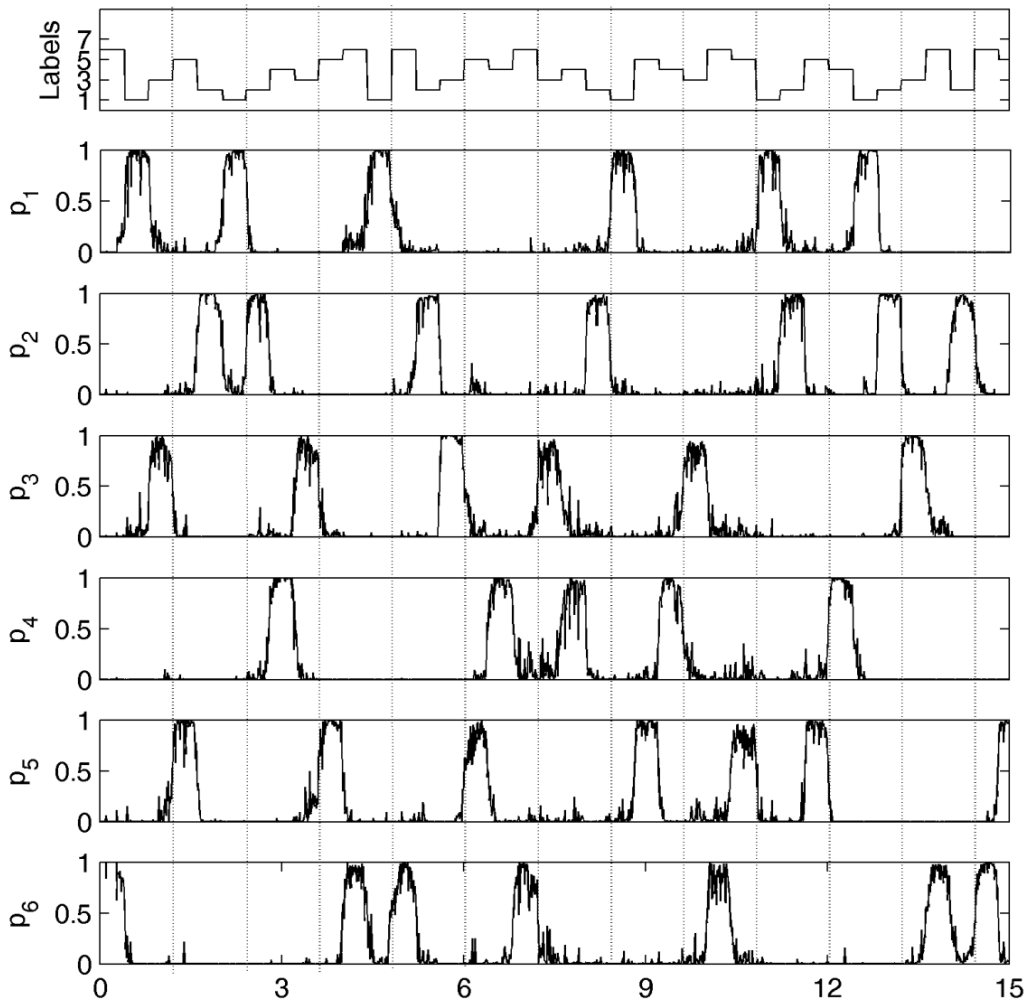

Figure 36. The posterior probability curves after independent component analysis is applied to data with the labeling system which is described in section 5.5

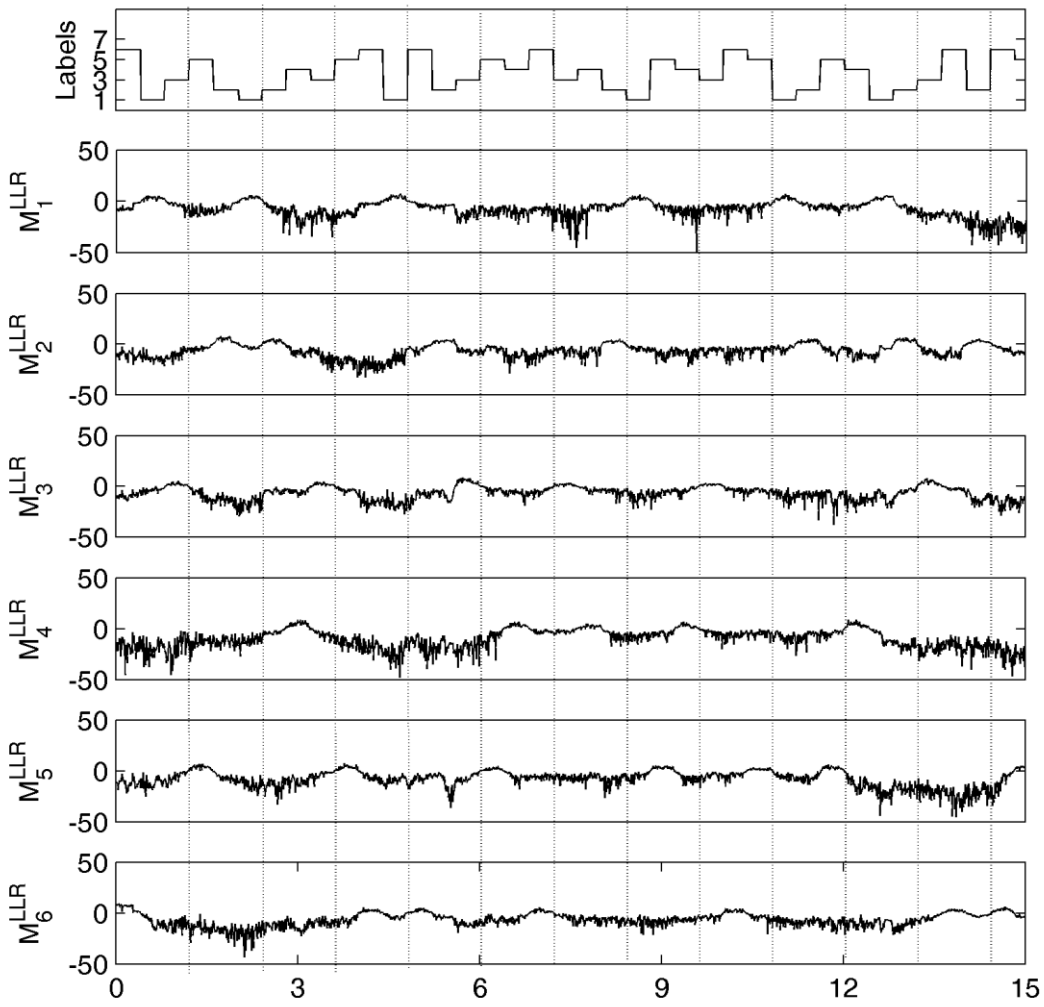

Figure 37. The class overlap measures according to stimuli after independent component analysis is applied.

The separation performance was improved further using EEG profiles based on the wavelet decomposition the optimal reference set size n was 48 (see Figure 38). The separation performance was calculated as %98.39.

Table 9. Confusion Matrix of Differences Between Visual Stimuli After Wavelet Decomposition Using Quasi-Supervised Learning Algorithm

		Estimated Labels						Total
		1	2	3	4	5	6	
Real Labels	1	302	0	0	0	1	2	305
	2	3	352	0	2	1	2	360
	3	0	0	297	2	9	0	308
	4	0	1	0	255	1	0	257
	5	0	0	2	3	327	0	332
	6	0	1	0	1	0	356	358
Total		305	354	299	263	339	360	1920

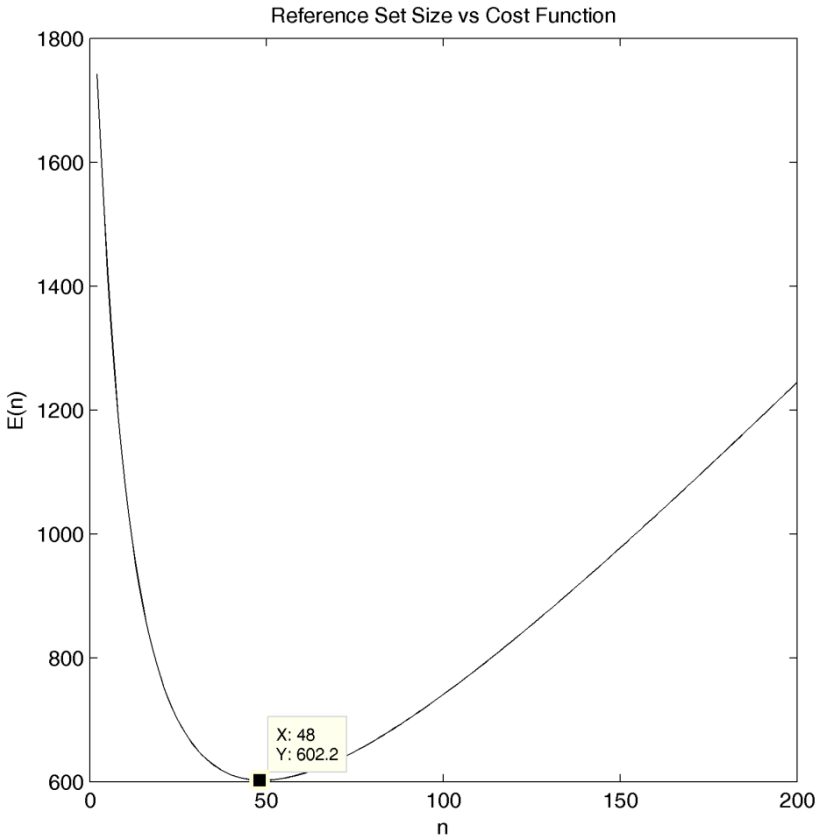

Figure 38. The plot of $\boldsymbol{E(n)}$ for n =1,...,200 for wavelet decomposition-based EEG profiles. The functional $\boldsymbol{E(n)}$ attained its minimum value at $\boldsymbol{n = 48}$.

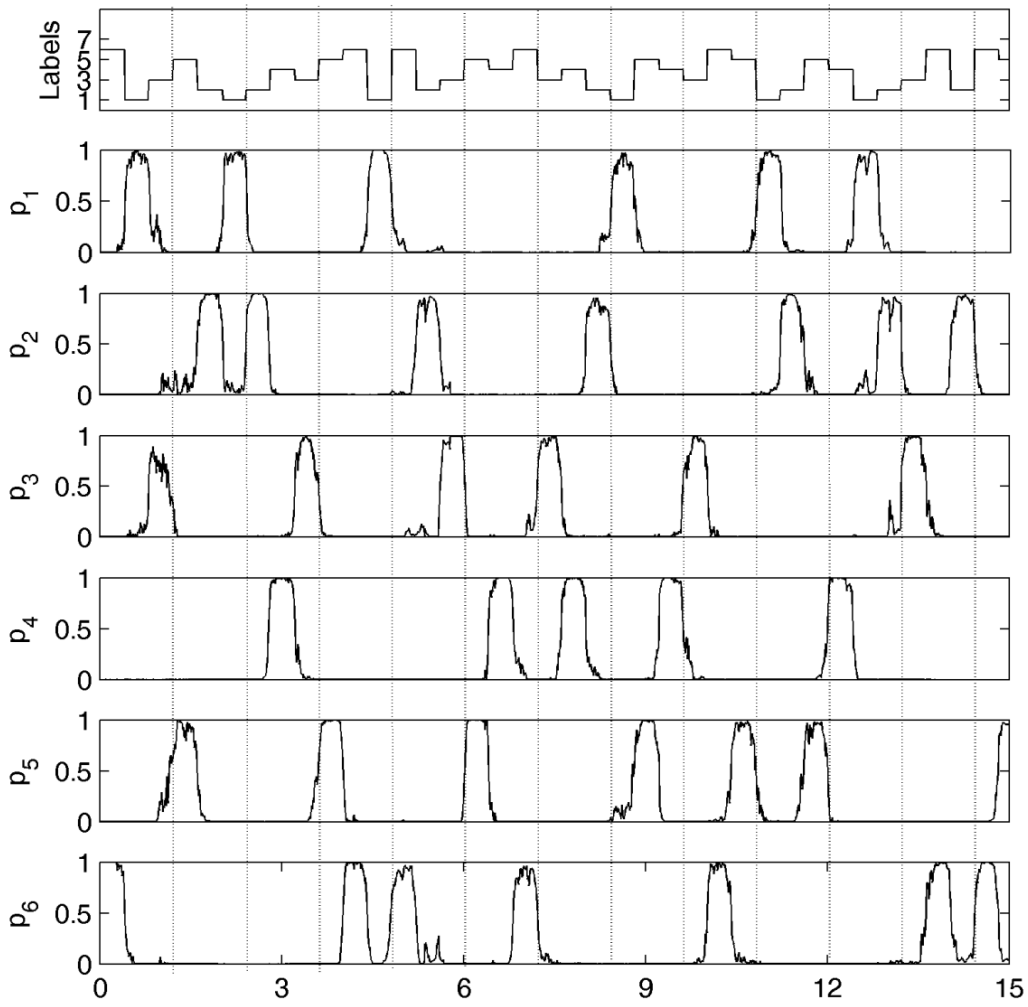

Figure 39. The posterior probability curves wavelet decomposition is applied to data with the labeling system which is described in section 5.5

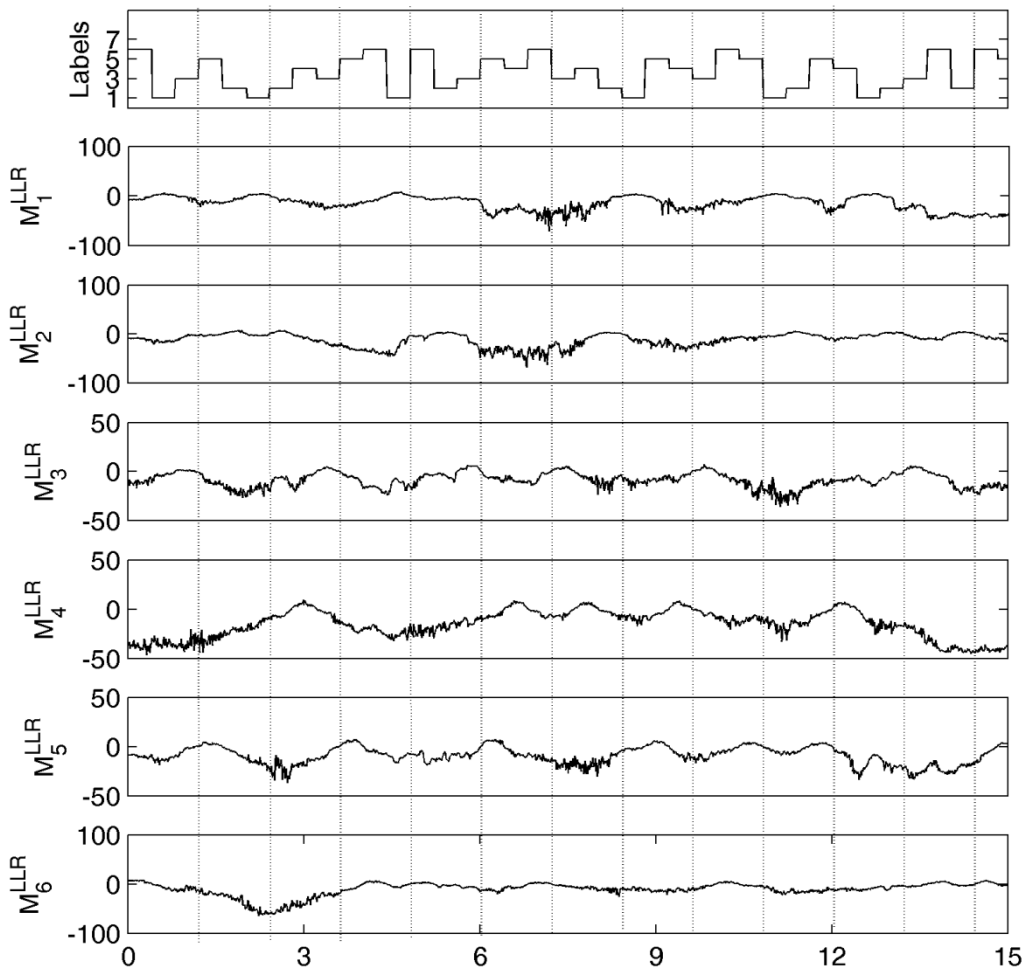

Figure 40. The class overlap measure according to stimuli after wavelet decomposition is applied.

The EEG profiles obtained by applying independent component analysis and the wavelet profiles produced greatest improvement on the separation performance between the different stimuli. The results show that separation performance increased to %99.64, with a reference set size of 77. The posterior probability curves reveal that the abnormal samples are between the 7-12 seconds.

Table 10. Confusion Matrix of Differences between Visual Stimuli after Wavelet Decomposition and Independent Component Analysis Using Quasi-Supervised Learning Algorithm

		Estimated Labels						Total
		1	2	3	4	5	6	
Real Labels	1	308	0	0	0	0	0	308
	2	0	358	0	2	0	0	360
	3	0	0	308	0	0	0	308
	4	0	0	0	257	0	0	257
	5	0	0	2	0	330	0	332
	6	0	0	1	1	1	355	358
Total		308	358	311	260	331	355	1920

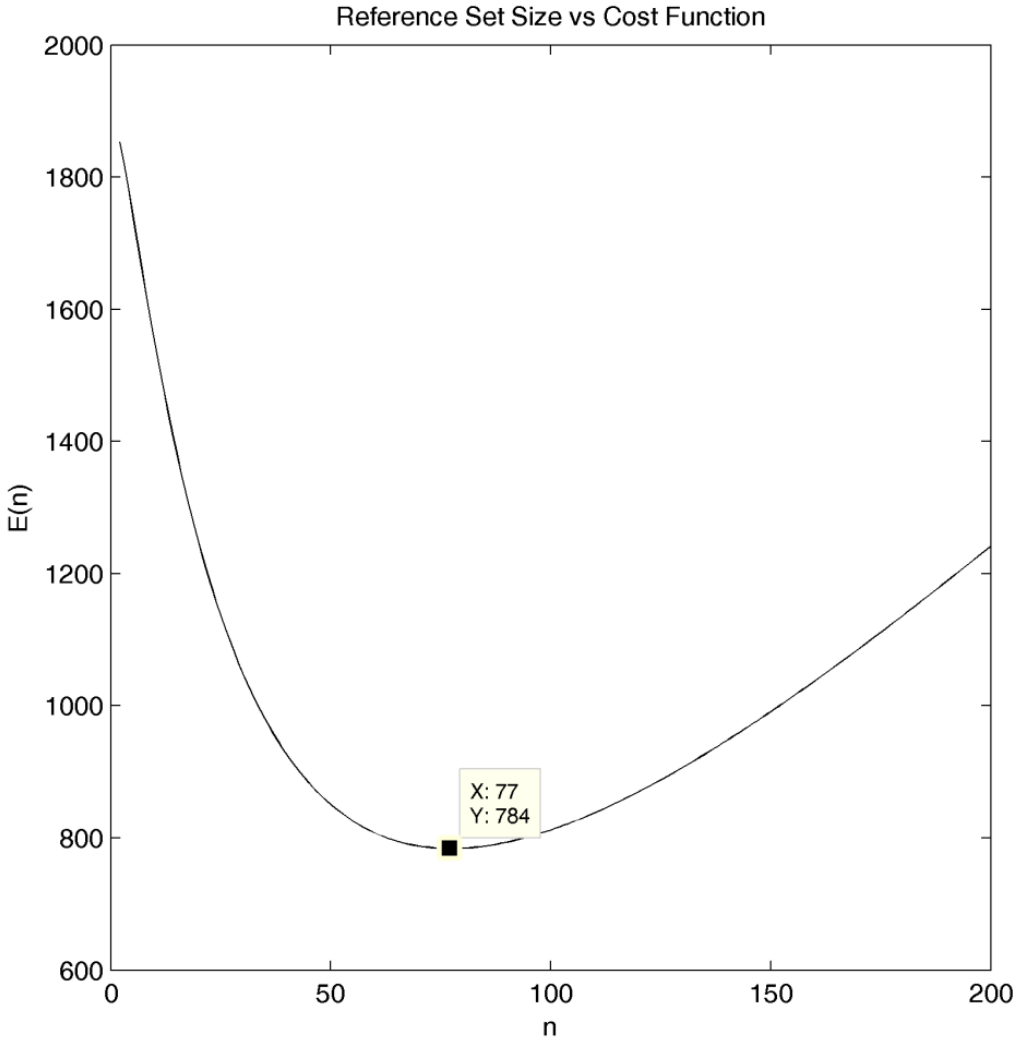

Figure 41. The plot of $\boldsymbol{E(n)}$ for n =1,...,200 for independent component analysis and the wavelet decomposition is applied to data. The functional $\boldsymbol{E(n)}$ is attained its minimum value at $\boldsymbol{n = 77}$.

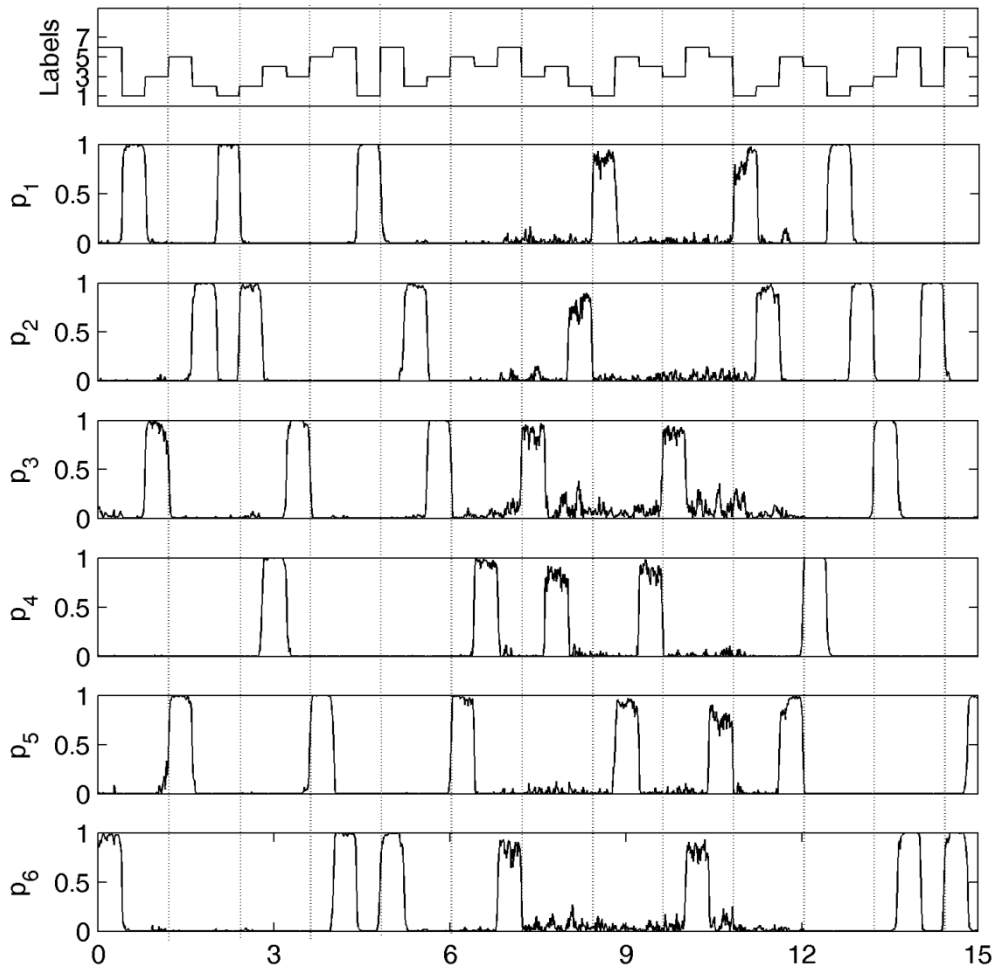

Figure 42. The posterior probability curves wavelet decomposition and independent component analysis were applied to data with the labeling system which is described in section 5.5

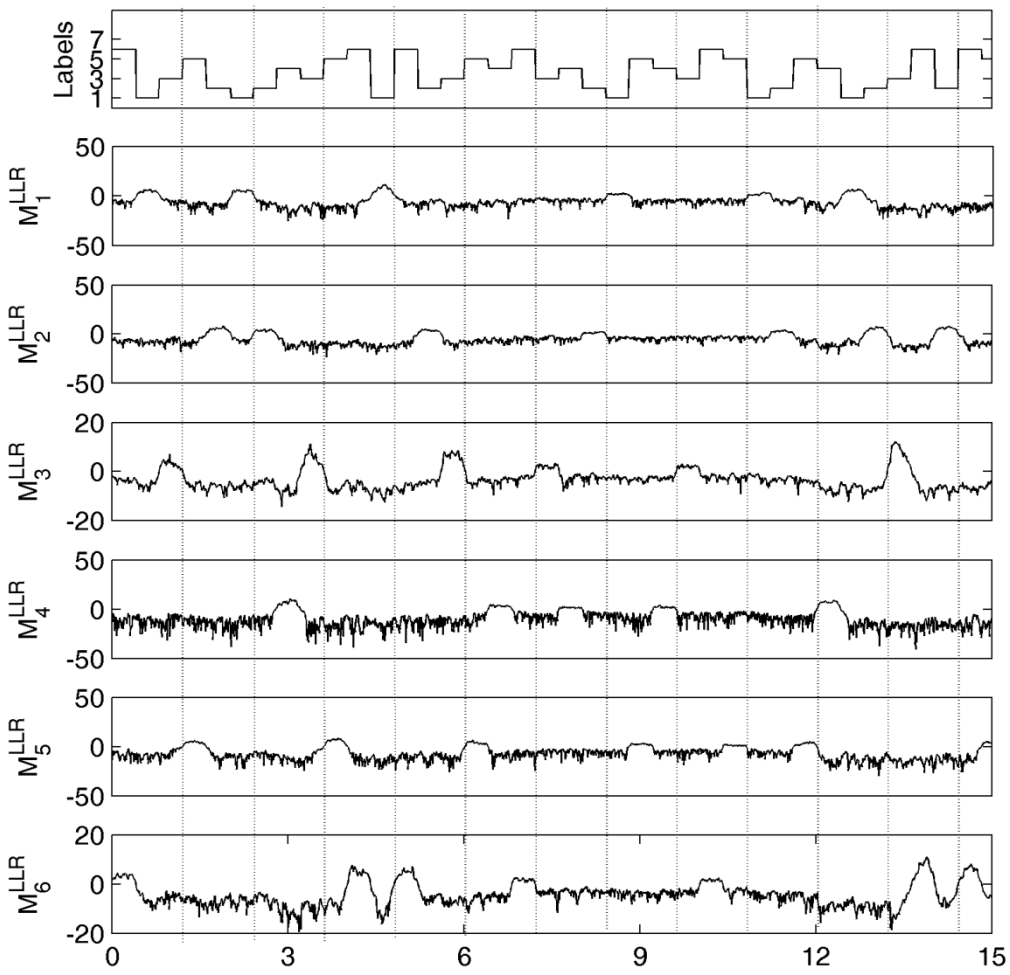

Figure 43. The class overlap measures according to stimuli after wavelet decomposition and independent component analysis were applied.

5.6. Target Image vs Other Images

In the experiment, the subject was asked to count the number of times the fourth image was shown. In order to use the separation between the expectant stimuli and the others, we have carried out the final experiment of this thesis. To this end, we labeled the samples recorded with the images as 1, and the target image interval were labeled as 2. Then we repeated the analytical steps used in previous sections.

We performed the binary quasi-supervised algorithm and identify the target image samples. The algorithm performance is calculated as %90.73.

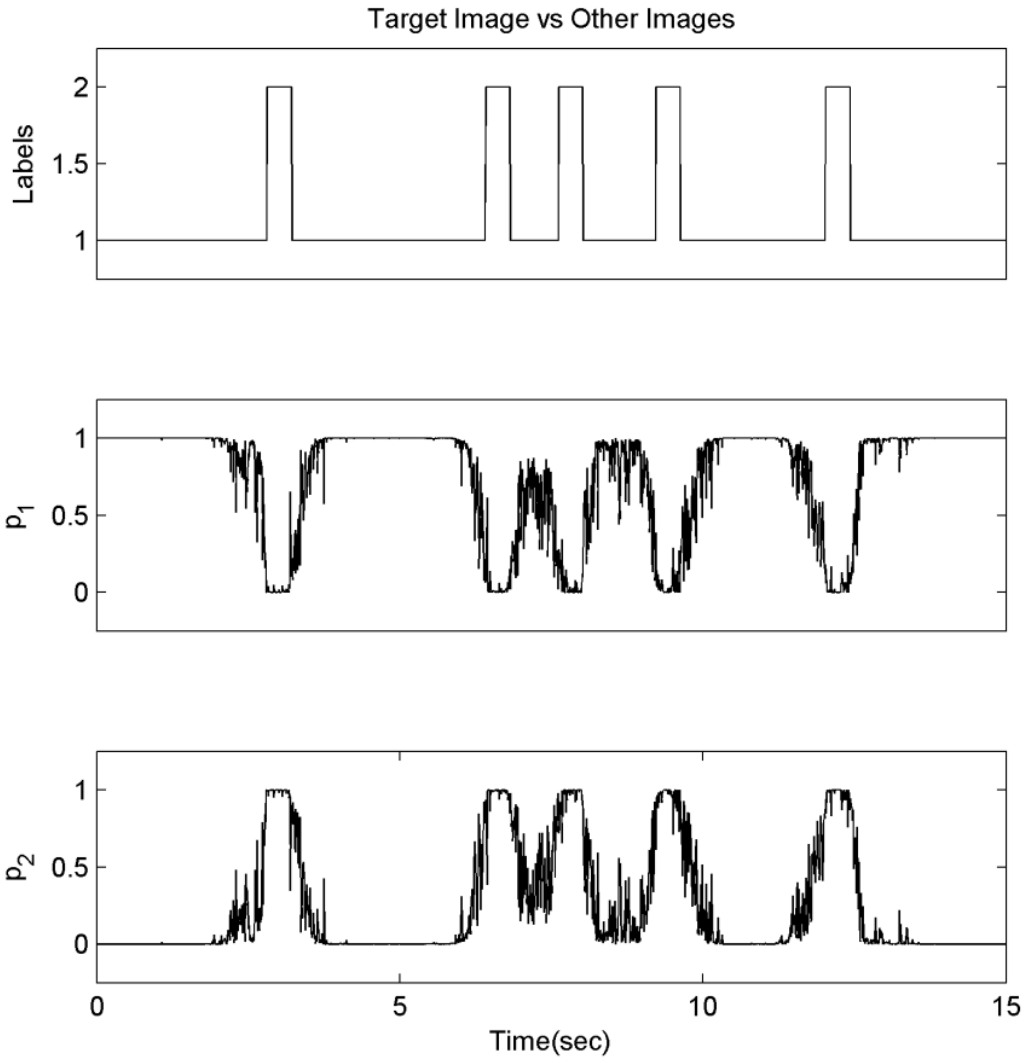

Figure 44. T he posterior probability curves of the data that the target image and other images compared.

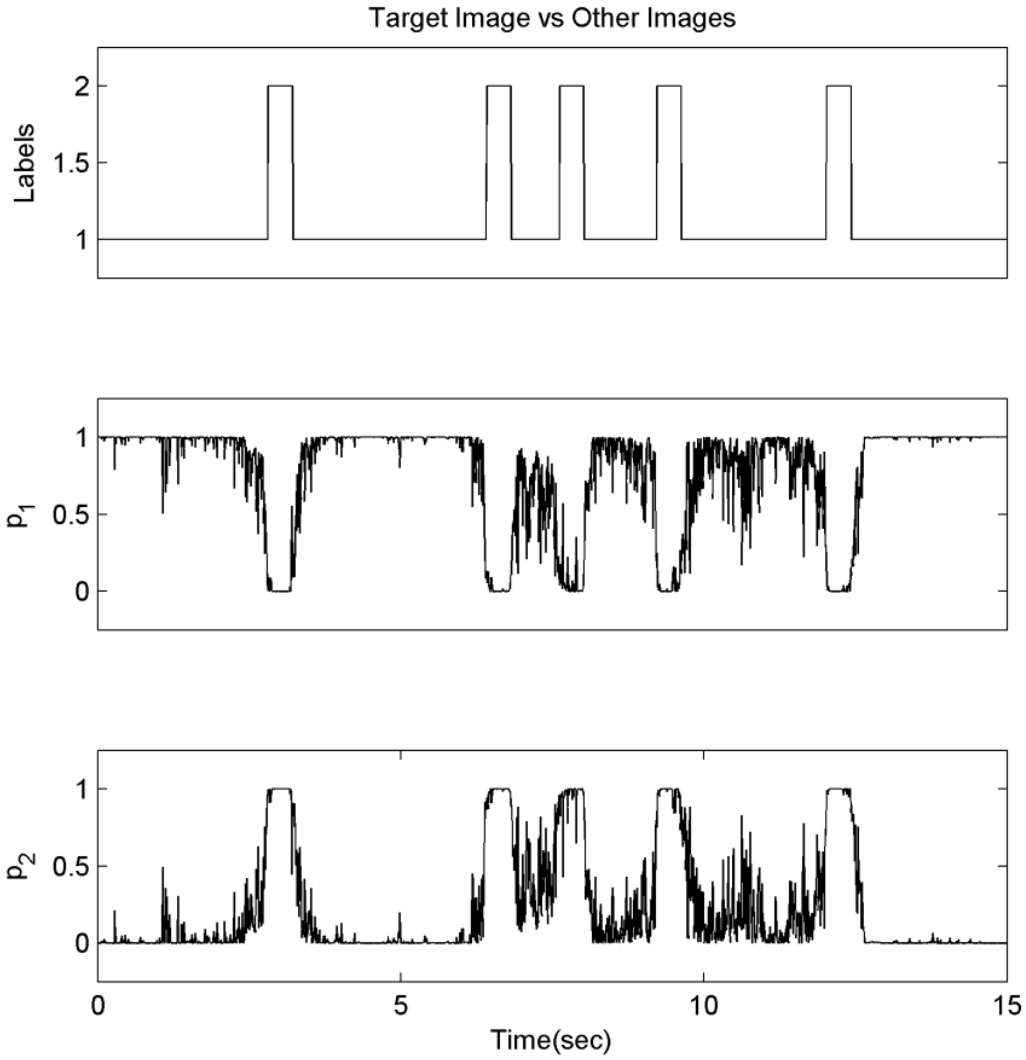

Figure 45. The posterior probability curves of the data which described in section 5.6 after independent component analysis.

As we know from the previous results, independent component analysis is successful for the determine features. So, in this experiment we used the independent component analysis before the quasi-supervised learning algorithm and the algorithm performance increase to %93.65.

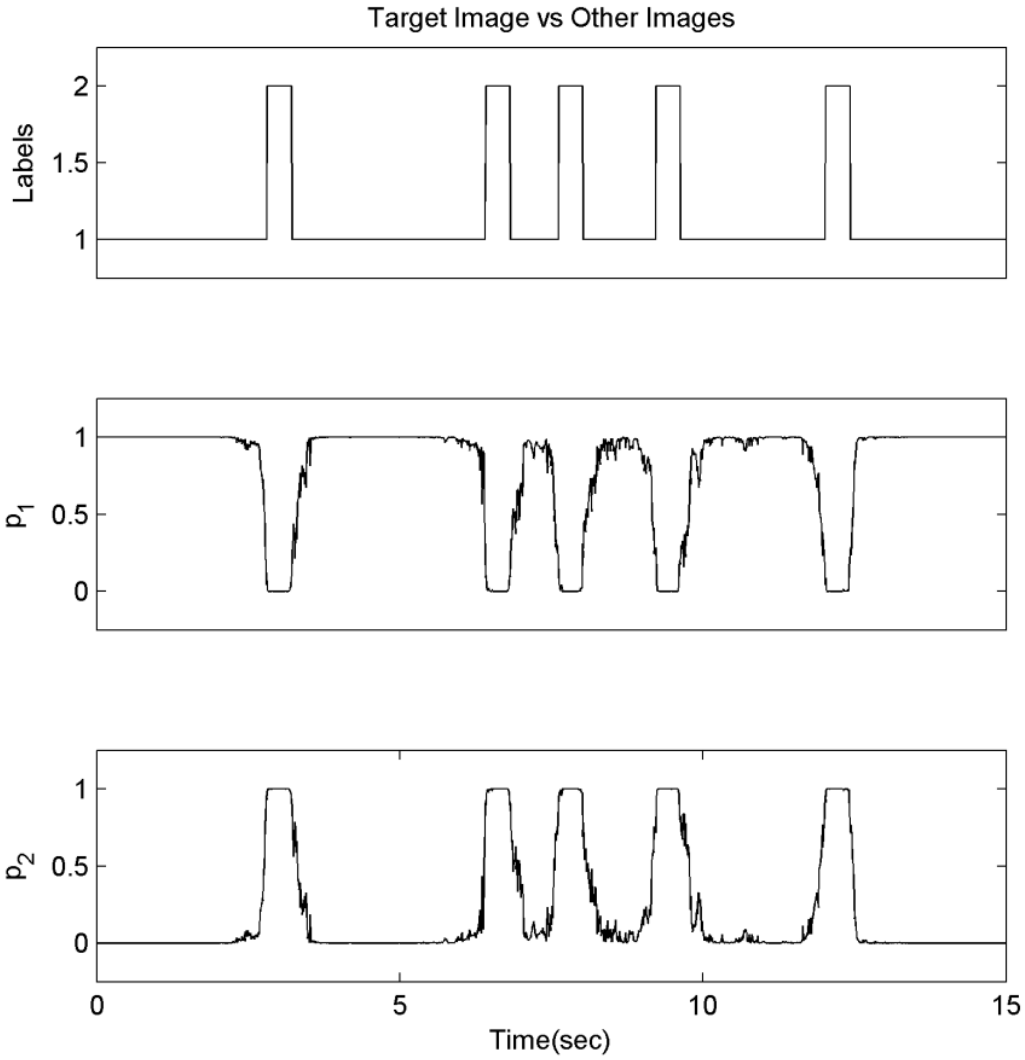

Figure 46. The posterior probability curves of the data which described in section 5.6 after wavelet decomposition.

In this section the data was expand again with the wavelet transform which was described in the previous sections and the quasi-supervised learning algorithm was applied. This process increased the algorithm performance to %94.95.

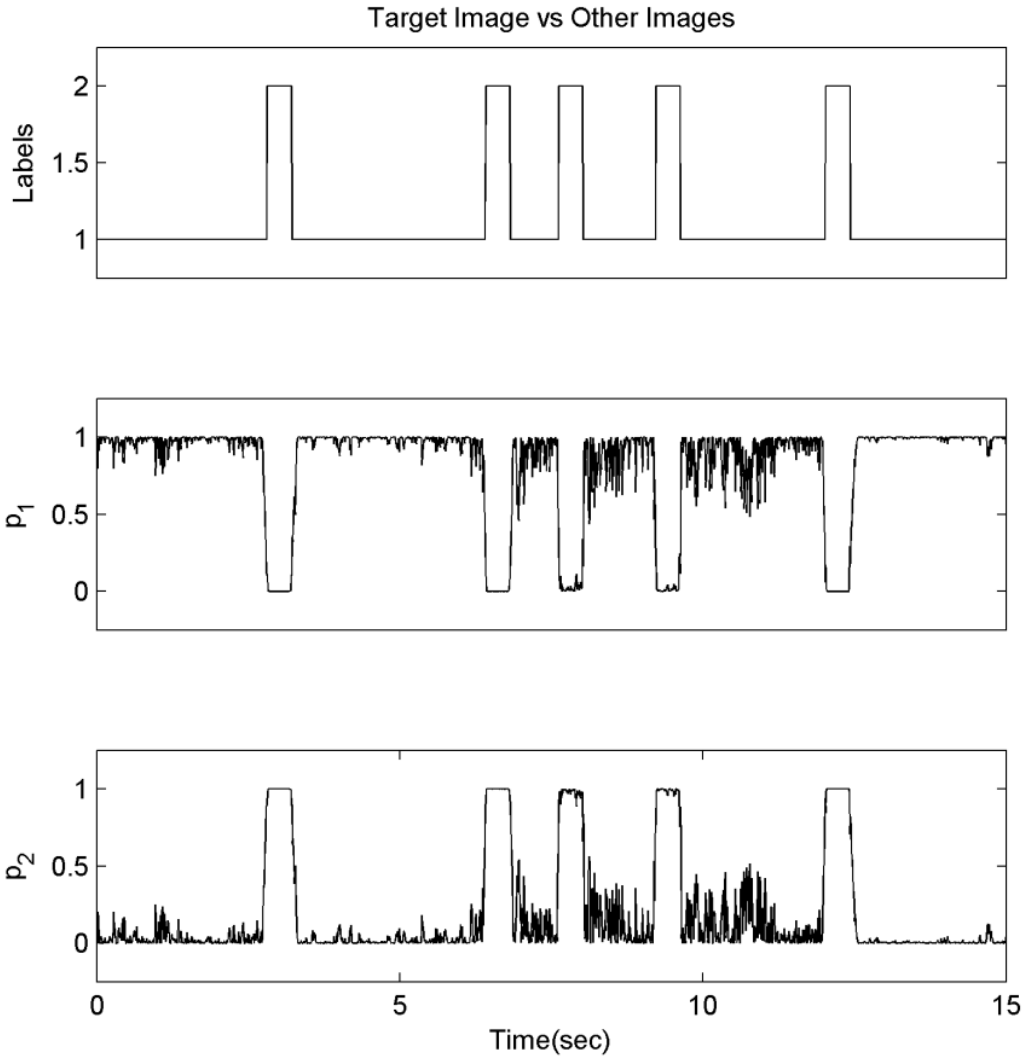

Figure 47. The posterior probability curves of the data which described in section 5.6 after wavelet decomposition and independent component analysis.

As we know, the combination of independent component analysis and wavelet decomposition. When we applied this algorithm to data, in the target image identification algorithm performance became %97.97.

CHAPTER 6

CONCLUSION

We demonstrated that binary and M-ary quasi-supervised learning algorithm is successful to automatically measure the separation of stimuli-specific brain activity patterns in EEG.

In this study, the EEG dataset recorded under randomly flashed six different images was used to obtain the separation of stimuli-specific brain activity patterns using quasi-supervised learning. Different EEG profiles were constructed by using the EEG data using independent component analysis and the wavelet decomposition. Separation between the samples of the data was measured using quasi-supervised learning for different scenarios with constructed EEG profiles. Identification independent signals and derivation of the multi-scale signals increased the separation performance of the quasi-supervised learning. The most effective distinction between samples was obtained with the combination of independent component analysis and wavelet decomposition when independent component analysis was applied to wavelet decomposition-based EEG profiles.

For future work, the separation of the different EEG dataset (e.g. recorded under both visual and auditory stimuli) using quasi-supervised learning can be investigated. This provides the automation of determination stimuli-specific brain activity patterns and brain-computer interface applications.

REFERENCES

[1] S. Ogawa, T. M. Lee, A. S. Nayak, and P. Glynn, "Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields," *Magnetic Resonance in Medicine,* vol. 14, pp. 68-78, 1990.

[2] D. Cohen, "Magnetoencephalography: evidence of magnetic fields produced by alpha-rhythm currents," *Science,* vol. 161, p. 784, 1968.

[3] G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, and E. M. Stadlan, "Clinical diagnosis of Alzheimer's disease," *Neurology,* vol. 34, pp. 939-939, 1984.

[4] R. Abrams and M. A. Taylor, "Differential EEG patterns in affective disorder and schizophrenia," *Archives of General Psychiatry,* vol. 36, p. 1355, 1979.

[5] D. Giannitrapani and L. Kayton, "Schizophrenia and EEG spectral analysis," *Electroencephalography and clinical neurophysiology,* vol. 36, pp. 377-386, 1974.

[6] B. Karaçalı, "Quasi-supervised learning for biomedical data analysis," *Pattern Recognition,* vol. 43, pp. 3674-3682, 2010.

[7] W. Klimesch, "EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis," *Brain Research Reviews,* vol. 29, pp. 169-195, 1999.

[8] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent component analysis* vol. 26: Wiley-Interscience, 2001.

[9] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 11, pp. 674-693, 1989.

[10] R. Vigário, J. Sarela, V. Jousmiki, M. Hamalainen, and E. Oja, "Independent component approach to the analysis of EEG and MEG recordings," *Biomedical Engineering, IEEE Transactions on,* vol. 47, pp. 589-593, 2000.

[11] J. Lofhede, "The EEG of the Neonatal Brain-Classification of Background Activity," PHD, Division of Biomedical Engineering Department of Signals and Systems Chalmers University of Technology, 2009.

[12] J. A. C. Saeid Sanei, *EEG Signal Processing*. West Sussex, England: Wiley, 2007.

[13] K Charand. Available: http://hyperphysics.phy-astr.gsu.edu /hbase/biology/actpot.html (24.11.2011)

[14] B. Nielson, "Towards 24-7 Brain Mapping Technology," Informatics and Mathematic Modelling, Technical University of Denmark, 2009.

[15] T. F. Collura, "History and evolution of electroencephalographic instruments and techniques," *Journal of Clinical Neurophysiology*, vol. 10, pp. 476-504, 1993.

[16] A. Cichocki, S. L. Shishkin, T. Musha, Z. Leonowicz, T. Asada, and T. Kurachi, "EEG filtering based on blind source separation (BSS) for early detection of Alzheimer's disease," *Clinical Neurophysiology*, vol. 116, pp. 729-737, 2005.

[17] F. Vialatte, A. Cichocki, G. Dreyfus, T. Musha, S. Shishkin, and R. Gervais, "Early detection of Alzheimer's disease by blind source separation, time frequency representation, and bump modeling of EEG signals," *Artificial Neural Networks: Biological Inspirations–ICANN 2005*, pp. 683-692, 2005.

[18] J. Osselton, "Electroencephalographic monitoring in epilepsy," *Clinical Physics and Physiological Measurement*, vol. 12, p. 203, 1991.

[19] R. M. Rangayyan, *Biomedical signal analysis : a case-study approach*. Piscataway, NJ
New York, N.Y.: IEEE Press ;
Wiley-Interscience, 2002.

[20] BCI2000. Available:

http://www.bci2000.org/wiki/index.php/User_Tutorial:EEG_Measurement_Setup (15.10.2011)

[21] S. Makeig, A. J. Bell, T. P. Jung, and T. J. Sejnowski, "Independent component analysis of electroencephalographic data," *Advances in neural information processing systems*, pp. 145-151, 1996.

[22] L. T. DeCarlo, "On the meaning and use of kurtosis," *Psychological Methods*, vol. 2, p. 292, 1997.

[23] P. J. Huber, "Projection pursuit," *The annals of Statistics*, pp. 435-475, 1985.

[24] A. Papoulis and R. V. Probability, "Stochastic Processes, 1991," ed: McGraw-Hill.

[25] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *Information Theory, IEEE Transactions on*, vol. 36, pp. 961-1005, 1990.

[26] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *Image Processing, IEEE Transactions on,* vol. 1, pp. 205-220, 1992.

[27] A. S. Lewis and G. Knowles, "Image compression using the 2-D wavelet transform," *Image Processing, IEEE Transactions on,* vol. 1, pp. 244-250, 1992.

[28] S. Santoso, E. J. Powers, and W. Grady, "Power quality disturbance data compression using wavelet transform methods," *Power Delivery, IEEE Transactions on,* vol. 12, pp. 1250-1257, 1997.

[29] N. Hazarika, J. Z. Chen, A. C. Tsoi, and A. Sergejew, "Classification of EEG signals using the wavelet transform," *Signal Processing,* vol. 59, pp. 61-72, 1997.

[30] M. M. Shaker, "EEG waves classifier using wavelet transform and Fourier transform," *brain,* vol. 2, p. 3.

[31] U. Hoffmann, J. M. Vesin, T. Ebrahimi, and K. Diserens, "An efficient P300-based brain-computer interface for disabled subjects," *Journal of Neuroscience methods,* vol. 167, pp. 115-125, 2008.

[32] M. M. S. P. G. o. É. P. F. d. Lausanne. Available: http://mmspg.epfl.ch/page-58322-en.html (15.10.2011)