EEG-Based Human Interface for Disabled Individuals: Emotion Expression with Neural Networks

THESIS
Submitted for the Master Degree

By
Antoine Choppin
(98M53014)

Under the Supervision of
Professor Dr. Yukio Kosugi

Department of Information Processing
Interdisciplinary Graduate School of Science and Engineering

TOKYO INSTITUTE OF TECHNOLOGY
Yokohama, Japan

August, 2000

Contents

Chapter 1

Introduction

1.1 EEG-based interface for disabled individuals

Motivation

Amyotrophic lateral sclerosis (ALS) is a devastating neuromuscular disease that strikes adults in the prime of life. ALS attacks motor neurons which control the movement of voluntary muscles, and progresses rapidly, leading to complete paralysis followed by death in 3 to 5 years [54]. To date, no effective treatement exist for curing ALS and people affected by the malady (about 5,000 cases diagnosed annually in the United States) are condemned to progressive paralysis and speech loss.

> Many people with ALS have written that losing the power of speech is the worst part of the disease. Speech is something most people take for granted, and it's a vital tool that keeps us connected with loved ones, friends and co-workers [55].

Several systems exist to compensate for the communication impairment, but most of these require a minimal movement ability in order to function. Hence they become unusable in the later stages of the disease, where the need for communication is probably the most important.

Brain-computer interfaces (BCI) may be the answer to this problem.

Brain-Computer Interfaces (BCI)

BCIs are systems which analyze the electroencephalogram (EEG) of the patient in order to allow communication with the outer world. BCIs may serve various purposes essentially falling in one of two categories: control of a wheelchair or a prothesis, and verbal communication.

The design of a BCI dedicated to locomotion or movement have been studied by several research groups [36, 63, 18]. More general systems, allowing for instance to move a cursor on a computer screen have also been proposed [67, 39, 50]. A device specially designed for verbal communication was developed in [42, 4].

A common characteristic of these systems is that they are based on a *voluntary mental command* initiated by the user. This command may be realized in different ways, including control of slow cortical potentials (SCG) [42, 4], control of the mu-rhythm (sensorimotor cortex) [50, 39], realization of particular mental tasks [18, 36] and imagination of movements [63]. In other words, we can say that these BCIs are all based on the analysis of *"controlled EEG"*, no attention being focused on *"uncontrolled EEG"*.

1.2 Emotion expression with neural networks

The "broken link"

> When facial muscles are paralyzed, it may be hard to "read" a person's feelings in his face, as you may be used to doing. You may have to ask the person how he is feeling or work out some other mood-indicating signal [55].

This research focuses on the development of a new type of BCI, for the purpose of helping a patient to *express emotion*.

As will be discussed in chapter 2, emotion perception may lead to emotion generation and emotional experience may in turn trigger emotional behavior or expression (see figure 1.1). Patients suffering from ALS or any similar disease actually generate emotion and can feel emotion but become unable to express there feelings.

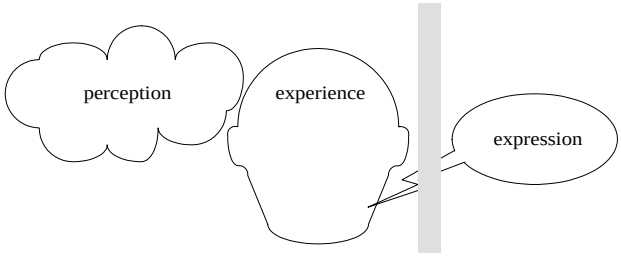


Figure 1.1: The "broken link".

The human-interface developed here is significantly different from conventional BCIs in the sense that it is based on the analysis of *"uncontrolled EEG"*. Our objective is to develop a system capable of detecting emotional experience from the EEG of the patient and to substitute for the "broken link" between emotional experience and expression.

Neural networks learn to recognize emotions

Desired characteristics for an effective emotion expressing human interface are capacity to adapt to the user, robustness and rapidity. We need a system capable of learning the complex relationship between EEG and emotion, capable to generalize and to adapt to new situations.

Artificial neural networks fulfill these requirements, thanks to their valuable characteristics: capacity to learn complex mappings from examples, non-linear representational power, and generalization ability.

Chapter 5 shows how neural networks are capable to learn the non-linear mapping between EEG features and emotion. It is also demonstrated that simple neural networks can be trained even with very few examples and yield an estimate of the emotional state of the patient.

Outline

This research is at the confluence of two important theoretical fields: brain science and information processing. The electroencephalogram is the link between these fields, as it constitutes a direct channel between the brain and the outside world.

This thesis is organized as follows. In chapter 2, emotion is studied in the light of neurophysiology and neuropsychology. Brain structures involved in emotion are identified and a

circuit of emotion is summarized. Chapter 3 concerns EEG digital signal processing. Techniques used for EEG recording, noise minimization and spectral feature extraction are presented. Experiments are reported in chapter 4. In a preliminary experiment, the influence of emotional stress on the EEG was studied. The main experiment allowed to observe the effect of emotion on the EEG and select features related to emotional experience. In chapter 5, neural network are shown to allow nonlinearly mapping EEG features onto emotion. Various issues are considered, including training, generalization and model selection. Finally, chapter 6 presents an online emotion expressing system integrating studied techniques.

Chapter 2

Emotion

"What is an Emotion ?" — This question was addressed by William James, more than a century ago. Although no complete answer is yet available today, we know much more about physiological and psychological aspects of emotion. On the one hand, brain research has progressed prodigiously the last decades, providing a better understanding of neurophysiological mechanisms underlying the generation of emotion. On the other hand, psychological theories have developped, allowing more precise definitions and useful representations of emotion. The first part of this chapter is devoted to the investigation of brain substrates of emotion. In the second part, psychological theories of emotions are considered, in order to clarify our approach.

2.1 Neurophysiological model of emotion

Neural mechanism of emotion can be described as a multi-layer system, in which each layer participates to the emotion process at a particular level of refinement. From an evolutionary perspective [20], at least three layers can be identified, ranging from the most primitive to the most elaborated: the limbic system, paralimbic association cortices, and the neocortex.

2.1.1 Limbic system

Buried within the depths of the brain, the limbic system is an ontogenically very ancient structure [34]. It is composed of several interconnected components including the hypothalamus, the amygdala, and the hippocampal system (Figure 2.1).

Hypothalamus: *the "motive force" of emotion*

The hypothalamus takes part to the emotion process at a very basic level. It is involved in homeostasis (endocrine, hormonal, visceral, and autonomic regulation) as well as in elementary emotions like hunger, thirst, pleasure, rage or aversion. It triggers very primitive, diffuse, undirected and unrefined emotions, which makes one "feel" happy or unhappy [34]. One component of the hypothalamus is the mamilliary body, through which it interconnects with the cerebral cortex. On the one hand, it receives afferent connections essentially coming from the fornix (itself connected to hippocampus, and to the cortex of the gyrus cinguli), and on the other hand, it sends efferent connections to the cortex along the mamillothalamic tract[61]. As a result of these strong interconnections, the rough feelings emanating from the hypothalamus appear to be under the inhibitory influence of higher order limbic nuclei (such as the amygdala and septum) [34] and of the cortex itself.

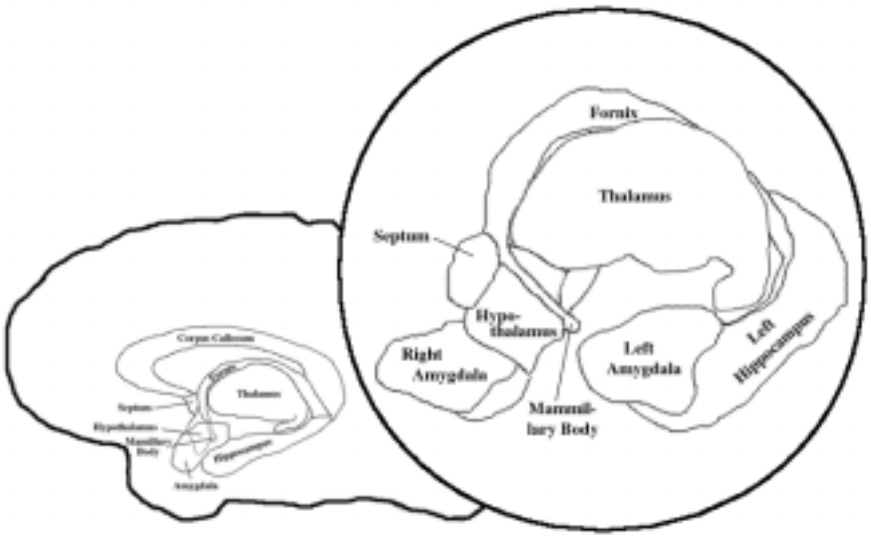


Figure 2.1: The limbic system

Hippocampus and septal nuclei: ***arousal, memory and behavioral inhibition***

The hippocampus connected through the fornix to the septal nuclei (or septum) participates in both cognitive and emotional functions [20]. Maintaining strong interconnections with the amygdala, it complements it in regard to attention, arousal, as well as in learning and memory. In conjuction with the hippocampus, the main role of the septal nuclei is to quite and dampen emotional drives emanating from other limbic components, couterbalancing their effect [34]. The "septohippocampal system" has also been termed as "behavioral inhibition system", as it inhibits motor behavior and increases arousal when receiving signals predicting non-reward or frustration [20].

Amygdala: ***emotional significance of external stimuli***

Although it was not a part of early circuits of emotion (e.g. Papez [61]), the amygdala is now known to be an essential component of the emotional system. In contrast with the hypothalamus, the amygdala is involved in higher order emotional activities. Because of its rich interconnections with the temporal and parietal cortices, it is a point of convergence of various sensory exteroceptive (auditory, somesthetic, visual), as well as interoceptive information. An important role of the amygdala is to determine the emotional significance of external events. In this context, the amygdala is involved in emotions like fear, anger and anxiety. It is probably also involved in the maintenance of mood states, by its action on the hypothalamus [34]. Maintaining direct two-ways connections with the cortex, the amygdala is capable of emotionally react to highly processed inputs, and to rapidly modulate subsequent cognitive processes in the cortex [20].

2.1.2 Paralimbic association cortices

Paralimbic cortices constitute the second layer of the emotional circuit. Figure 2.2 shows the location of paralimbic cortices, directly surrounding the limbic system and covered by the neocortex. Paralimbic cortices are mostly association cortices. They can be divided into two

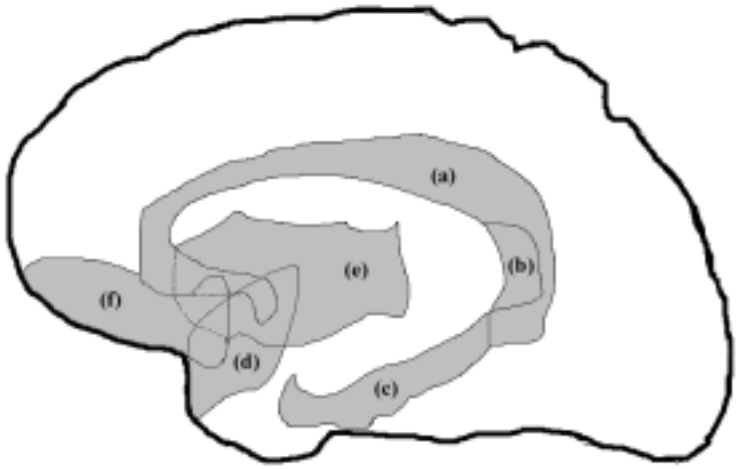


Figure 2.2: Paralimbic cortices constitute the second layer of the emotional circuit. They include (a) the cigulate gyrus, (b) retrosplenial cortex, (c) parahippocampal cortex, (d) temporal lobe, (e) insular cortex and (f) orbitofrontal cortex.

classes [20]: (1) cortices centered around the hippocampus, including the cingular, retrosplenial and parahippocampal cortices, and (2) cortices centered around the amygdala, including the temporal lobe, insular cortex and orbitofrontal cortex.

Paralimbic cortices centered around the hippocampus

The *cingulate cortex*, also called (cortex of the) gyrus cinguli, borders the corpus callosum on the median side of each hemisphere. Papez already pointed out its involvement in the mechanism of emotion: "the gyrus cinguli is the seat of dynamic vigilance by which environmental experiences are endowed with an emotional consciousness" [61]. Derryberry and Tucker [20] report the involvement of both the cingulate cortex and the *parahippocampal cortex* in spatial and motivational functions.

The *retrosplenial cortex* is included in the posterior cingulate cortex. Several neuroimaging studies have shown retrosplenial cortex activation associated with emotionally salient events. It may however be more closely related with emotion perception than with the generation of emotional responses [48].

Paralimbic cortices centered around the amygdala

The *temporal pole* (located at the anterior tip of the temporal lobe) maintains rich connections with the amygdala, and is critically involved in fear and anxiety [20] and in panic attack disorders [35].

The *insular cortex*, not visible on the surface of the brain, occupies the medial wall of the lateral sulcus. It is primarily involved in processing interoceptive information, and possesses for this purpose a topographic primary visceral sensory area. Like the amygdala, the insular cortex is a highly polymodal region, but its representations are more flexible [20]. The insular cortex includes a major zone dedicated to gustatory processing, which is probably related to its activation when processing disgust faces [15].

The *orbitofrontal cortex* is located in the medial and ventral surfaces of the frontal lobe. It is concerned with general arousal reaction [35], as well as with the coordination of exteroceptive

and interoceptive domains, in order to correct responses as conditions change [20].

2.1.3 Neocortex

The neocortex has shown to play a critical role in the interpretation of many stimuli that induce emotional experience [25]. Neocortical emotional processing finds place at a very high level, and is mixed with other cognitive processes. Individual differences are also particularly present at that level. The neocortex can be subdivided in two complementary ways: into cortical lobes, and laterally into left and right hemispheres.

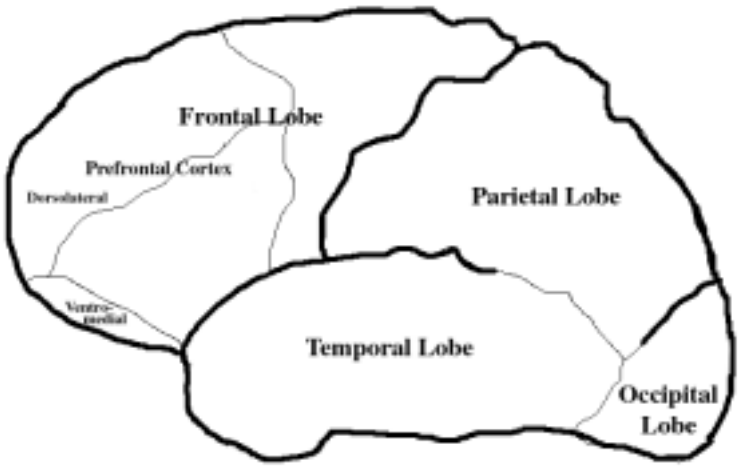


Figure 2.3: Cortical lobes.

Cortical lobes

Frontal lobe. It is widely recognized that the frontal region plays a role in cognitive behavior and motor planning. Particularly thought to be involved in emotion, is the prefrontal association cortex (PFC). The PFC can be subdivided into the dorsolateral sector and the ventromedial sector (see figure 2.3), which coincides with the orbitofrontal association cortex (see 2.1.2).

The dorsolateral PFC is probably involved in the representation of goal states toward which elementary positive or negative states are directed. The ventromedial PFC is probably involved in the representation of positive and negative states in the absence of immediate stimuli. This rests among others on the evidence that patients with bilateral lesions of the ventromedial PFC cannot anticipate future positive or negative consequence of their actions. In other words, the PFC plays the role of an "affective working memory" [15]. The dorsal frontal sector maintains strong interconnections with the posterior parietal lobe and cingulate gyrus, with which it mediates spatial and motivational information processing [71].

Finally, as pointed by Stuss et al. [71], the role played by the frontal cortex in motivation, organization and integration, and motor expression makes it extremely important with regard to the expression of personality.

Temporal lobe. It is in the temporal lobe that lies the primary auditory cortex. It is also involved in memory related functions. However, besides its involvement in cognitive functions,

the temporal lobe plays a major role in affective and emotional behavior. In particular, the temporal lobe seems to be involved in speech prosody (affective component of language, consisting in the intonation of speech). Observations in patients with temporal lobe epilepsy also suggest that it is involved in personality, and in sexual and social behavior [35].

Parietal lobe. Various sensory informations are processed in the parietal lobe, including somatic and visual sensory information. The parietal lobe is particularly involved in high-level sensory functions, as well as in language [35]. Perceptual and conceptual processing performed in the parietal cortex is modulated by projections emanating from the paralimbic association cortices [20]. The inferior parietal lobe seems to play an important role in attention and arousal [25]. It is in this polymodal association cortex that the meaning of a collection of modality-specific inputs is determined. The significance of these sensory inputs is then determined in the frontal cortex, with respect to motivation and long-term goals (the inferior parietal lobe is indeed strongly interconnected not only with the limbic system, but also with the frontal cortex). Moreover, Denny-Brown and Chambers [19] suggest that the parietal lobe be involved in approach behavior, based among other on ablation studies and on the observation of patients with parietal lesions.

Occipital lobe. The role played by the occipital lobe in emotion essentially concerns the early modulation of visual sensory information, in the light of emotional information generated by the limbic system and other cortical areas. For example, backward projections from the temporal pole may facilitate the early object-processing in terms of the current motivational state [20]. Besides, Lang et al. [45] report results of an fMRI analysis where increased activation of the visual cortex was observed when viewing emotional pictures.

Hemispheric differences

Hemispheric specialization[1] or lateralization is observable in emotional processing, but not as clearly defined as in cognitive functions [6]. There are essentially two kinds of theories accounting for hemispheric specialization in emotion: the *right hemisphere* theories, hypothesizing the right hemisphere to be dominant in emotional processing; the *valence* theories, according to which the left hemisphere would be associated with positive emotions (approach-related behaviors), and the right hemisphere with negative emotions (withdrawal-related behaviors). Importantly, many observations showing the dominance of the right hemisphere concern emotional perception (particularly spatial), rather than emotional experience [e.g. 7]. Besides, some opinions vary among valence theories, regarding to whether positive (resp. negative) emotion is associated to approach (resp. withdrawal) behavior [81], or not [25].

2.1.4 Circuit of emotion

How the components of the limbic system interact with paralimbic cortices and with the neocortex to process and generate emotion still remains not completely understood. However, several propositions have been made for a possible "circuit of emotion". Early proposals were made by Papez [61], and later by MacLean [47]. More recent studies gave new insights in the mechanisms of emotion and particularly in the important role of higher level neocortical brain regions.

[1] Hemispheric specialization refers to the fact that certain higher functions are differentially represented in the two hemispheres.

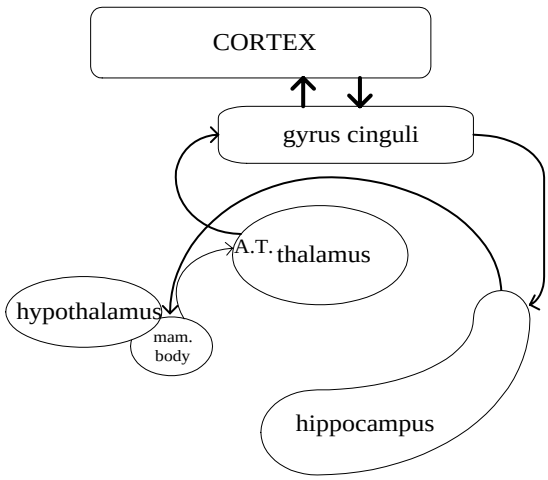


Figure 2.4: Papez circuit of emotion.

Papez circuit

The circuit proposed by Papez [61] is depicted in figure 2.4. It can be summarized as a two-way connection between the hypothalamus and the cingulate gyrus. According to Papez, emotion can arise in two ways: as a result of psychic activity and as a consequence of hypothalamic activity. In the first case, emotion is triggered by high-level processes, coming from the cortex of the cingulate gyrus, descending through the hippocampus and the fornix and acting on the hypothalamus, which in turn acts on various efferent centers (e.g. the endocrine system). In the latter case, emotion is triggered by the hypothalamus itself, and sent through the mamillothalamic tract and the anterior thalamic nuclei to the cortex, where emotional information influences higher-level cognitive processes.

Integrated view of emotion

Among the earliest additions to Papez circuit was the inclusion of the amygdala, of which the importance in determining the significance of incoming sensory inputs has been described above. Papez circuit lacked this "eye turned outward", as termed by Joseph [34], allowing emotions to be generated in response to sensory information. The amygdala probably also accounts for various emotional coloring that Papez had trouble to explain.

In addition to the important role of the cingulate gyrus in allowing low-level emotional information to be relayed to higher level centers, other association cortices have been presented in section 2.1.2. The role of paralimbic association cortices with regard to emotion could be summarized as follows: to refine rough emotional information coming from deep limbic nodes, with regard to various aspect of cognition (attention, motivation, reaction to input stimuli[2],...). Rough and extreme emotional states (such as rage, fear or sexual activity), which have been shown to be possibly obtained just by electrically stimulating limbic components (like the hypothalamus or the amygdala) are "softened" by the intermediate layer of paralimbic cortices. This intermediate processing is however not a "forced way", since the amygdala maintains direct projections with neocortical areas, and is therefore capable in some cases to directly excert emotional influence on high-level cognitive processes.

[2]Paralimbic cortices, rather to work on a single-modality input, use polymodal information to emotionally respond to a given situation.

Information processing in the neocortex is both precisely localized and highly distributed. Most neocortical areas are hence dedicated to precise tasks, but all participate to the whole brain processing. In the same way, when a particular emotion is triggered, the whole neocortex participates to reacting appropriately (avoidance, withdrawal in case of a negative emotion, and approach, persistence in case of a positive emotion). The action of various neocortical centers is essentially cooperative: sensory cortices integrate the emotional significance of a situation, while the frontal cortex acts as to generate an appropriate motor response.

2.2 Psychological views of emotion

Emotion is very broad concept and it is now time to look at how one can precise what is meant in the context of this study. There are obviously various kinds of emotions, but how can we identify and classify them ?

2.2.1 Emotional perception, experience and behavior

A first useful distinction concerning the meaning of "emotion" was already pointed out by Papez [61]:

> The term "emotion" as commonly used implies two conditions: a way of acting and a way of feeling. The former is designated as *emotional expression*; the latter, as *emotional experience* or subjective feeling[3].

Besides, Davidson [13] mentions another important subdivision of emotion, namely the difference between emotional perception and emotional experience. Perceiving emotional information is problably something different than actually experiencing emotion (although the perception of an emotional stimulus can of course trigger emotional experience). Still another possible subdivision [24] is to distinguish between emotional behavior, emotional communication (through words, prosody, gestures) and emotional feeling or experience. However, I prefer to consider the first two categories as included in emotional behavior.

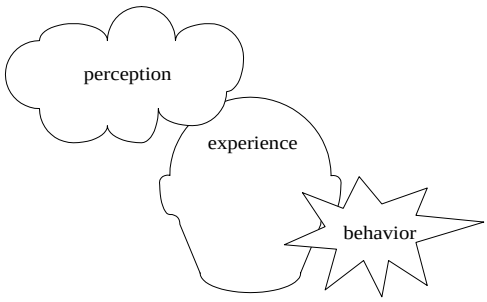


Figure 2.5: Three components of emotion: perception, experience and behavior.

Figure 2.5 summarizes a possible subdivision of emotion, in the form of three successive "steps": emotion perception, emotional experience and emotional behavior. These steps often occur in series, although this is not necessary the case. First, an information is processed and its emotional content is assessed. In case of an emotionally colored stimulus, it may trigger

[3]Italic added.

emotional experience, the actual feeling of emotion. If this feeling is strong and if other conditions are met, this emotion will probably be expressed and yield emotional behavior (be it a way of acting –e.g. crying– or any form of communication). Of course, this scheme is oversimplified: not all emotional stimuli will lead to emotional experience, nor will emotional experience always be exteriorized through behavior. Furthermore, emotional experience can happen without the presence of any exterior stimulus.

The present research focuses on the expression of felt emotions. For patients suffering from ALS or any other disease impairing communication with the outer world, the link between emotional experience and emotional expression or behavior is cut: although such patients are able to process stimuli from various sensory modalities, to analyze their emotional significance, and to actually experience an emotional feeling, the expression of this feeling is made impossible by the malady. An important focus of the present research is to detect emotional experience in order to replace the lacking function of emotional expression.

2.2.2 Representation of emotion

Focusing now on experienced emotions, an important issue is to look for an efficient way to represent them, since we are not only interested in detecting any emotion, but also in *what kind* of emotion it is. Two approaches have traditionally be used to classify emotions: basic emotions and dimensions (or components) of emotion.

Basic emotions

A first class of approaches to analyze emotion postulate the existance of a number of emotions called "basic", because they can be considered as "primitive" or "irreducible" either psychologically or biologically. Ekman is famous for belonging to the defenders of the basic emotions approach [21]. His argument for the existence of basic emotions is essentially based on physiology, and particularly on the existence of clearly distinct facial expressions associated with a particular feeling. Emotions considered as basic by Ekman, as well as by Friesen and Ellsworth are: anger, disgust, fear, joy, sadness, surprise. However, antagonists of this theory point out that other researchers supporting this view arrive to different conclusions regarding to what basic emotions are and which emotions are basic [76]. Furthermore, there are also some dissensions regarding to whether or not basic emotions are to be viewed as building blocks of more complex emotions.

Dimensional views of emotion

A second approach to emotion is based on the hypothesis that any particular emotion results from the combination of a number of "components" or "dimensions". Osgood et al. [59], using the semantic differential, found that most of the variance of verbal assessments of emotional judgments could be accounted for by three major dimensions: evaluation (pleasant/unpleasant), potency (dominant/dominated) and activity (excited/calm). Evaluation, activity and potency can alternatively be called valence, arousal and control (or dominance). Heilman [24] proposes a similar set of components: valence, arousal and motor activation (approach, avoid, neither). For Lang [43], all emotions can be located in a two-dimensional space, as coordinates of affective valence and arousal (in later studies he also considered dominance,

as a third dimension). For example, sadness is characterized by negative valence and low arousal.

Discussion

Some controversy arose between the partisans of both approaches [76, 21, 31], probably because these theories actually aim at describe *what are* emotions, rather than *how to represent* emotions. I will however not argue in that discussion, since my only interest in the context of this research is to find a suitable representation of emotions.

In this work I opt for a two-dimensional representation of emotions, in terms of valence and arousal. This choice was motivated mainly for two reasons:

- *Simplicity:* Any felt emotion can easily be expressed in terms of valence and arousal, while it is sometimes more difficult (and uncertain) to try to express it in terms of basic emotions[4].
- *Universality:* There is little (probably no) dissension regarding the first two dimensions against which emotions should be expressed: valence and arousal appear as natural candidates to express a basic feeling. The meaning of valence and arousal are universal, independent of any cultural factor.

An additional reason why I used this bi-dimensional representation of emotion is that it is widely used for ratings emotional stimuli like pictures or sounds (see description of the IAPS/IADS systems in section 4.2.2).

[4]It can be argued that the representation is not unique: some distinct emotional states are mapped on the same couple (valence, arousal). However, we are interested in a clear, simple and intuitive representation, rather than in precision or unicity. Self-assessment of emotion has been reported to be a difficult task (see chapter 4). For this reason, a simple representation system was used, to minimize ratings uncertainty.

Chapter 3

EEG Digital Signal Processing

This chapter aims at describing the methods I used in two experiments presented in chapter 4, as well as underlying theoretical foundations. In the first section, I explain how data acquisition and preprocessing were carried out. Methods used for extracting spectral features from the recorded signal are presented in section 3.2.

3.1 Acquisition and preprocessing

3.1.1 The electroencephalogram

The electroencephalogram (EEG) is the electric potential recorded at the surface of the scalp, resulting from the electrical activity of large ensembles of neurons in the brain. The non-invasive brain imaging technique based on EEG recording is called electroencephalography. Figure 3.1 (adapted from Kandel et al. [35]) illustrates how the EEG is related to the electrical activity of groups of neurons.

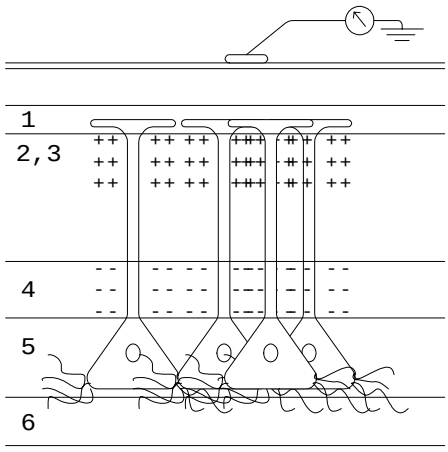


Figure 3.1: The electroencephalogram (EEG) results from the electrical activity of large ensembles of neurons in the brain.

Although the EEG reflects neuronal electrical activity, the electric potential measured at the surface of the scalp is extremely different from the one that would be recorded directly at the cortex. This is because of the presence of several intermediate layers (essentially cerebrospinal fluid (CSF), skull and fat), which are at the origin of both attenuation and volume

conduction. Attenuation can be as high as 5000:1 [10], although typical values are generally much lower. Volume conduction causes a general "smoothing" of the electrical activity over the whole head surface.

Typical EEG is characterized by amplitudes ranging from about 10 to 100μV and frequencies from DC to 60Hz. A few dominant frequency bands are generally observed: alpha (8–13Hz), beta (13–30Hz), delta (0.5–4Hz) and theta (4–7Hz). Each of these frequency bands is associated to particular brain rhythms of which the generating mechanisms are not yet completely understood.

3.1.2 EEG acquisition

The EEG acquisition system used in this research is depicted in figure 3.2.

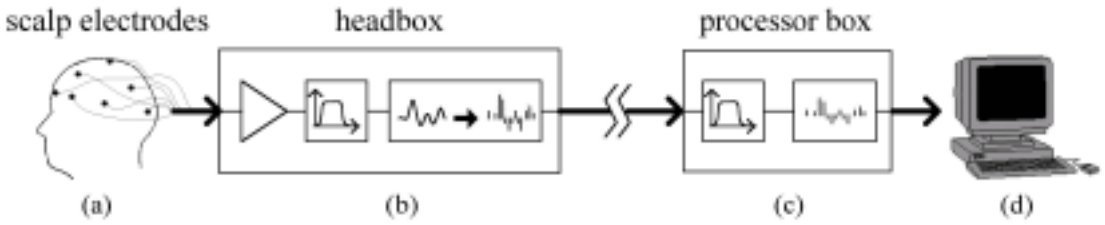


Figure 3.2: EEG acquisition system.

The system consists of the following successive elements:

- *Scalp electrodes* are used to measure the electric potential at the surface of the head (electrode montage is discussed below).
- A *headbox* amplifies, filters (HP filter at 0.53Hz and anti-aliazing LP filter at 250Hz) and samples the analog EEG multichannel signal at 1kHz.
- A *processor box* is responsible for collecting the digital signal coming from the headbox, possibly applying digital filters, and performing decimation (re-sampling).
- The EEG is recorded by a *personal computer*.

Measurements were carried out with a Digital Bio-Amplifier 5200 (NF Electronic Corp.), consisting of a EEG Headbox 5202 (14 EEG channels, sensitivity: 409.6μVpp/FS, 2 EMG channels, sensitivity: 10.24mVpp/FS, time constant: 0.3s) and a Processor Box 5201 (IIR digital filtering, decimation).

In the main experiment, all measurements were carried out inside a shield room (Nihon Itagarasu Kankyo Ameniti NEA Corp., model "Mag Savor 15") ensuring an attenuation of 60dB for electric waves (0.2∼18MHz) and of 40dB for magnetic waves (0.2∼1.9MHz). Both the processor box (c) and the recording computer (d) were located outside the shield room, where the digital signal was collected via insulated connectors.

Electrode montage

Electrodes were placed according to the "International 10–20 System" shown in figure 3.3.

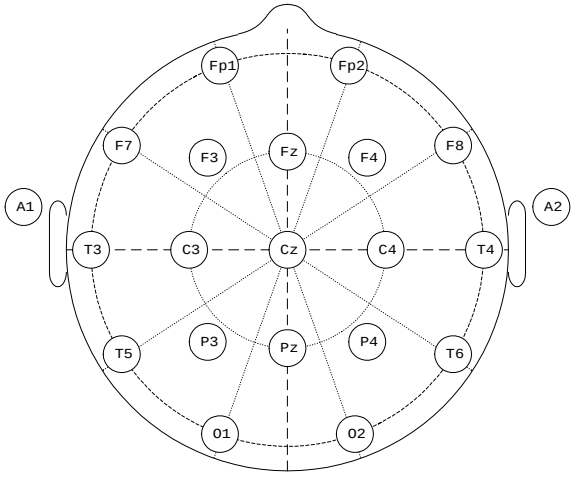


Figure 3.3: The "International 10–20 System" for electrode placement.

The ground electrode was placed at the vertex Cz, and all channels were referenced to the left earlobe (A1). The choice of an appropriate reference is an important issue, especially with respect to inter-electrode coherence [58]. Ideally, one should use a reference located at an infinite distance of recording electrodes. In practice, several options are possible. Single ear reference is a common option, which I adopted in the first experiment (section 4.1). In the second experiment (section 4.2), I used a "digitally linked ears reference", that is, recordings were referenced at the left ear V_{A1}, but the right ear potential (referenced to the left ear) was also recorded: $(V_{A2} - V_{A1})$. Subsequently, half of this quantity was subtracted from each single referenced channel V_j:

$$(V_j - V_{A1}) - \frac{1}{2}(V_{A2} - V_{A1}) = V_j - \frac{1}{2}(V_{A1} + V_{A2}),$$

which results in referencing each channel to an averaged reference $V_A = 1/2(V_{A1} + V_{A2})$. Note that this is in general different from a "physically linked ears reference", for which the reference potential on the wire connected to the two ears is:

$$V'_A = \frac{R_2 V_{A1} + R_1 V_{A2}}{R_1 + R_2},$$

where R_1 and R_2 are the two ear electrode contact impedances. V'_A can approximate V_A only if R_1 and R_2 are nearly equal [58].

3.1.3 Artifact minimization

Because of the small amplitude of electric potentials measured at the surface of the scalp, EEG recording is very easily influenced by any other electric signal. Fluctuations in the recorded signal that do not result from brain activity are called *artifacts*. Artifacts affecting EEG recording belong two categories: artifacts caused by the subject and artifacts not related to the subject [72].

Artifacts not caused by the subject essentially result either from the mixture of external parasitic signals with scalp potentials, or from variation in the subject environment. Parasitic electrical signals may result, among other, from static electricity or from electromagnetic

radiation produced by surrounding appliances. In the main experiment, artifacts exterior to the subject were almost inexistant, because all EEG measurements were performed in a shield room blocking electromagnetic radiation from outside. The subject was alone in the shield room, excluding possible artifacts due to other people movements[1].

Artifacts caused by the subject are of two types: artifacts provoked by subject's movements and artifacts provoked by biological electric phenomena. In order to reduce artifacts caused by movements (gestures, respiration, etc.), it is important that the subject be relaxed and asked not to move during the experiment. The experiments always consisted in short recording phases interleaved with rest phases where the subject was allowed to move and relax. Biological phenomena that cause EEG artifacts are: eye movements, muscle activity, cardiac beat, and sweat [72]. In the following, I discuss the most important: eye movement and muscle artifacts.

Eye movement artifacts

Eye movement artifacts (or ocular artifacts) result from the contamination of the EEG by the electrooculogram (EOG), a potential produced by movement of the eye or eyelid. Several methods have been proposed for removing ocular artifacts from the EEG, most of which make use of a separate EOG record. In the experiments, I used the simple EOG recording scheme depicted in figure 3.4. With this montage two EOG channels are recorded, related to vertical and horizontal eye movements (these channels are denoted by EOG_V and EOG_H in the following).

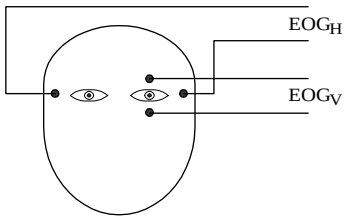


Figure 3.4: Electrooculogram electrode placement.

Classical methods for removing EOG artifacts can be classified into rejection methods and subtraction methods [32]. Rejection methods simply consist in discarding contaminated EEG, based on either automatic or visual detection. There success, however, crucially depends on the quality of the detection. Moreover, this can lead to an unacceptable loss of data. Subtraction methods are based on the assumption that the measured EEG is a linear combination of background (actual) EEG and EOG. The original EEG is hence recovered by subtracting separately recorded EOG from the measured EEG, using appropriate weights (reflecting the influence of the EOG on particular EEG channels). More recently, new methods for artifact minimization have been proposed, based on the concept of blind source separation (BSS) [78].

In the following, two methods used in the experiments are presented. The first one is a classical subtraction method, which I used in the main experiment because it allows on-line EOG minimization. The second method uses a recently developed BSS technique called independent component analysis (ICA). This technique was used successfully in the preliminary experiment, for off-line EOG minimization.

[1] In a preliminary experiment, although recordings were performed in a large room, with almost no electrical appliances in the direct surrounding of the subject, subsequent power line noise removal was needed (HAM filtering).

Recursive least squares

The classical least square method for EOG subtraction relies on the assumption that each measured EEG channel y_k is a linear combination of background EEG[2] e_k and EOG channels x_j ($j \in \{V, H\}$):

$$y_k(i) = \sum_{j \in \{V,H\}} \theta_{k,j} x_j(i) + e_k(i) \qquad i = 1, 2, \ldots, T, \;\; k = 1, 2, \ldots, n.$$

Here, T is the length of the measured signals (number of samples). This expression can be re-written in matrix form:

$$Y = \Theta X + E, \tag{3.1}$$

where Y, Θ and X are $(n \times T)$, $(n \times 2)$ and $(2 \times T)$ matrices, respectively (n is the number of EEG channels and 2 is the number of EOG channels). In this method, an appoximation $\hat{\Theta}$ of the Θ is sought, so that the original EEG can be estimated as

$$\hat{E} = Y - \hat{\Theta} X. \tag{3.2}$$

Equation 3.2 can be solved by minimizing the sum of squared background EEG:

$$\hat{\Theta} = Y \cdot (X^T X)^{-1} X^T = Y \cdot X^+ \tag{3.3}$$

where X^+ is the pseudoinverse of X.

Ifeachor et al. [30] proposed an online version of this technique (based on recursive least square (RLS) approximation), which allows to adapt the coefficients Θ each time a new sample $\mathbf{y}(i) = [y_1(i) \ldots y_n(i)]^T$ (EEG) and $\mathbf{x}(i) = [x_V(i)\ x_H(i)]^T$ (EOG) becomes available.

Note that, as the EOG propagates through the scalp and intermixes with the EEG, the inverse is also true. Because the recorded EOG is susceptible to contain EEG as well, EOG was low-pass filtered with a cut-off frequency of about 15Hz, in order to remove high frequency EEG and noise.

A possible weakness of LS based methods is that they become less reliable if the quantity of EOG is very low. In that case, E cannot anymore be considered as noise. In general, however, this method yields more than acceptable results. I used LS subtraction in the main experiment because it is (1) simple and efficient, (2) completely automatic (requires no human intervention), (3) implementable online (e.g. in an emotion expressing system like the one described in chapter 6).

Independent component analysis

Independent component analysis (ICA) is a relatively recent method for blind source separation (BSS), which has shown to outperform the classical principal component analysis (PCA) in many applications. In particular, it has been applied for the extraction of ocular artifacts from the EEG [79].

ICA assumes the existence of n signals that are linear mixtures of m unknown independent source signals. At time instant i, the observed n-dimensional data vector $\mathbf{x}(i) = [x_1(i) \ldots x_n(i)]^T$ is given by the model [78]

$$\mathbf{x}(i) = \sum_{j=1}^{m} \mathbf{a}_j s_j(i) = \mathbf{A}\mathbf{s}(i), \tag{3.4}$$

[2]Justification for the term "background EEG" is that the EOG amplitude is substantially higher than the EEG (about 5 to 10 times), so that the actual EEG can be considered as background "noise".

where both the independent source signals $s_1(i), \ldots, s_m(i)$ and the mixing matrix $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_m]$ are unknown. Other conditions for the existence of a solution are (1) $n \geq m$ (i.e. there are at least as many mixtures as the number of independent sources), and (2) up to one source may be Gaussian. Under these assumptions, the ICA seeks a solution of the form

$$\hat{\mathbf{s}}(i) = \mathbf{B}\mathbf{x}(i), \tag{3.5}$$

where $\mathbf{B}$ is called the separating matrix.

FastICA is a fixed-point algorithm for ICA, proposed by Hyvärinen and Oja [29]. It solves the problem presented above in two steps:

1. Whitening: this is basically a PCA projection, which transforms mixtures $\mathbf{x}(i)$ into uncorrelated (or *white*) vectors $\mathbf{v}(k)$, with unit variance.

2. Fixed-point computation of the *orthogonal* separation matrix $\mathbf{W}$ of pre-whitened vectors: this is done by maximizing a criterion of non-Gaussianity, the fourth order cumulant (also called *kurtosis*) and defined as $\text{kurt}(s_j) = E\{s_j^4\} - 3[E\{s_i^2\}]^2$ ($E\{\cdot\}$ denotes mathematical expectation).

I used the FastICA algorithm (implemented in the FastICA package for MATLAB$^{\text{TM}}$ [28]) for extracting independent EOG components from the EEG. In order to make convergence faster and more robust, two separate EOG channels were recorded, as described above, which were used together with the recorded EEG channels, as mixture signals. As mentioned previously, EOG recordings are susceptible to contain a (very low amplitude) EEG component as well, which justifies the inclusion of EOG among the mixture signals $\mathbf{x}(i)$.

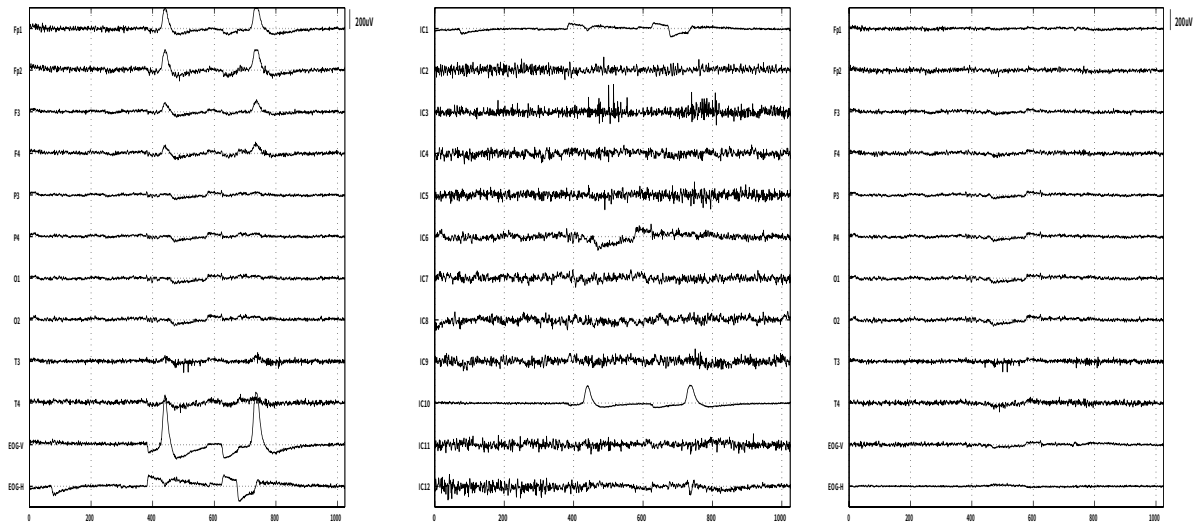
An example of EOG suppression using ICA is shown in figure 3.5 (Only 5.12 seconds are shown, of the 1 minute recording on which the algorithm has been applied). Recorded signals, including 10 EEG and 2 EOG channels (a) are used as inputs (FastICA 'deflation' approach was used, with 'tanh' nonlinearity; the algorithm converged in 42 steps). (b) shows independent components $s_1(i), \ldots, s_m(i)$, denoted by IC1 up to IC12. In almost all cases, the algorithm converged without problem, yielding two (sometimes three) independent components representing eye artifacts. Eye movement related ICs were selected manually and the signals reconstructed by excluding the selected ICs (c).

Visual inspection of the mixing matrix $\mathbf{A}$ can be done easily, as shown in figure 3.6. The upper plot shows the contribution of each IC to the vertical EOG channel (i.e. the 11th row of the separation matrix $\mathbf{A}$, $[a_{11,1} \ldots a_{11,n}]^T$.). Clearly, IC10 appears to mostly represent EOG_V [3]. Next, the lower plot shows the quantity of IC10 in each EEG channel (i.e. the 10th column of the separation matrix $\mathbf{A}$, $\mathbf{a}_{10}$). The highest quantity is of course in EOG_V itself. The frontopolar electrodes Fp1 and Fp2 (located just above the eyes) are the most contaminated. The quantity of IC10 then decreases whith the distance to the front side of the head (minimal for O1 and O2) [4].

These results illustrate the suitability of ICA for removing ocular artifacts from the EEG. There are however some unsolved issues, related to the possibility that ICA does not converge (this happened very rarely), or yields more EOG-related ICs than the number of EOG channels (e.g. extracting independent components from the EOG rather than separating EOG

[3] Also note the negative contribution of IC1, which happens to represent EOG_H, the horizontal EOG. This expresses that some horizontal EOG has been recorded by the vertical eye movement electrodes. ICA has successfully separated independent horizontal and vertical components.

[4] Note the slightly higher contribution of EOG_V to left electrodes Fp1, F3, P3, O1 and T3, compared to their right homologue. This is probably because EOG_V has been recorded at the left eye.



(a) (b) (c)

Figure 3.5: ICA for eye movement artifact minimization: (a) original EEG channels, (b) independent components, (c) EEG channels re-constructed without IC1 and IC10.

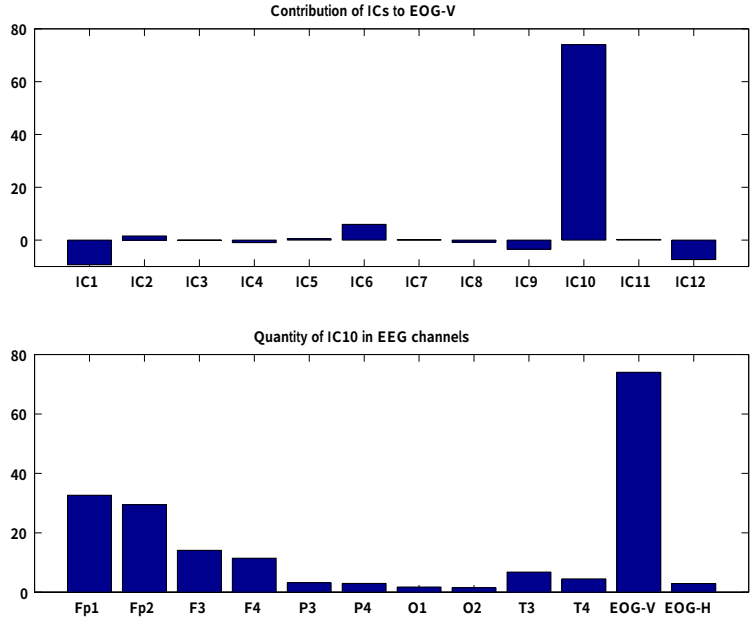


Figure 3.6: Selection of ICs related with eye movement artifacts, by visual inspection of the mixing matrix.

from EEG). The need for visual inspection is also still a problem, although if the quality of the convergence is satisfying, automatic (e.g. threshold based) IC selection can probably be performed.

Finally, let us mention that the use of ICA in EEG analysis is not restricted to artifact detection and that it can probably successfully used for other purposes (for instance, event related potential (ERP) extraction or source localization).

Muscle artifacts

The second type of artifact discussed here is due to muscle activity. Either conscious or unconscious muscle activity produces an electric potential called electromyogram (EMG). Muscle artifacts can be classified according to their spread in space (broad or localized) and time (transient or permanent) [72]. Compared to normal EEG activity, muscle artifact is characterized by high frequencies (over 15Hz) and often by high amplitude.

Unlike eye movement artifacts, muscle artifacts are especially difficult to remove from the recorded signal. Several approaches (like in [60]) are based on (possibly non-linear) digital filtering of measured signals. One possible approach is to filter out high frequency EEG (e.g. above 20Hz). This approach was used in the preliminary experiment.

Filtering high-frequency components is however undesirable as underlying EEG activity could be used in analysis. For this reason, the approach adopted in the main experiment was to detect and reject EEG contamined by muscle artifacts. van de Velde et al. [77] reviewed several criteria that can be used for muscle artifact detection:

- Time domain: slope differentiator, max/min threshold;
- Frequency domain: absolute and relative power over 25Hz (beta2), spectral edge frequency.

According to their results, I opted for absolute power over 25Hz for detecting EMG artifacts. The power was log-transformed to normalize the spectrum distribution and the threshold was set arbitrarily, on basis of visual inspection. 10-seconds EEG segments were tested against the rejection criterion. A further measure was taken, in order to properly reject permanent artifacts: for each EEG channel recorded in a given experiment, if more than 50% of the segments were marked for rejection, the whole channel was rejected. This allowed to apply more rigorous rejection for channels where EMG was continuous (opposed to burst muscle artifacts). Fronto-polar and temporal electrodes were the most sensitive for this test. Figure 3.7 shows a very short record (2.64 second) of a few EEG channels, where automatic rejections have been marked. In this case, three channels (Fp1, Fp2 and O2) are contamined by permanent muscle artifact (The percentage of contaminated segment of Fp1, Fp2 and O1 for this subject were 78%, 97% and 97% respectively; eight channels contained no muscle artifacts and three channels contained less than 25% contamined segments).

3.2 Feature extraction

The success of EEG analysis essentially relies on the quality and the relevance of the information extracted from raw records. This section presents the signal processing methods used for feature extraction.

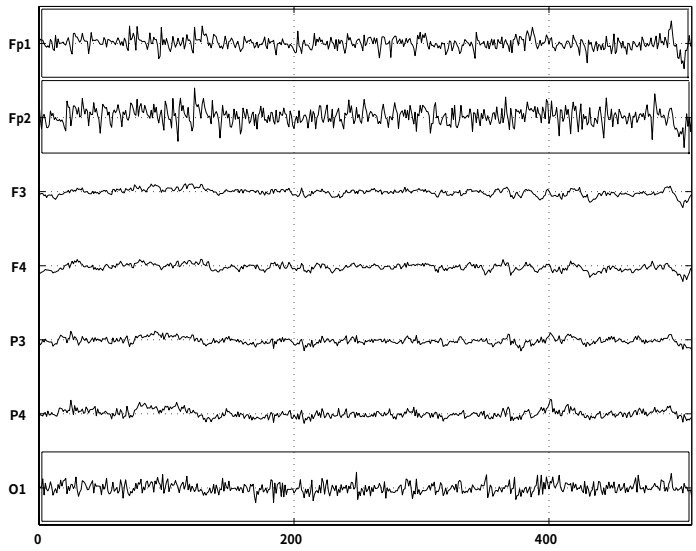


Figure 3.7: Automatic muscle artifact detection (channels marked for rejection are outlined).

3.2.1 EEG spectral estimation

The existence of particular brain rhythms underlying electroencephalographic activity naturally motivates the investigation of spectral characteristics of the EEG. Fourier analysis constitutes a theoretical base for spectral estimation of stationary processes. Still, can the EEG be considered as stationary ?

Time interval

If one looks at long records of EEG, clear changes can be observed and it would be erroneous to consider that statistical properties of the EEG do not change over time. In general, the EEG is thus not stationary. However, when reducing the time window over which one observes it, the EEG unsurprizingly tends to exhibit more stable statistical properties. For instance, many more different brain states are indeed likely to be reached in a day than within a minute. Similarly, fewer distinct emotions are likely to be experienced within a small time interval. In other terms, for the stationarity hypothesis to be acceptable, one needs to consider sufficiently small time interval.

On the other hand, another problem arises, from a practical point of view: for spectral estimation to be reliable (i.e. the variance of the estimate be small enough), a sufficiently long time interval is needed. In practice, Davidson [13] suggests to use a minimum of 10–15 seconds of EEG signals in order to obtain stable estimates.

In the preliminary experiment, the tasks performed by the subject (computer game playing, see section 4.1) were relatively long (1∼3 min). Each EEG record was therefore segmented into small slices of about 5 seconds, and spectral features were computed for each slice. A problem of this approach, however, is that long mental tasks usually involve a succession of various brain states. Successive slices may therefore have very different spectral characteristics, which makes it difficult to draw valid conclusions.

For this reason, the timing of the main experiment was designed as to obtain short (about 10 sec) distinct EEG epochs, each of which being associated to a well-defined brain state (i.e. emotion, see section 4.2). Spectral features were then computed for each epoch.

Power spectral density

The *power spectral density function* $h(\omega)$ of a stationary stochastic process $X(t)$ can be expressed[5] as the Fourier transform of its autocovariance function $R(\tau)$ [65]:

$$h(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega\tau} R(\tau)\, d\tau. \tag{3.6}$$

When it exists, $h(\omega)$ has the following interpretation,

$$\begin{aligned} h(\omega)\, d\omega \;\; = \;\; & \text{average (over all realizations) of the contribution} \\ & \text{to the total power from components in } X(t) \\ & \text{with frequencies between } \omega \text{ and } \omega + d\omega. \end{aligned} \tag{3.7}$$

In practice, there are essentially two restrictions: (1) we can only *observe* a particular realization X_t of $X(t)$, and (2) this observation is finite. $h(\omega)$ can therefore only be *estimated*. Methods for spectral estimation can be classified into parametric methods and FFT-based methods. In this research, I chose an FFT-based technique, basically for simplicity and efficiency reasons. Simple FFT-based methods allow on-line computation of spectral estimates of acceptable quality. Although the on-line estimation of simple (e.g. AR) models would also be possible, the quality of the estimation becomes questionable if a too simple model is used.

Periodogram estimation. Given N observations $X_1, X_2, \ldots, X_N$, the *(modified) periodogram*, or *sample spectral density function* of X_t is defined as follows:

$$I_N(\omega) = \frac{1}{2\pi N} \left| \sum_{t=1}^{N} X_t\, e^{i\omega t} \right|^2. \tag{3.8}$$

The periodogram, which is the square of the finite Fourier transform, can be shown to be an asymptotically unbiased estimate of $h(\omega)$ [65]:

$$\mathrm{E}\{I_N(\omega)\} = h(\omega) + O\left(\frac{\log N}{N}\right). \tag{3.9}$$

The periodogram itself is however an extremely poor estimate of the spectral density function, because (1) $\mathrm{var}\{I_N(\omega)\}$ does not tend to zero as $N \to \infty$, and (2) as a function of ω, $I_N(\omega)$ typically has an erratic and wildly fluctuating form [65].

Welch's averaged periodogram method. This method proposed by Welch [80] is illustrated in figure 3.8. The signal X_t is first broken into m short sections (a). A non rectangular data window is then applied to each section (b), before computing their periodogram (c). Obtained modified periodograms are finally averaged. Note that short sections can also be overlapped, in order to increase their number, and thereby reduce the variance of the estimate.

Key parameters of Welch's method are the window size and, possibly, the overlap size. A less crucial issue is the choice of the data window. An overlap of 50% was used systematically in the measurements, as it is commonly done. The data window was a Hamming-type window.

[5] provided it exists for all ω

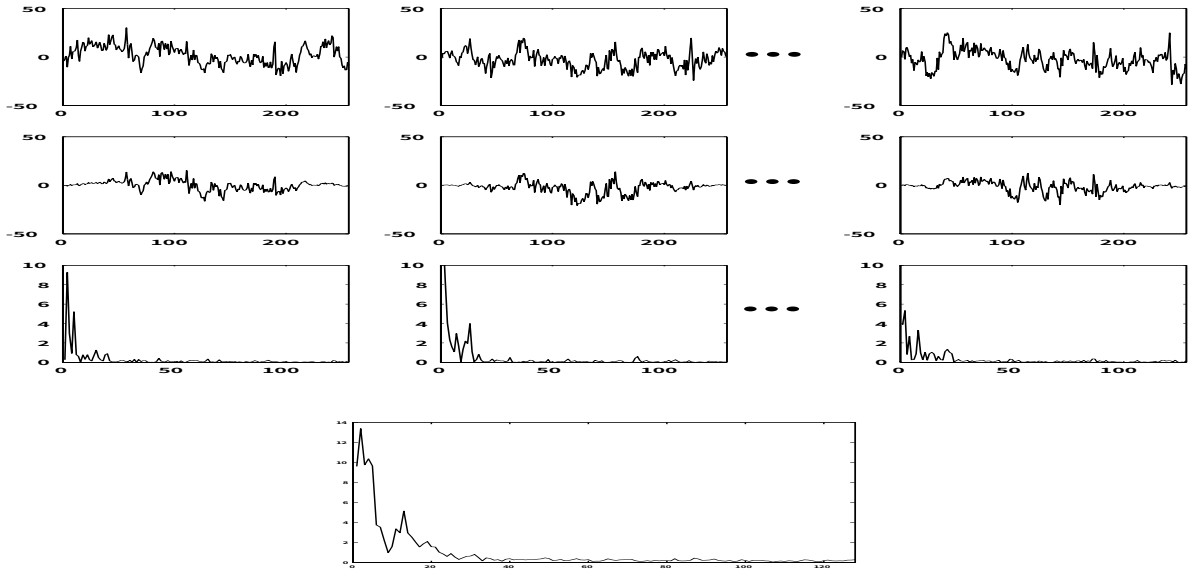


Figure 3.8: Welch's averaged periodogram method for spectral estimation.

Segment size. Regarding the choice of a segment size, it should be noted that although long segments yield a better definition in the frequency domain, the number of segments used in the average decreases when their size increases (at fixed total length N). There is thus a trade-off between definition in the frequency domain and quality of the estimate (the variance of the estimate decreases when the number of segments increases). In this case, 256- and 512-sample segments (1.28 and 2.56 sec respectively) were used in both experiments. In general short (i.e. 1.28 sec) segments were preferred, because they provide a more reliable estimate, with a still acceptable definition in the frequency domain (about 0.79Hz).

Coherence

Besides analyzing spectral properties of each channel separately, it is interesting to inspect joint spectral properties of pairs of EEG channels. Given two jointly stationary processes $X_1(t)$ and $X_2(t)$, the *cross-spectral density function* $h_{12}(\omega)$ can be expressed as the Fourier transform of the cross-covariance function $R_{12}(\tau)$ [65]:

$$h_{12}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega\tau} R_{12}(\tau)\, d\tau. \tag{3.10}$$

Using 3.10, the *coherence function* is defined as:

$$w_{12}(\omega) = \frac{|h_{12}(\omega)|}{\{h_{11}(\omega)h_{22}(\omega)\}^{1/2}}, \tag{3.11}$$

where $h_{11}(\omega)$ and $h_{22}(\omega)$ are power spectral density functions of $X_1(t)$ and $X_2(t)$, respectively. From this definition, w_{12} is comprised between 0 and 1 and can be interpreted as the correlation coefficient between the random coefficients of the components in $X_1(1)$ and $X_2(t)$ at frequency ω [65].

Estimation of the coherence function. The estimation of the coherence function requires the estimation of $h_{11}(\omega)$, $h_{22}(\omega)$, and $h_{12}(\omega)$. $h_{11}(\omega)$ and $h_{22}(\omega)$ can be estimated, using Welch's method presented in the previous section.

For the estimation of $h_{12}(\omega)$, it is also possible to use Welch's method. Let $X_{1,1}, X_{1,2}, \ldots, X_{1,N}$ and $X_{2,1}, X_{2,2}, \ldots, X_{2,N}$ be observations of $X_1(t)$ and $X_2(t)$, respectively. The *cross-periodogram* $I_{N,12}$ between $X_{1,t}$ and $X_{2,t}$ can be defined as the product of the sample spectral density function $\zeta_{X_1}(\omega)$ of $X_{1,t}$ and of the conjugate of the sample spectral density function $\zeta_{X_2}(\omega)$ of $X_{2,t}$:

$$I_N(\omega) = \zeta_{X_1}(\omega)\, \zeta^*_{X_2}(\omega) \tag{3.12}$$

where

$$\zeta_{X_j}(\omega) = \frac{1}{\sqrt{2\pi N}} \sum_{t=1}^{N} X_{j,t}\, e^{i\omega t}, \qquad \text{for } j = 1, 2.$$

The estimation is done as follows. The two signals are divided into m (possibly overlapping) sections. After having applied a data window to each segment, the periodogram of each segment is computed as well as the cross-periodogram of each pair of corresponding segments. Next, both the periodogram and the cross-periodogram are averaged over all sections. Finally, the coherence function is computed according to equation 3.11.

For coherence estimation, a shorter segment size (0.64 sec) was juged necessary, for the quality of the estimate to be acceptable.

Higher-order spectra

Higher-order spectra do not only reveal amplitude information about a process (like do second-order statistics), but also reveal phase information. They are especially useful for dealing with non-Gaussian and nonlinear processes [51]. For this reason, higher-order spectra seem particularly well-suited for EEG analysis, seen the nonlinear nature of scalp potentials.

Investigations were limited to third-order statistics, namely *bicoherence*. Given three jointly stationary processes $X_1(t)$, $X_2(t)$ and $X_3(t)$, the *third-order cross-cumulant* of $X_1(t)$, $X_2(t)$ and $X_3(t)$ is defined as:

$$C_{123}(\tau_1, \tau_2) = E\{X_1^*(t)\, X_2(t+\tau_1)\, X_3(t+\tau_2)\}. \tag{3.13}$$

The *third-order spectral density function*, or *bispectrum* is defined as the Fourier transform of the third-order cumulant:

$$h_{123}(\omega_1, \omega_2) = \frac{1}{4\pi^2} \int\int_{-\infty}^{\infty} e^{-i(\omega_1\tau_1 + \omega_2\tau_2)} C_{123}(\tau_1, \tau_2)\, d\tau_1\, d\tau_2. \tag{3.14}$$

Finally, the *bicoherence function* of $X_1(t)$, $X_2(t)$ and $X_3(t)$ is defined as:

$$w_{123}(\omega_1, \omega_2) = \frac{|h_{123}(\omega_1, \omega_2)|}{\{h_{11}(\omega_1 + \omega_2) h_{22}(\omega_1) h_{33}(\omega_2)\}^{1/2}}. \tag{3.15}$$

Estimation of the cross-bicoherence function. The same kind of method was used to estimate cross-bicoherence as for second-order spectral estimation. The observed time series $X_{1,t}$, $X_{2,t}$ and $X_{3,t}$ are segmented into possibly overlapping sections; the mean is removed from each record, a time-domain window is applied, and the sample spectral density function (FFT) $\zeta_{j,k}(\omega)$ computed. The cross-bispectrum of the kth section is computed as $\hat{w}_{123}(\omega_1, \omega_2) = \zeta^*_{1,k}(\omega_1 + \omega_2)\zeta_{2,k}(\omega_1)\zeta_{3,k}(\omega_2)$ (the k index of $\zeta_{j,k}$ refers to the kth section). Periodograms are computed as usual. Finally, the spectral and cross-biscpectral estimates are averaged accross records, and the cross-bicoherence estimated according to equation 3.15.

Probably the most important drawback to the use of higher-order spectral methods is that they require longer data lengths than do second-order spectral methods, in order to reduce the variance of the estimate [51]. I used higher-order spectra only in the main experiment, where the time interval was limited to 10 seconds. The estimation was done with extremely short windows (64 samples, that is 0.32 seconds), in order to have a sufficiently small variance. This drastic measure had the effect to yield a relatively poor definition in frequency domain: 3.125Hz. Nevertheless, this allowed to get an idea about higher-order phenomenon in broad frequency bands.

3.2.2 EEG features

To summarize, the following features were computed for each EEG record[6]:

1. *Log PSD*:* The power spectral density function was estimated and integrated over a number of frequency bands of interest. Furthermore, the obtained coefficients were log-transformed, in order to obtain a Gaussian distribution [23].

2. *Log power asymmetry*:* Difference of particular log-transformed PSD coefficients were computed, to express the relative 'shift' in activity between two sites or from one frequency band to another.

3. *Coherence*:* The coherence function was estimated and integrated over frequency bands of interest. The computed coefficients (termed as *coherencies* in the following) express the degree of agreement between the activity of two cerebral locations in a given frequency.

4. *Peak frequency:* The frequency for which the PSD function reaches a maximum within the alpha band (8∼13Hz) was determined.

5. *Cross-bicoherence:* The cross-bicoherence function was estimated. Then, values of the cross-bicoherence at particular loci in the two-dimensional frequency domain were extracted (the function was not integrated, since frequency 'bands' usually contained only one estimated point).

[6]Only the features marked with an asterisk (*) were used in the preliminary experiment.

Chapter 4

Experiments

This chapter presents two EEG recording experiments aiming at exploring influences of emotion on the EEG.

In the first section, a preliminary experiment is presented in which stress was studied as a particular emotion. It was also the opportunity to get familiar with basic techniques for EEG recording and analysis.

The main experiment, presented in the second part of this chapter, focused on effects of general emotion on the EEG. Results are presented and extensively discussed from a psychophysiological point of view.

4.1 EEG-based stress assessment

4.1.1 Introduction

In a preliminary experiment, I focused on *emotional stress* as a particular emotional condition, and investigated the influence of stress on the EEG. This issue is very important, as it could be used in stress assessment applications, like aircraft pilot stress monitoring, or patient postoperative stress evaluation in hospitals. Classical stress assessment techniques include analysis of the electrocardiogram [53], of blood pressure [3] and other physiological factors [37]. EEG analysis provides an alternative assessment method, which makes use of higher level information than most existing methods.

In this study, video game playing was used to elicit stressing situations. Certain video games are indeed known to cause stress increase, by requiring quick reactions and key movements. For instance, Johnstone [33] reports the use of computer games to elicit emotional speech.

The objective of this experiment was to determine changes in the spectral properties of the EEG of humans, associated with stress increase. Because of important inter-subject differences in EEG patterns [49], this objective is twofold: (1) describe individual EEG responses to stress increase, and (2) point out general trends in EEG changes observed among subjects.

4.1.2 Methods

Experimental setup

Stress was elicited in nine healthy subjects (4 male, 5 female) aged from 21 to 28 (mean 24.3), by means of video game playing. Eight games were selected on basis of their capacity

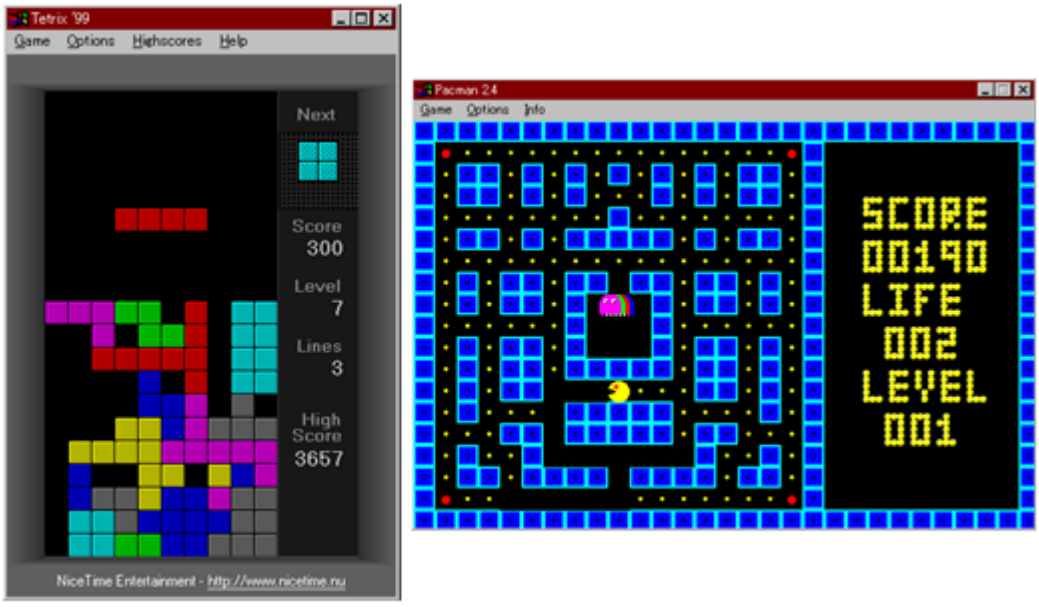


Figure 4.1: Examples of computer games used to elicit stress.

to elicit stress by requiring fast key responses (see figure 4.1). Seven games were played on a portable computer (LCD display –diagonal: 14"– to minimize radiations), and one on a GameBoy^{TM}. The experiment protocol, summarized in Table 4.1, is as follows. The subject, comfortably sitting at the desk, was asked to play all games in turn for 5 minutes each (5 games included 3 levels of difficulty, which were played in turn: 3 minutes for the easiest level and one minute for the medium and difficult levels). Before playing each game, a rest period was taken for 30 seconds. After each game, the subject was asked to evaluate stress experienced while playing, by ranking the game on a "stress scale" (from 1 –not stressing at all– to 10 –extremely stressing–), and answering a few questions.

Table 4.1: Experimental protocol.

For each game,		
1.	Rest (eyes closed, hands on knees)	30 sec
2.	Level 1	3 min
3.	Level 2	1 min
4.	Level 3	1 min
5.	Stress evaluation	...

EEG processing and data analysis

Ten channels of EEG (Fp1, Fp2, F3, F4, P3, P4, O1, O2, T3 and T4) and two EOG channels were recorded as described in section 3.1.2. After preprocessing for artifact minimization (see section 3.1.3), EEG records were sliced into 10.24- and 5.12-second segments. For each segment, power spectral density (PSD) and inter-electrode coherence functions were esti-

mated and spectral features computed as described in section 3.2. The total studied spectrum (4∼15Hz) was subdivided into three main frequency bands: (1) theta (5Hz∼7Hz), (2) alpha (8Hz∼12Hz), and (3) low beta (13Hz∼15Hz). A further subdivision of the alpha band into low alpha (8Hz∼10Hz), narrow alpha (9.5Hz∼10.5Hz) and high alpha (10Hz∼12Hz) was also considered.

From the PSD and coherence coefficients (which are called "single features" in the following), "integrated features" were computed by summing single features on particular functionally differentiated areas of the brain: frontal lobe (Fp#, F#), fronto-temporal lobe (Fp#, F#, T#), parieto-occipital lobe (P#, O#), temporal lobe (T#) and whole brain (all electrodes). Integrated features are of three types: (1) total power (sum of PSD coefficients), (2) right-left asymmetry (difference of PSD coefficients at right and left homotopic electrodes) and opposed hemisphere coherence (sum of coherence coefficients involving electrodes located in different hemispheres).

In order to compare the influence of stress among the subjects, three games of increasing stress (the same for all subjects) were selected on basis of self-reports: L (low stress), M (moderate stress) and H (high stress). Were retained only those features of which the median increased or decreased monotonically among the three games, that is, for which

$$m_L \leq m_M \leq m_H \quad \text{or} \quad m_L \geq m_M \geq m_H$$

where m_L, m_M and m_H denote the median of the feature under consideration for games L, M and H, respectively.

4.1.3 Results and discussion

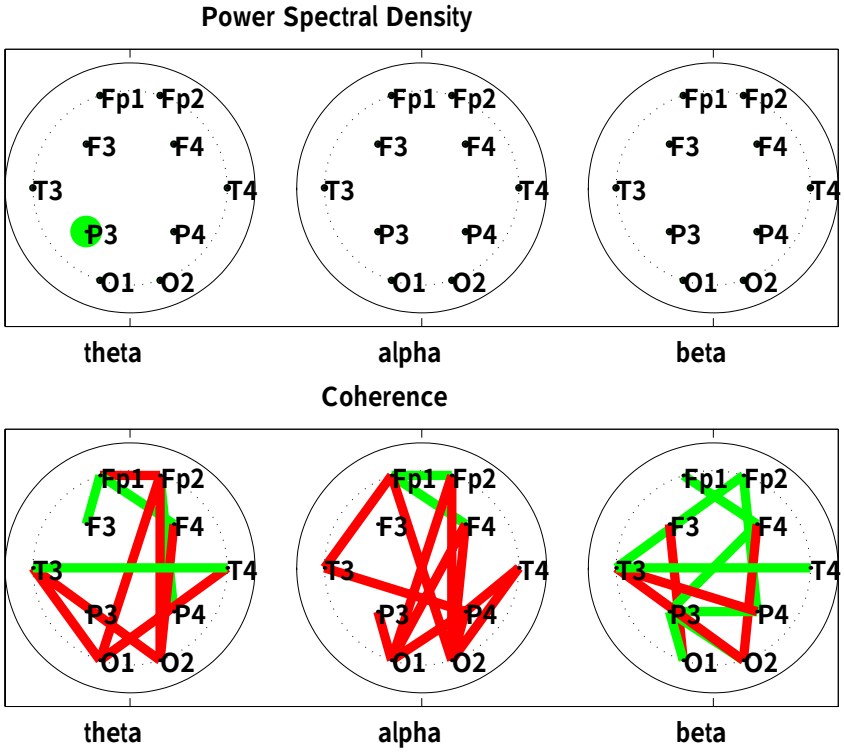
Single features

Figure 4.2 shows significant monotonic changes in PSD and coherence for two subjects (time interval: 10.24 sec). For each subject, the upper graph shows significant monotonic changes in PSD (marked by a circle). The lower graph shows significant monotonic changes in inter-electrode coherence (marked by a line). An increase is marked in green and a decrease in red.

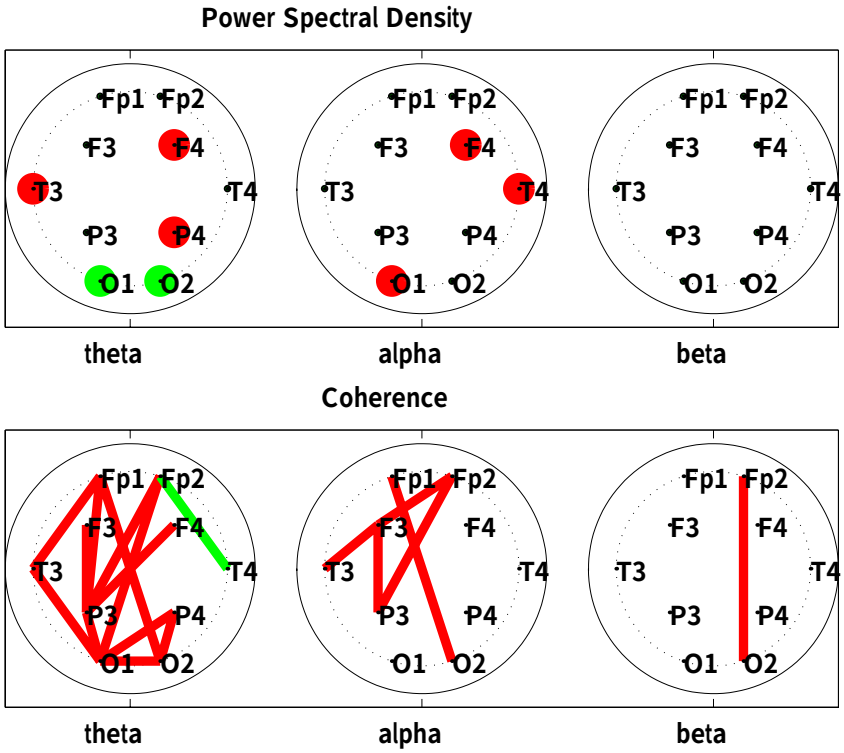
These graphs can be interpreted as follows: for example, figure 4.2 (a) shows that the PSD in theta increases at the left-parietal electrode (P3) for subject 'KR', when stress increases. Similarly, an increase of stress for 'KR' is associated with an increase in temporal coherence (T3–T4), both in theta and low beta. These two graphs (as well as results for other subjects) show that single features significantly changing with stress are considerably subject-dependent. This can be explained by the following factors: (1) stress experienced by subjects undertaking similar stimuli differ from person to person, (2) stress is experienced in different ways by different subjects, and (3) individual differences in the EEG are important.

The choice of the computer games used in this analysis has been performed so that the first factor has a minimal influence, that is, so that the experienced stress reported by subjects is lowest for game L and highest for game H. This selection has been done by using *relative* ranking of different games, rather than absolute stress index (because the way in which different people use a scale may vary considerably).

A second factor concerns the nature of the feeling of stress. The experience of stress is rarely an unmixed feeling. In other words, a subject reporting an increase in stress might also feel some kind of excitation, which is not necessarily a negative feeling, by opposition to



(a)



(b)

Figure 4.2: Significant monotonic changes in PSD and coherence with increasing stress for subjects 'KR' (a) and 'CH' (b).

the negative nature of stress. Some difference in the EEG might be explained by a mixed emotional state.

But the most important factor explaining inter-subject differences in single features correlated with stress is probably the fact that the EEG substantially varies from one individual to another. This has been reported, among others, by Maremmani et al. [49] and Wheeler et al. [81], who discuss personality and affective style reflected in the EEG. However, these results do not exclude the possibility to perceive general trends in EEG changes, associated with stress increase.

To summarize, the analysis of single features allow to precisely study changes in spectral properties of the EEG, accompanying a condition of increasing stress. Single features are suited for an individual analysis and could be used in a stress-detection system, tuned for each user in particular, and hence achieving more precise assessments.

Integrated features

Integrated features, computed from single features, are expected to show general trends in EEG changes related to increase in reported stress. It is convenient to summarize monotonic changes of integrated features in a table, were all subjects are listed. Since integrated features are always computed over a given brain region, one table is produced for each region. Table 4.2 shows results for the "whole brain" region (all electrodes). Similar tables were constructed for the other regions. The notation is as follows: T, A, A1, A2, An and B refer to the six frequency bands (theta, alpha, low alpha low, high alpha, narrow alpha and beta, respectively); a monotonic increase (resp. decrease) in the studied integrated feature is indicated by a '+' (resp. '–') sign; each row corresponds to one subject.

Table 4.2: Integrated features computed over the "whole brain" region.

Subj.	total power						right-left asymmetry						opposed hemisphere coherence					
	T	A	A1	A2	An	B	T	A	A1	A2	An	B	T	A	A1	A2	An	B
LA								–		–	–							
KR			–					–	–	–	–	–				+		
HU													–	+		+		
CH	–	–	–	–			+					+				–	+	
HA	+	–	–		–						–			+		+		
KE	+							–	–				+	+	+	+		+
HR	+	+	+	+	+	+	–	–		–			–	–	–	–	–	
ED			+					–	–					+	–	+		+
MA	+	+	+	+	+	+		+	+	+	+		+	+	+	+		+

A column-wise inspection of such a table gives some information about similarities and differences between subjects. For example, the first column of table 4.2 indicates that four subjects (vs. 1) have shown a monotonic increase in theta PSD over the whole brain region. A systematic inspection of table 4.2 and of other tables of integrated features computed over different brain regions was performed. Global changes noticed among all tables are summarized in table 4.3.

Table 4.3: Global changes in integrated features, when reported stress increases.

brain region	**integrated feature**	**trend**	**# cases (# opp. cases)**	
frontal	theta opp. hem. coherence	↗	4	(1)
fronto-temporal	theta opp. hem. coherence	↗	4	(0)
temporal	alpha r-l asymmetry	↘	4	(1)
	beta r-l asymmetry	↘	4	(1)
parieto-occipital	alpha r-l asymmetry	↘	4	(1)
whole	theta total power	↗	4	(1)
	alpha r-l asymmetry	↘	5	(1)
	alpha opp. hem. coherence	↗	5	(1)
	high alpha opp. hem. coher.	↗	6	(2)

Before discussing these results, the importance of integrated features in comparing EEG of different individuals is pointed out. Individual differences in single features, and the apparent difficulty to generalize changes in the EEG related to reported stress increase, motivate the search for more global features. Integrated features result from the summation of single features over functionally differentiated brain areas. This simple operation can be justified by the fact that, while summation reduces the spatial resolution of the features (from one feature per electrode, we obtain only one feature per brain region), it also integrates similar effects in a given region, and by this, increases the agreement between individuals.

Results from table 4.3 are now discussed. First, the most striking trend seems to be a decrease in right-left asymmetry, which appears in the parieto-occipital and temporal regions, for four subjects (vs. one), as well as for the whole brain (5 subjects vs. 1). Intuitively, a decrease in alpha right-left asymmetry denotes a shift in alpha activity, towards the left hemisphere, which, considering the alpha rhythm as inversely related to activation [13], would suggest that there is a shift in cerebral activity from the left to the right hemisphere. This is in agreement with the findings of Davidson [13], associating the right hemisphere with negative emotions (and withdrawal behaviors). Another trend is the increase of alpha opposed hemisphere coherence, which indicates an "alignment" between the two hemispheres in the frequency content in the alpha band, occuring in response to stress increase. Finally, the slight increase in total theta power may be related to the findings of Walter, summarized by Nierdmeyer [57, p. 144]: "Walter (1959) has associated the theta activity with emotional processes and (...) he also attributed runs of theta waves to the emotional correlated of disppointment and frustration because of its appearance at the interruption of a pleasurable stimulus." However, further investigations would be necessary to verify whether the theta activity actually occurred in bursts at critical times in game playing.

4.1.4 Conclusion

The first study focused on mental stress as a particular emotional state. Single EEG spectral features (PSD and coherence coefficients) were computed and have shown to allow a detailed description of the effect of stress on the EEG. Single features are precise but vary considerably among individuals. On the other hand, integrated features (total power, right-left asymmetry and opposed hemisphere coherence) allow to describe general trends in EEG changes accom-

panying stress increase, among different persons.

This preliminary experiment however has the following shortcomings:

- *Long tasks:* The tasks during which the EEG was recorded (i.e. computer game playing) lasted from 1 to 3 minutes. As discussed in section 3.2.1, although long EEG segments theoretically would yield better spectral estimates, long mental tasks usually involve a variety of brain states, making the stationarity hypothesis questionable.

- *No reference EEG:* Investigations for EEG changes associated with stress increase concerned three selected games of increasing stress. Although a low-stressing game (L) was used, no actual reference EEG (known to involve no stress at all) was recorded for comparison.

- *Abundance of muscle artifacts:* Game playing are quite "active" tasks, requiring quick movements from the subjects. This inevitably yielded many muscle artifacts. As discussed in section 3.1.3, since no efficient minimization technique is known, the only solution was to filter out frequencies above 15Hz.

In the next experiment, a more general form of emotion was studied and the above-mentioned shortcomings were eliminated.

4.2 Emotion and EEG

4.2.1 Introduction

Literature review

Electroencephalography has been used by several researchers as a brain imaging technique to study how emotion is generated in the brain. Many studies are based on power spectrum analysis, and particularly on the analysis of asymmetry between left and right hemispheres. Dawson [16] studied the relationship between frontal asymmetry and emotion expression in infants, focusing on the 6∼9Hz frequency band (equivalent of the alpha band in infants). Davidson [13, 14], Davidson and Irwin [15], Wheeler et al. [81] have studied extensively the relationship between anterior cerebral asymmetry and emotion, on basis of electroencephalographic data, but also of other data (e.g. fMRI). These investigations, limited to the power in alpha band in the frontal lobe, are however not in agreement with [16] nor [11], neither with the current results. As discussed later on, the sole alpha power asymmetry is believed to be insufficient for discriminating between emotional states. Besides, Maremmani et al. [49], who also studied EEG asymmetry (in total power, this time), suggested the influence of personality on task-dependent EEG asymmetry. A more detailed spectral analysis was conducted by Crawford et al. [11], who investigated changes in regional EEG activity in several frequency bands, in response to self-generated emotions. Although hemispheric differences were observed between happy and sad emotions, they concerned only the low alpha activity and were localized in the parietal lobe, rather than in the frontal lobe, as suggested by Davidson [13].

Studies about emotion and EEG involving other features than the power spectrum are rather infrequent. Hinrichs and Machleidt [26] studied how basic emotions are reflected in EEG-coherences, during self-elicited emotions. There description is however qualitative, and concerns only a subset of the data, selected on an arbitrary basis. Musha et al. [56] showed that EEG inter-electrode coherence could be used as features for discriminating between emotions. The problem of artifacts was however not clearly dealt with and emotional stimulus were not standard, making results difficult to reproduce. An attempt to observe the bispectrum in positive and negative emotional states was made by Kim et al. [38], who unfortunately obtained no consistent results among subjects, because of noise and individual differences. Instead of looking at the power spectrum, Kostyunina and Kulikov [40] considered shifts in peak frequency in the alpha band, in various emotional states. Increase and decrease in peak alpha frequency are however not confirmed by [26], neither in the current experiments. Finally, besides spectral analysis techniques, let us mention the possible use of dimensional analysis of the EEG, as proposed by Aftanas et al. [1, 2]. Yet, these techniques were not used in the present study.

Objective

The objective of the main experiment was to investigate how emotion influences the EEG, so that this information could be used in an emotion expressing interface for the disabled.

Rather than understanding neural mechanisms of emotions, it was aimed at finding EEG features allowing a good discrimination between different emotional states, and possibly, a quantification of emotional intensity.

As previously mentioned, an obvious difficulty in trying to determine EEG-patterns associated to particular emotional states is the existance of individual differences in the EEG. For this reason, rather than a “general emotion recognizer” which could be used “as is” with

different subjects, the objective is to design an emotion recognition system which could be trained and tuned to a particular person. The design of such a system can be divided in two successive steps:

1. Find suitable features for discriminating between emotions, independently of a particular person.

2. Design a system capable of efficiently learning a particular mapping between EEG features and emotion.

The first step is the main concern of this section. The second issue is addressed and discussed extensively in chapter 5.

4.2.2 Methods

Emotional stimulation and evaluation

Emotional stimulation

Various methods can be used to elicit emotional experience. Table 4.4 is a non-exhaustive list of methods, which have been used principally in experiments involving EEG measurement.

Table 4.4: Methods used to elicit emotion.

Method	Reference	# of sub.
mental technique		
imagination technique	Hinrichs and Machleidt [26]	32
self-induced imagery of past personal experiences	Crawford et al. [11]	31
	Kostyunina and Kulikov [40]	12
visual stimuli		
pictures (IAPS)	Lang et al. [45]	12
	Canli et al. [9]	14
pictures (EPPP)	Sutton et al. [74]	24
emotional faces	Borod et al. [7]	36
auditory stimuli		
auditory stimuli (pleasant and unpleasant sounds)	Kim et al. [38]	18
verbal stimulus and music	Maremmani et al. [49]	12
speech prosody (declarative sentences with happy, sad or neutral intonation)	Pihan et al. [64]	16
	Borod et al. [7]	(36)
audio-visual stimuli		
video clips	Aftanas et al. [2]	76
	Wheeler et al. [81]	81
other stimuli		
lexical stimuli (emotional words and sentences)	Borod et al. [7]	(36)
pleasant and neutral touch	Francis et al. [22]	4
taste and smell	Francis et al. [22]	(4)

The present experiment involved two very common sensory channels: visual and auditory. The stimuli were of three types: pictures, sounds, and combinations (of one picture and one sound). Pictures and sounds were taken from the International Affective Picture System (IAPS) and International Affective Digitized Sounds (IADS), developed by the CSEA-NIMH at the University of Florida [44, 8]. The IAPS and IADS are standardized sets of more than 700 digitized photographs and 100 digitized sounds including, for each stimulus, affective ratings determined according to a standardized procedure administrated to a large number of subjects.

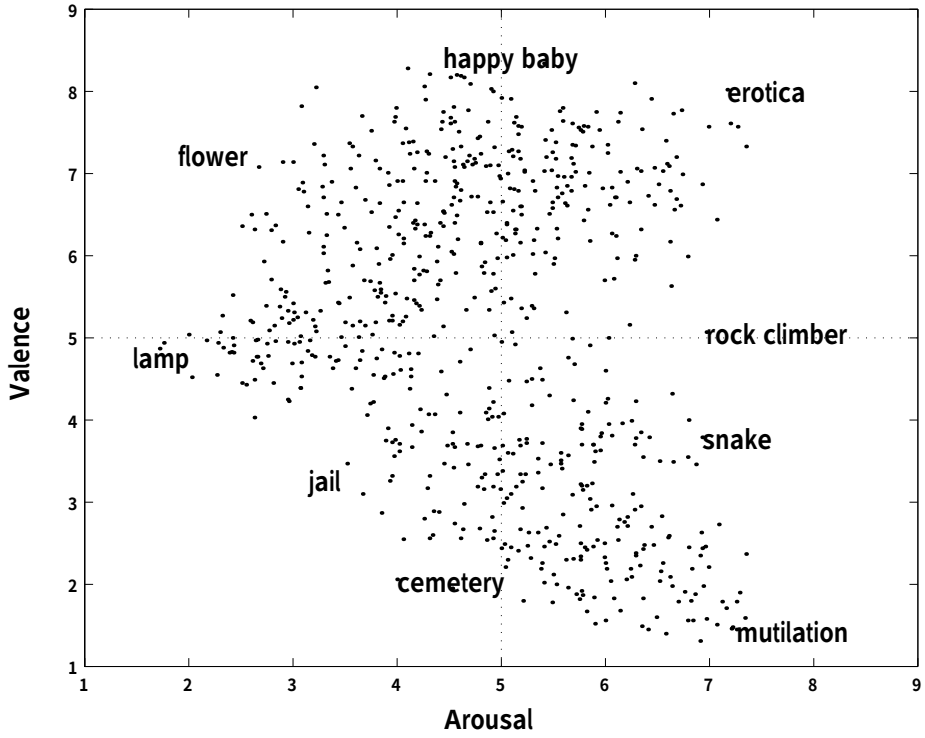


Figure 4.3: Affective ratings for the International Affective Picture System (IAPS). Individual ratings were made on a nine-point scale: for valence, 1 = "completely unhappy" and 9 = "completely happy"; for arousal, 1 = "completely calm" and 9 = "completely aroused".

As discussed in section 2.2.2, I opted for a simple and universal representation of emotion along two dimensions: *valence* (happy/unhappy) and *arousal* (excited/calm). These two dimensions are a subset of the three-dimensional representation used by Lang et al. [44] and Bradley and Lang [8] for collecting affective ratings for the IAPS and IADS. Figure 4.3 shows mean affective ratings against valence and arousal for IAPS pictures.

On basis of these ratings, 32 pictures and 32 sounds were selected, uniformly distributed along the valence and arousal axes. Moreover, 30 combinations were constructed by selecting 30 additional pairs of one picture one sound, each combination being presented as a whole to the subject during the experiment. The three sets of stimuli are listed in appendix.

Evaluation of experienced emotion

It is important to distinguish between emotional stimulus and actually experienced emotion. IAPS/IADS stimuli are associated with "standard ratings", but it is important to remember that these are *average* ratings, which are not necessarily predicting the actual emotional state of a subject to which the stimuli are presented.

For this reason, the subjects were asked to rate their own emotional experience while being presented each stimulus. Each subject was told about the importance of these ratings, with particular emphasis on the importance to rate *how he/she actually felt* while viewing each picture and hearing each sound.

For collecting emotional ratings, a standardized rating system called "Self-Assessment Manikin" (SAM) [43] was used, which allows to easily rate emotion against both valence and arousal scales. The computer-based version of SAM specially developed for this experiment is

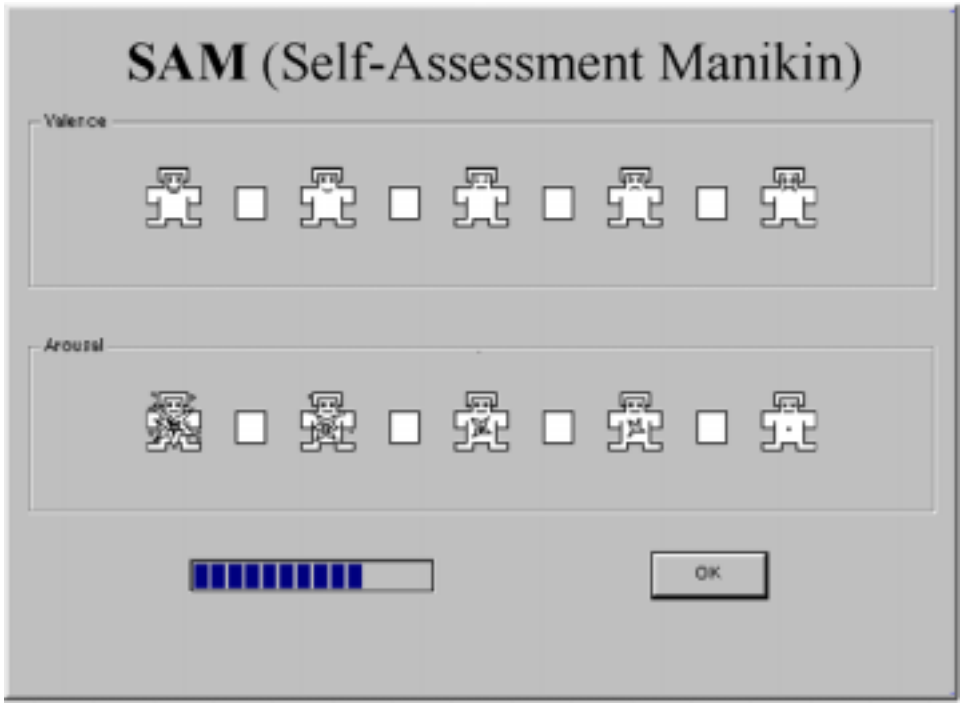


Figure 4.4: Computer-based version of the "Self-Assessment Manikin" (SAM) system.

shown in figure 4.4.

Experimental setup and procedure

EEG recording setup

EEG recordings were carried out in a shield room (see section 3.1.2). The experimental setup is depicted in figure 4.5.

Thirteen channels of EEG were recorded from electrodes Fp1, Fp2, F3, F4, F7, F8, T3, T4, P3, P4, O1, and O2, placed according to the 10–20 system (see figure 3.3) all referenced at the left earlobe A1. As discussed in section 3.1.2, the right earlobe potential A2 was recorded as well, for referencing all channels to a "digitally linked ear". A vertical and an horizontal EOG channel were also recorded for artifact minimization (see section 3.1.3).

Subjects

Twenty volunteer subjects (13 male, 17 female) participated to the experiment. Subjects were aged from 22 to 33 (mean: 24.5) and healthy. For each subject, a written consent was collected before starting the experiment.

Experiment protocol

The experiment consisted in three successive phases of about 15 minutes, each of which involving a particular type of emotional stimulus. (1) pictures only, (2) sounds only, (3) combinations of one picture and one sound. Recordings during sound stimuli (phase II) were performed with closed eyes.

After electrodes placement was completed, the experimental procedure was explained to the subject by way of an automatic demonstration software (including spoken instructions

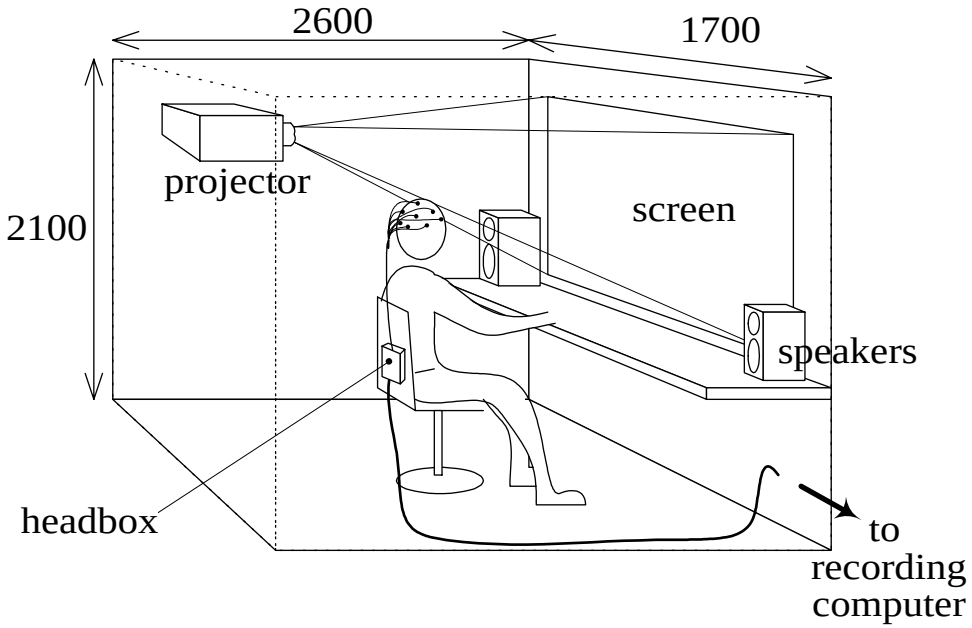


Figure 4.5: Experimental setup for EEG recording with emotional stimuli.

and visual examples). Instructions about the use of the computer-based SAM system (derived from [44]) were included in the demonstration. Just after having received the explanations, the subject was asked to practice on three examples (one picture, one sound and one combination, not part of the previously mentioned sets).

The experiment was driven automatically by the recording computer, which was also responsible for presenting stimuli (pictures and sounds), and synchronizing the recording. Between each of the three phases, a short pause was allowed for relaxation.

Each phase consisted in the following sequence, carried out in loop (see table 4.5). First, a warning message appeared on the screen, telling the subject to prepare for the next stimulus. Next, the stimulus (picture, sound, or combination) was presented to the subject (duration of the stimulus, pictures: 10 s, sounds: 6 s). When the stimulus stopped, the "rating window" (shown in figure 4.4) appeared on the screen. Fifteen seconds were allowed for rating (depicted by a progress bar), but the subject could use the "OK" button to go directly to the next stimulus, provided the rating was completed.

Table 4.5: Procedure and timing for phases I, II and III.

For each stimulus,		
1.	Warning message	5 sec
2.	Stimulus	6 or 10 sec
3.	SAM rating (valence, arousal)	max 15 sec

The first phase consisted of 32 picture stimuli, each of which was presented for 10 seconds. The second phase consisted in 32 sound stimuli, each of which lasting 6 seconds. The subject was asked to close his/her eyes from the apparition of the warning message on the screen, until the end of the sound stimulus. The third phase consisted in 30 combination stimuli. The

sound lasted only for 6 seconds, but the image remained on for 10 seconds. The subjects was asked to continue looking at the picture until it disappeared, before making his/her rating.

EEG processing

The recording procedure described before yielded, for each subject, three sets of short EEG segments (also called "epoch" in the following). Segment length was 10 second for phases I and III, and 6 seconds for phase II. Unlike in the previous experiment, no further segmentation was necessary. For computational reasons, however, the segment length was adjusted as follows. For segments of phases I and III, it was set to 2048 samples, that is 10.24 s[1]. For segments of phases II, only the last 5.12 seconds were kept, yielding 1024 sample segments.

Next, each segment was processed according to the following steps: (1) re-referencing to a "digital linked ear"; (2) eye movement artifact minimization; (3) muscle artifact detection; (4) spectral feature computation. Computed features were PSD, coherence and cross bicoherence in various frequency bands, as well as peak frequency in alpha. Steps (1) to (4) are detailed in chapter 3.

4.2.3 Results

After explaining how emotional ratings were handled, results are presented separately for valence and arousal dimensions.

Emotional ratings

As a result of the experimental procedure, each of the 94 epochs collected for every subject was associated with a bidimensional affective rating against valence and arousal on a scale from 1 to 9.

Individual differences in the use of rating scales. Different people sometimes tend to utilize rating scales very differently, even after having received exactly the same explanation about the meaning of the scale. Figure 4.6 illustrates this, showing the repartition of the 94 valence ratings on the 9 point scale, for three subjects: 'HU', 'KU' and 'RY'.

While most subjects display near-Gaussian rating distributions, some subjects tend to privilege extremes of the scale, or not to make use of all available values. For this reason, it is important not to use absolute ratings, but rather relative values, which tend to be more consistent.

Two-class comparison. This chapter concentrates on the search for features that efficiently discriminate between emotions. Because of uncertainty in subject ratings, rather than looking at ratings as samples of a continous function, the ratings were used to define *well separated classes of emotional states*. For each subject, some EEG epochs were selected for each phase and grouped into two classes, according to three subdivisions presented below. Each subdivision resulted in two classes of 6 samples. The reason for taking only a little subset ($2 \times 6 = 12$ epochs out of the 30 or 32 available) is that, at first, only "extreme" epochs were considered (that is, epochs corresponding to a clearly marked emotion, as explained below).

[1]Most segments lasted a bit longer than 10 seconds; if a segment was too long, the very first samples were discarded; when a segment was too short, it was padded with a few samples just preceding the stimulus.

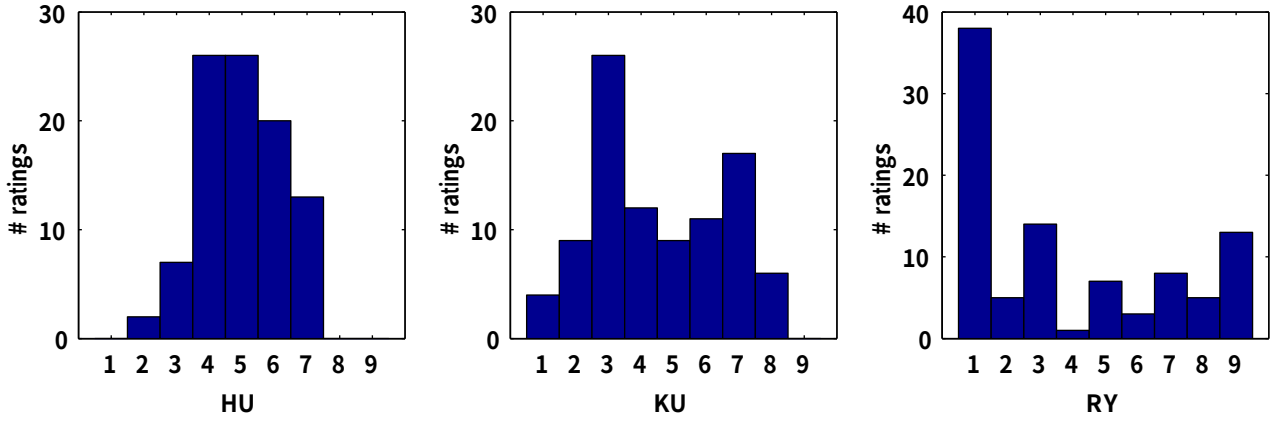


Figure 4.6: Repartition of valence ratings on the 9 point scale, for three subjects.

Let us now look at the three subdivisions of the epochs, according to the ratings (the two first subdivisons concern the valence dimension; the third subdivision concerns the arousal dimension):

- *"Emotional" and "neutral" states.* The class of emotional epochs contains 3 epochs of maximum valence ("happy") and 3 epochs of minimum valence ("unhappy"). The class of neutral (non-emotional) states contains 6 epochs of medium valence (that is, closest to 5, the "neither happy nor unhappy" state).
- *"Positive" and "negative" emotional states.* Six epochs of maximal valence ("happy") and six epochs of minimal valence ("unhappy") are selected, and the median values are ignored.
- *"Excited" and "calm" emotional states.* Six epochs of maximal arousal ("excited") and six epochs of minimal arousal ("calm") are selected, ignoring the median values[2].

The choice between "equivalent" epochs. A problem arises in some cases, when trying to select, for instance, the 3 epochs of maximal (or minimal) valence. Indeed, it is obvious from figure 4.6(c), that more than 3 epochs be attributed the *same* rating. Which epochs should then be selected ? In this analysis, the IAPS/IADS standard results were used to resolve this conflict: in order to choose between several epochs rated equally by some subject, the epochs were selected for which the given stimulus was rated higher (or lower) on the scale, referring to the average ratings.

Comparison between classes. For each computed feature, the equality of means between two classes was tested for each subject (significance levels: P=0.10 and P=0.05) and features of which the mean was significantly different for the two classes were detected.

Results will now be presented for the three subdivisions defined above.

[2]The arousal scale was not further subdivided, because of the apparent absence of a median state: "neither calm, nor excited". Besides, this was also reported by several subjects.

Valence: "emotional" vs. "neutral" epochs

Analysis of power spectral density (PSD) revealed a significant increase in beta activity in the mid-frontal (F3·F4) and parietal (P3·P4) lobes, for several subjects. A decrease in alpha activity in these regions was also sometimes observable. The log-transformed ratio of PSD in beta by PSD in alpha was particularly correlated with emotional experience.

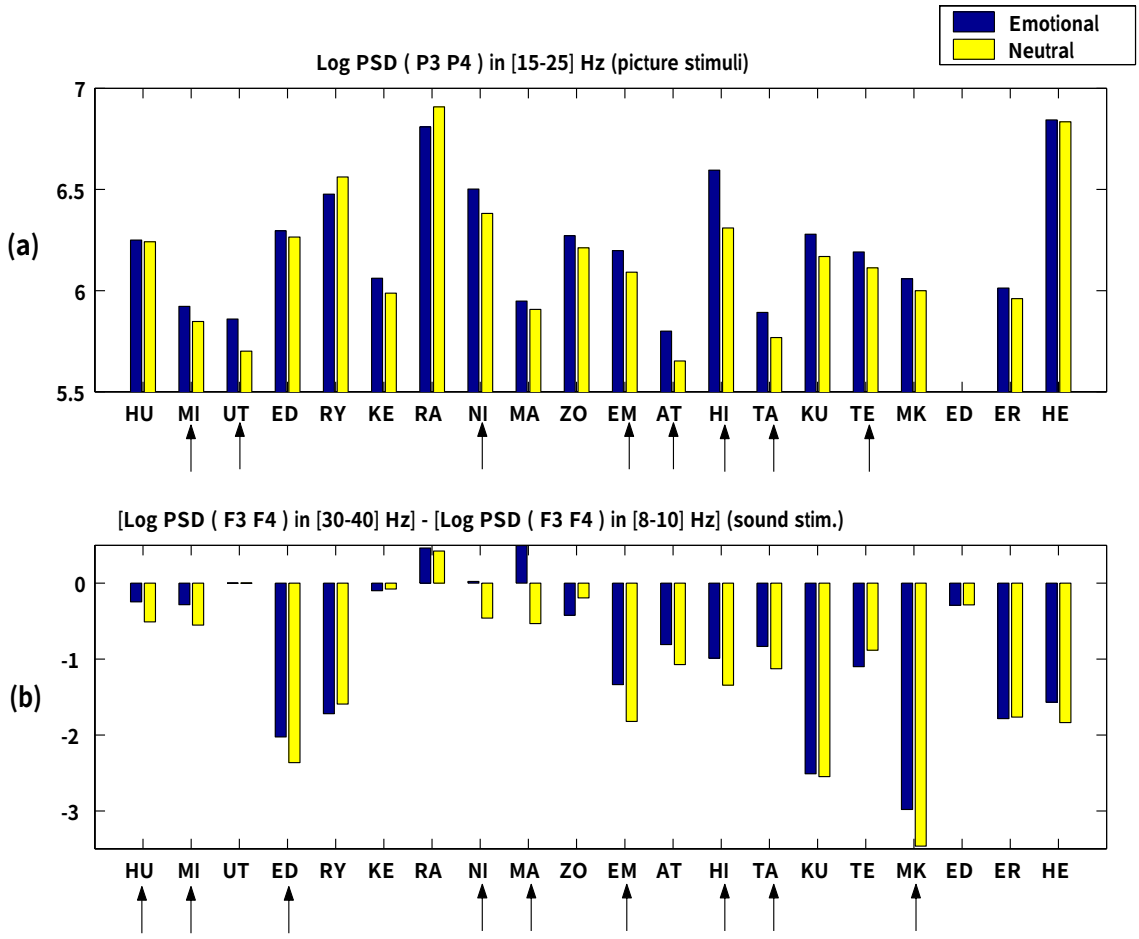


Figure 4.7: Significant changes between "emotional" and "neutral" epochs for picture (a) and sound (b) stimuli. Subjects showing a significant increase accompagnying emotional experience are indicated with an arrow.

Pictures. Figure 4.7 (a) shows the mean of log PSD at parietal electrodes in the [15-25] Hz beta band, for emotional and neutral epochs. This quantity was significantly higher for emotional epochs than for neutral epochs in 8 subjects at P=0.10 (4 subjects at P=0.05), out of 19. Note that no subject presented a significantly lower parietal beta power for emotional epochs. Besides, the log-transformed ratio of PSD in [15-20] Hz (low beta) by PSD in [9-11] Hz (central alpha) at mid-frontal electrodes was also significantly higher for emotional epochs than for neutral epochs in 6 subjects at P=0.10 (5 subjects at P=0.05) out of 17.

Sounds. Figure 4.7 (b) shows the mean of log-transformed ratio of PSD in [30-40] Hz (high beta) by PSD in [8-10] Hz (low alpha) at mid-frontal electrodes. This quantity increased

significantly with emotional experience for 9 subjects at P=0.10 (6 subjects at P=0.05), out of 19. Again, no subject showed the opposite trend. Besides, the log-transformed ratio of PSD in [30-35] Hz (high beta) by PSD in [7.5-13.5] Hz (alpha) at parietal electrodes was significantly higher for emotional epochs than for neutral epochs in 9 subjects (vs. 2) at P=0.10 (4 subjects vs. 1 at P=0.05) out of 20.

Combinations. The tendency of an increase of the ratio of beta activity by alpha activity was also present for combination stimuli, but less marked than for picture and sound stimuli. The log-transformed ratio of PSD in [15-25] Hz (low beta) by PSD in [11.5-13.5] Hz (high alpha) at mid-frontal electrodes was significantly higher for emotional epochs in 6 subjects (vs. 2), out of 19 (P=0.10). Besides, the log PSD in [30-40] Hz at parietal electrodes was significantly (P=0.05) higher for emotional epochs in 6 subjects (vs. 1) out of 20.

Valence: "happy" vs. "unhappy" epochs

Although significant changes in PSD were not observable for more than one third of the subjects in phases I and III, important changes in PSD were observed in phase II. These changes essentially concerned the parietal lobe, where the PSD in central alpha was significantly higher for happy epochs. Less importantly, increase of PSD in high beta and decrease of PSD in theta were also observed.

Difference in inter-electrode coherence were consistently observed for several subjects. The most important effect observed with valence increase was an increase in alpha coherence between distal frontal electrodes (F7·F8), although this was not observed for sound stimuli (i.e., with eye closed).

Bicoherence also showed some significant changes accompagnying stress, but for fewer (usually four or five) subjects.

Pictures. For happy epochs, PSD was significantly lower in [5-8] Hz (theta) at P3 and P4, for 6 subjects (vs. 1), out of 19. In addition, a significantly higher PSD in [30-40] Hz was observed at F3 and F8 for 5 subjects (vs. 1), out of 18.

More essential changes concern inter-electrode coherence. Figure 4.7 (a) shows the mean coherence between F7 and F8 in the [10-13] Hz (high alpha) band, for happy and unhappy epochs. The coherence was significantly higher for happy epochs compared to unhappy epochs, in 10 subjects at P=0.01 (5 subjects at P=0.05) out of 13 subjects. Note that subject 'KU' showed the opposite trend.

Several effects have also been observed for bicoherence. For example, the auto-bicoherence at F3 about the [10,39] Hz^2 bi-dimensional frequency was significantly lower for happy epochs, compared to unhappy epochs, for 5 subjects (vs. 1), out of 17. The same decrease was observed at F4 and F8, as well.

Sounds. The PSD in [9-11] Hz (central alpha) at right parietal electrode P4 was significantly higher for happy than unhappy epochs, in 11 subjects (vs. 1) at P=0.10 (5 subjects vs. 1 at P=0.05), out of 20. Similar increase was also observed at the left homotopic site P3, for 8 subjects at P=0.10 (6 subjects at P=0.05), out of 20.

Figure 4.7 (b) shows the mean coherence between P3 and P4 in the [27.5-30] Hz (medium beta) band, for happy and unhappy epochs. The coherence increased significantly with happiness for 8 subjects (vs.1) at P=0.10 (3 subjects at P=0.05), out of 20.

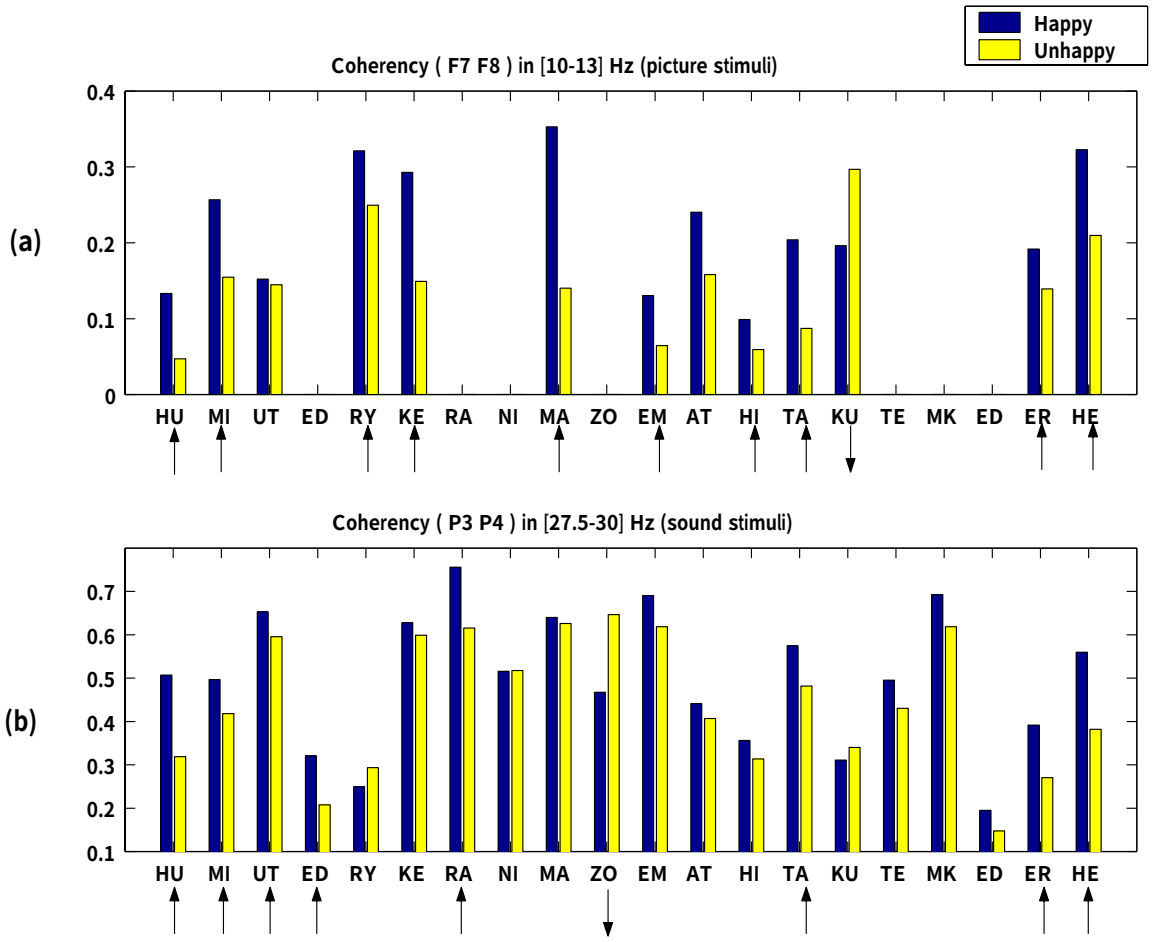


Figure 4.8: Significant changes between "happy" and "unhappy" epochs for picture (a) and sound (b) stimuli. Subjects showing a significant increase (resp. decrease) accompagnying happiness are indicated with an arrow.

In addition to some changes in auto-bicoherence, a significant decrease of the peak frequency in alpha (8-13 Hz) with happiness was observed for 6 subjects (vs.1) out of 19.

Combinations. As for picture stimuli, epochs of phase III did not show many consistent EEG feature changes with valence increase. An increase in high beta was nevertheless observed in the right hemisphere, at mid-frontal electrode F4 for 5 subjects, and at parietal electrode P4 for 6 subjects (vs. 1) out of 19.

For inter-electrode coherence, results were similar to those of figure 4.7 (a). The most important change was a significant increase in coherence between F7 and F8 in the [9-11] Hz (central alpha) band, for 11 subjects at P=0.10 (5 subjects at P=0.05) out of 16. No subject presented the opposite trend.

As for picture stimuli, a significant decrease in auto-bicoherence at F8 about [10,39] Hz^2 was observed for 6 subjects (out of 18). Other observed effects concerned only auto-bicoherence.

Arousal: "excited" vs. "calm" epochs

High arousal was generally characterized by an increase in parietal beta PSD and a decrease in overall alpha power, although both effects were not observable with all stimulus types. Additional effects were decrease in alpha coherence in frontal lobe, or increase in beta coherence in parieto-occipital lobe. As for "happy" vs. "unhappy" epochs, changes in bicoherence were also observed, but with less agreement among subjects than second order spectral features.

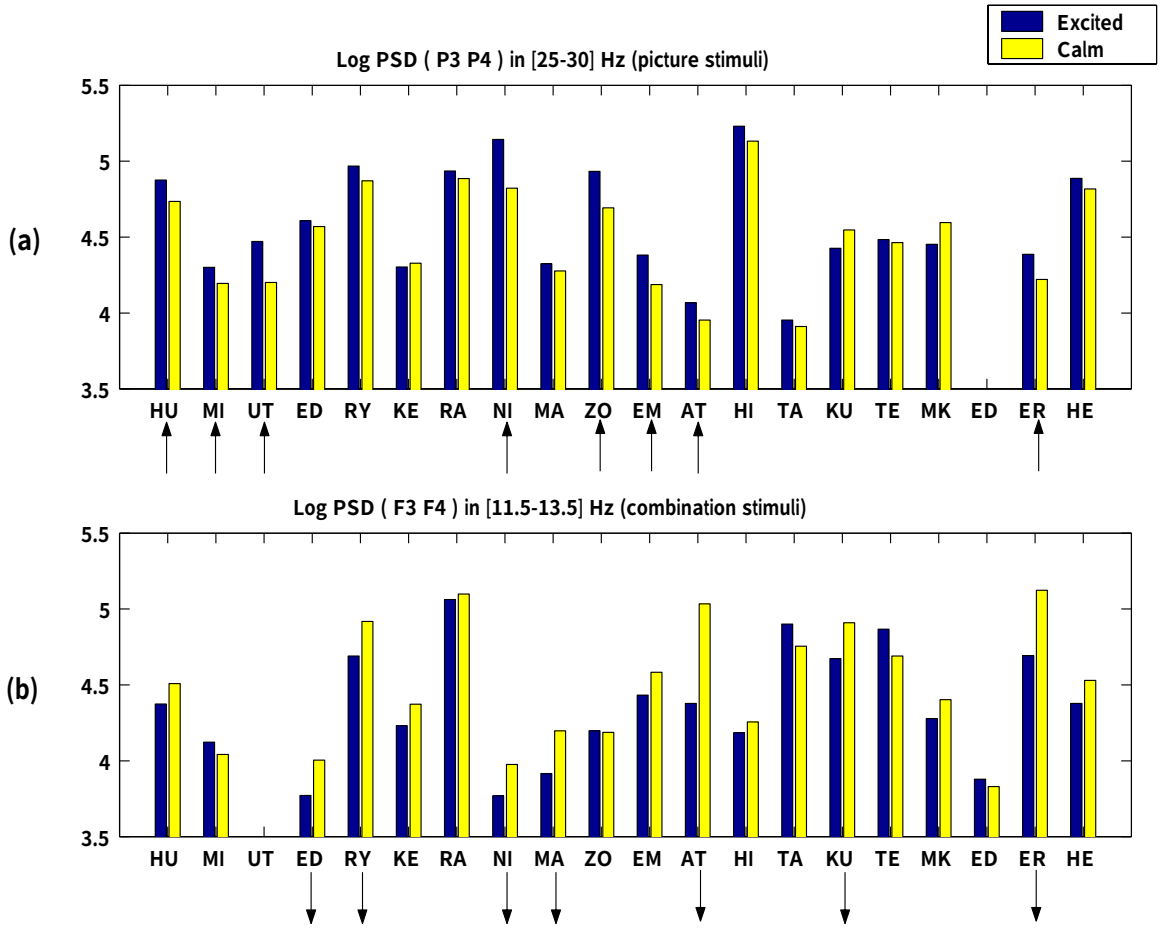


Figure 4.9: Significant changes between "excited" and "calm" epochs for picture (a) and combination (b) stimuli. Subjects showing a significant increase (resp. decrease) accompagnying arousal are indicated with an arrow.

Pictures. Figure 4.9 (a) shows the mean log-transformed PSD at parietal (P3·P4) electrodes in the [25-30] Hz (medim beta) band, for excited and calm epochs. Log PSD increased significantly with arousal (i.e. excitation) for 7 subjects at P=0.10 (4 subjects at P=0.05), out of 19. In addition, (1) central alpha (9.5-11.5 Hz) coherence between F7 and F8 significantly decreased for 7 subjects out of 13, and (2) medium beta (25-30 Hz) coherence between P3 and P4 and between O1 and O2 increased significantly for 6 subjects out of 19, and 6 subjects (vs. 1) out of 16, respectively.

Sounds. The change in parietal PSD in beta was not observable for sound stimuli. Instead, a generalized decrease in alpha power was clear. In particular, the log PSD at left parietal electrode (P3) in the [9.5-11.5] Hz (central alpha) band was significantly lower for excited epochs compared to calm epochs, for 8 subjects (vs. 1) at P=0.10 (4 subjects at P=0.05), out of 20 subjects. Besides, right frontal intra-lobe (F4·F8) coherence in low alpha (7.5-9.5 Hz) was significantly lower for excited epochs, for 7 subjects (vs.1) out of 18.

Combinations. With combination stimuli, excitation was characterized by a decrease in high alpha over all lobes, in both hemispheres. Figure 4.9 (b) shows the mean log-transformed PSD at mid-frontal (F3·F4) electrodes in the [11.5-13.5] Hz (high alpha) band. This quantity was significantly lower for excited epochs than for calm epochs, in 7 subjects at P=0.10, out of 19. No subject showed the opposite trend.

An increase in beta coherence was observable between electrodes located in opposed hemispheres in all lobes. The most marked effect was a significant (P=0.10) increase of coherence between P3 and O2, and between P4 and O1, in the [18-22] Hz (median beta) band, for 7 subjects (vs. 1) out of 20, and 8 subjects (vs. 1) out of 16, respectively. This effect was still significant at P=0.05 for 6 and 4 subjects, respectively.

4.2.4 Discussion

EEG features correlated with emotional experience, common to several subjects

Individual differences. The fact that different people have very different EEG and that the same reported feeling may lead to dissimilar EEG patterns in different individuals has already clearly appeared in the preliminary experiment reported in the beginning of this chapter. Also in this experiment, individual differences made it difficult to draw general rules about how emotion influences the EEG. Inter-subject diversity can be formulated at two levels:

1. *Emotional report:* two people feeling exactly the same way are likely to report experienced emotion differently.

2. *Biological realization:* assuming two people could feel exactly the same emotion, it is probable that this emotion will be neurophysiologically realized in a different way in the two individuals.

The first issue has strongly motivated the choice for a simple and universal way to represent emotions. However, notwithstanding the choice of an extremely simple representation, several subjects reported the difficulty to evaluate "how they felt" when being presented a stimulus. Besides, figure 4.6 clearly shows that different persons sometimes tend to use the same system in very different ways.

Concerning the variety of EEG patterns, a look at figures 4.7, 4.8 and 4.9 indeed shows that inter-individual differences are of a different order of magnitude than the difference due to different emotional states for a given subject.

Beyond diversity. As stated in section 4.2.1, the objective of this first analysis was to concentrate on EEG influences of emotion that are *common to several subjects*. This ambitious objective partly justifies the procedure adopted in this analysis (section 4.2.3), based on two-class subdivisions. Considering the great variety of possible emotions, one might have been

tempted to divide epochs into more than two classes. However, because of the issues enumerated above, this would probably have prevented to find any agreement between subjects.

Similarly, the use of a quite lenient degree of significance in most hypothesis tests (P=0.10) was motivated by the hope to discover even slight EEG changes related to emotion, at the risk (10%) to also detect spurious changes.

Emotion and EEG: results interpretation

In this section, an attempt is made to interprete results presented above and to related these findings with neurophysiological facts and with other findings.

Before entering into detail, I would like to comment the difference between feature changes detected with picture or combination stimuli (phases I and III), on the one hand, and with sound stimuli (phase II), on the other hand. The main reason for this difference is the fact that the former were recorded with open eyes, and the latter with closed eyes. Closed eyes almost automatically triggered a high ampltitude alpha component (probably emanating from the occipital lobe, but spreading across the whole head), which intermixed with alpha components of other origins, masking some effects normally observed with open eyes, or on the contrary, enhancing some effects not observable with open eyes. Another difference between phases I and III and phase II is the segment length: 10 s in the former case and 5 s in the latter. As discussed in section 3.2.1, spectral estimation performed over different time intervals may capture different brain processes.

Although this analysis handles separately the three subdivisions of epochs, it is important to keep in mind that these subdivisions are closely related. This is obvious for the first two subdivisions, since the class of emotional epochs contains both happy and unhappy epochs. Moreover, the third subdivision according to the arousal dimension is not completely independent from the first two subdivisions, because some subjects present a significant (either positive or negative) correlation between valence and arousal (see figure 4.3).

Valence: "emotional" vs. "neutral" epochs

The effect of emotional experience on the EEG can be summarized as: (1) an increase in the ratio of beta by alpha activity in the frontal lobe, and (2) an increase in beta activity at the parietal lobe.

The former phenomenon can be paraphrased, roughly speaking, by saying that emotional experience triggers *frontal lobe activation*. This is in agreement with findings presented in section 2.1.3, regarding the important role played by the prefrontal cortex in emotion, particularly, as an "affective working memory", independently of valence (that is, related to positive as well as negative emotions).

Increased beta activity was observed in both frontal and parietal lobes when emotion was reported. More precisely, the observed activation was related to low beta (below 25 Hz) for phases I and III, and to high beta (30∼35 Hz) for phase II (high beta activation was also observed during phase III). Several other studies [11, 17] report increased unilateral beta activity related to happy or sad emotions. It is therefore not surprising that emotional experience (positive or negative) is characterized by a general increase in beta activity (regardless of the side).

Valence: "happy" vs. "unhappy" epochs

In phases I and III, the most important effect observed for happy epochs was a significantly higher coherence betwen F7 and F8 in the central alpha band, compared to unhappy epochs.

This phenomenon confirms findings of Hinrichs and Machleidt [26], who report a decrease of alpha coherence during sadness and an increased alpha coherence during joy. Interestingly, this effect was observed almost exclusively between F7 and F8, rather than between F3 and F4, for example. This could be explained by the fact that F7 and F8 are the closest electrodes to the limbic system, hence reflecting direct influences of bilateral projections from limbic components (mainly the amygdala) to the neocortex.

In phase II, the increased coherence between F7 and F8 was not observed. Instead, the PSD in central alpha was significantly higher for happy than unhappy epochs. In the same line as the brief discussion concerning the differences between phases I and III and phase II, the following explanation is proposed. The high amplitude alpha rhythm triggered by eye closing, originating from the occipital lobe and propagating to the frontal lobe, was masking the alpha rhythm observed in phases I and III, of another origin (possibly from the limbic system). Besides, the higher PSD in central alpha, not observed in phases I and III, probably resulted from the enhancement of the occipital alpha rhythm due to happy emotion (or reduction of the occiptal alpha rhythm due to unhappy emotion).

In happy epochs of phase II, an increased coherence between parietal electrodes in medium beta (27.5∼30 Hz) was observed for almost half of the subjects, and accompanied by an increased right parietal PSD in medium beta. The latter observation compares with findings of Schellberg et al. [69], Stenberg [70], who reported higher right temporal beta activity in positive than negative emotion. Increase in right parietal PSD in high beta was also observed in phase III, but not in phase I.

Other changes in PSD between happy and unhappy epochs were observed, but the lack of agreement between subjects (usually not more than 5 or 6) suggest that PSD alone are not sufficient to reliably discriminate between positive and negative emotion.

Changes observed in bicoherence mainly concerned auto-bicoherence. Among other, the increase of auto-bicoherence at some frontal electrodes around (10,39) Hz^2 during happy emotion could be interpreted as the increased influence of frequency component around 10 Hz on a higher beta component around 39 Hz, in happy emotion compared to unhappy emotion. These quadrature phase coupling effects however need to be further investigated in order to be fully explained.

Arousal: "excited" vs. "calm" epochs

While higher frequency EEG, commonly termed as "beta activity" may indicate different kinds of cognitive processes, consistent findings associate the beta rhythm with arousal and attention [12]. Increased beta (25∼30 Hz) activity observed at the parietal lobe in excited epochs is therefore not surprising. However, this effect was observed only during phase I, were visual attention was probably the most important. Rather than a simple increase in beta power, the effect of arousal during phases I and III was characterized by an increased coherence between left and right parieto-occipital lobes. The importance of the inferior parietal lobe in attention and arousal was mentioned in section 2.1.3. Besides, the occipital cortex has been shown to be functionally activated by emotional arousal [45]. The complementarity of processes realized by the parietal and occipital lobes justifies the increased coherence observed between them in excited epochs.

Since phase II did not involve visual attention, it is normal that the phenomena described before were not observed in epochs recorded with closed eyes. Instead, the left parietal alpha power was lower for high arousal. Again, the alpha power measured at the parietal lobe with closed eyes essentially originated from the occipital lobe. The excited state had the effect to reduce the right alpha component, due to the activation of this region. In addition, lower

alpha PSD in excited epochs was not limited to the parietal lobe, but occurred (less markedly, though) over the whole brain in phase II and over the frontal lobe in phases I and III. The latter observation indicates an increased activation of the frontal lobe in high emotional arousal conditions.

4.2.5 Conclusion

The current experiment studied general influence of emotion on the EEG. Shortcomings of the preliminary experiment have been successfully overcome:

1. Short EEG epochs were recorded, involving simple mental task (i.e. a particular emotion).
2. Recorded epochs included reference EEG, corresponding to neutral emotion.
3. No movement was required from the subject, resulting in few muscle artifacts. Remaining artifacts were carefully removed.

The objective of this experiment was to find EEG features changes correlated with variations in emotional experience. Several features were successfully detected that presented consistent changes with particular aspects of emotion. These changes can be coarsely summarized as follows:

1. Strong emotional feeling (regardless of its sign) was associated with activation of the frontal cortex.
2. Positive emotion presented higher frontal coherence in alpha and higher right parietal beta power, compared to negative emotion.
3. Excitation was characterized by higher beta power and coherence in parietal lobe, and lower alpha activity.

Chapter 5

Neural Networks for Emotion Expression

5.1 Introduction

Objective

In chapter 4, influences of emotion on the EEG were discussed. Features were extracted from the EEG and several features varying with emotional state were presented. These features were significantly related with emotion for several subjects. This chapter focuses on the second step in the design of an "emotion expression system", stated in section 4.2.1:

> Design a system capable of efficiently learning a particular mapping between EEG features and emotion.

Let us remind that the same system could hardly be used by different users since, although several EEG features have been shown to reflect emotion better than other, individual differences require that the system be tuned to a particular user in order to be efficient.

An "emotion expressing system" is basically a special-purpose brain-computer interface (BCI) and should therefore possess the following characteristics:

- Capacity to adapt to the user, not obliging the user to adapt to the system. The system should therefore possess the ability to *learn* how the EEG of a particular patient are mapped onto emotion. It is worth saying this mapping is non-linear, due to the high interdependance of different EEG features and to the complexity of brain processes that generate emotion.

- Robustness, that is, minimization of error rate in emotion estimation from EEG features. This is however not as critical as in a BCI which commands a wheelchair, for example. Besides, it is desirable to have an idea about the uncertainty of estimated emotion.

- Rapid response. The system should be able to function in real time, so that it can complement other channels of communication. This issue will be further adressed in chapter 6.

The following sections demonstrate how neural networks can be used to efficiently learn the non-linear mapping between EEG features and emotion, and provide robust estimation of the emotion state.

Working strategy

Many approaches can be adopted to design a neural network for solving the problem described above. In this study, I made the following practical choices:

1. Neural networks were always trained *for a particular subject*.
2. *Inputs:* The same EEG features were selected for all subjects.
3. *Outputs:* Emotional state was restricted to 3×2 classes: happy-neutral-unhappy and excited-calm.
4. Neural networks were all *multi-layer perceptron (MLP)*.

Justification

The first choice, consisting in designing a system tuned to a particular subject, was motivated in the preceding section.

The decision to consider a fixed set of features (the same for all subjects), was taken because of the existence of a set of features that present significant variations with emotion for several subjects, as presented in chapter 4. Although it could be argued that features presenting agreement among individuals are maybe not the most suited for each particular subject, I decided to orient the discussion in this chapter towards the neural network *mapping*, rather than the quality and relevance of the inputs. Moreover, the use of non-optimal inputs (i.e. features) for some subjects allowed to test the robustness of neural networks to less important (or irrelevant) inputs.

The third choice can be re-formulated by saying the focus was set on *classification* rather than on a continous *function approximation*. This decision was motivated by the high uncertainty of emotional ratings. As explained in section 4.2.2, emotional ratings were collected on a 9-point scale. Intuitively, one would be tempted to think that the more points we have on scale, the more precise the result will be. However, Hirota [27], addressing the problem of vagueness in questionnaire enquiries from an informational point of view, provided the following interesting result. It is in a 3-point scale that the information content is the highest. The "informational weight" decreases then for 4, 5 and more choices, and is eventually lower for a 7-point scale than for a binary choice !

Finally, the choice of MLP was motivated by the following considerations: (1) it allows supervised learning and can be used for classification, (2) rather than RBF-like networks, for example, it is based on asymmetric non-linearities (e.g. sigmoid function), (3) it has been extensively studied and its theoretical fundations became relatively well known. The second point means that MLP allows to partition the input space by defining "half-spaces" rather than "clusters", which was intutively juged more appropriate to learn the mapping of EEG features onto emotions (but further verification would be interesting).

Practical issues

The training of an MLP to learn the mapping of EEG features onto emotions raises the following issues:

- *Training examples.*
  1. Small training set. As explained in section 4.2.2, about 30 EEG epochs were recorded and corresponding ratings collected for each type of stimulus, for each subject. Taking into account that some epochs were rejected because of muscle artifacts (see section 3.1.3), this yielded training sets of about 20 to 30 samples, which is extremely small for training a neural network. Particularly, the number of needed examples is expected to increase (exponentially) with the dimensionality of the input space (see section 5.2.2). The main concerns of using small training sets are (1) the *generalization ability* of the network may be poor if the network uses only few examples for training, and (2) the number of examples that can be spared for *testing* is limited, making it difficult to evaluate network performance.
  2. Noisy data. Both inputs and outputs of the examples presented to the network are subject to noise: (1) inputs are *estimates* of spectral characteristics of the EEG, of which the precision is limited by the segment size, and (2) outputs are emotional ratings, dependent on subjectivity and sensitivity (the reduction of ratings to a small number of classes however allowed to strongly reduce this uncertainty).

- *Choice of an MLP architecture.* The choice of a particular network architecture is sometimes also refered to by the more generic appellation *"model selection"*. The definition of an MLP architecture essentially involves choosing (1) the number of hidden layers, and (2) the number of neurons in each hidden layer. This study was limited to 3-layer MLP (that is, with one hidden layer), because it is conjectured that the mapping is not too complicated.

- *Performance evaluation.* The choice of a neural network architecture requires to be able to evaluate the actual performance of a trained network. The most common evaluation method consists in using a *test set*. As previously mentioned, the small number of data examples available for the current application does not favor this approach.

- *Choice of a training algorithm.* Besides the famous back-propagation (BP) algorithm, plenty of methods exist to train MLPs, several of which can be regarded as general optimization methods. These algorithms differ from each other in several aspects, including computational complexity, rapidity of convergence, memory use, number of parameters, etc.

- *Parameter setting.* Closely related to the choice of a training algorithm is the necessity to find suitable values for the training parameters, so that the convergence actually occurs in a reasonable time, and that training objectives are met. Because this issue is relatively cumbersome, an effort was made to reduce as much as possible the number of parameters to be adjusted "by hand".

- *Output uncertainty.* Classically, a trained neural network is able to estimate output values for new, unseen input patterns. In this application, it is also desirable to be provided with an estimate of the confidence that can be given to the estimated output.

5.2 Methods

5.2.1 Definitions

Before presenting the methods used in this analysis, some notations are introduced.

Inputs and outputs

The selection of EEG features correlated with emotion was discussed in section 4.2.4. A subset of, say, p features correlated with emotion was used as input of the neural networks studied in this chapter (the generation of data sets is further described in section 5.2.2). For each subject, a vector of p features was estimated for every EEG epoch i:

$$\mathbf{x}_i = [x_{i1},\, x_{i2},\, \ldots,\, x_{ij},\, \ldots,\, x_{ip}] \qquad i = 1,\, \ldots\, n. \tag{5.1}$$

Each *feature vector* $\mathbf{x}_i$ (the index i will sometimes be omitted for clarity) belongs to one of the three valence classes (happy: H, unhappy: U or neutral: N), and to one of the two arousal classes (excited: E or calm: C). The appartenance of $\mathbf{x}_i$ to these classes is represented by a *label vector* of five components: $\mathbf{y}_i = [\mathbf{y}_{i,val}\, \mathbf{y}_{i,aro}] = [y_{iH},\, y_{iU},\, y_{iN},\, y_{iE}\, y_{iC}]$, where y_{ic} takes the value 1 if $\mathbf{x}_i$ belongs to class c. Obviously, only one component of $\mathbf{y}_{i,val}$ and one component of $\mathbf{y}_{i,aro}$ will take the value 1 for a given feature vector.

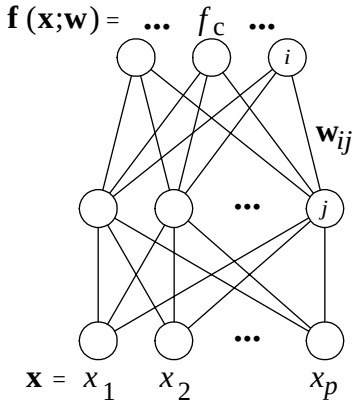


Figure 5.1: Multi-layer perceptron for classification.

Neural network

In the following, MLPs (see figure 5.1) are trained to *classify* feature vectors, that is, given $\mathbf{x}_i$, to produce an estimate $\hat{\mathbf{y}}_i$ of the label vector $\mathbf{y}_i$. In practice, separate networks were used for estimating valence and arousal class membership probabilities $\mathbf{y}_{i,val}$ and $\mathbf{y}_{i,aro}$. The function realized by each network is defined by the following equations (where c takes the value H, U, N for valence and E, C for arousal estimation):

$$\begin{aligned} \mathbf{f}(\mathbf{x};\mathbf{w}) &= [\ldots,\, f_c(\mathbf{a}),\, \ldots], \\ a_c &= \sum_{k \in W_{hid}} w_{kc}\; g\left(\sum\nolimits_{j \in W_{in}} w_{jk}\, x_j\right), \end{aligned} \tag{5.2}$$

where $\mathbf{w} = [w_{ij}]$ is the *weight vector* (w_{ij} links neuron i to neuron j), $\mathbf{a} = [a_c]$ is the vector of output layer activations, and W_{in} and W_{hid} are sets of indices of the neurons belonging to the

input layer and hidden layer, respectively. Two types of nonlinearities are used: $g(\cdot) = \tanh(\cdot)$ in the hidden neurons, and $f_c(\cdot)$ in output neurons. The latter is the *'softmax'* function, defined as:

$$f_c(\mathbf{a}) = \frac{\exp(a_c)}{\sum_{c'} \exp(a_{c'})} \tag{5.3}$$

which Bishop [5] recommends to use when the network outputs are to be considered as probabilities of class membership, because it ensures that the sum of probabilities is correctly normalized.

Binary classification. When there are only two classes (this is the case, e.g. for the "arousal network", trained to classify features vectors either in the "excited" or in the "calm" class), a single output neuron is sufficient since the two class membership probabilities sum to 1. Hence, for binary classification the classical sigmoid activation function was used, which output is restricted to the interval [0, 1]:

$$f(a) = \frac{1}{1 + \exp(-a)}. \tag{5.4}$$

Classification decision. According to equations 5.4 and 5.3, each component f_c of the output $\mathbf{f}(\mathbf{x}; \mathbf{w})$ of the neural network takes values between 0 and 1. Since we are interested in an estimate of the network performance e.g. in terms of correct classification rate, we eventually need to decide the outcome of the classification for each input, that is, to convert the network output into a binary vector. This is simply done by selecting the most probable class. In the case of a binary classification, this comes down to setting the only output to 1 if $f(\mathbf{x}; \mathbf{w}) > 0.5$ and to 0 otherwise. In the general case (eq. 5.3), a vector $\mathbf{f}' = [\ldots, f'_c, \ldots]$ is constructed so that only the component corresponding to the most probable class is set to 1, and the remaining components are set to 0:

$$f'_c = \begin{cases} 1 & \text{if } f_c = \max_{c'} f_{c'}(\mathbf{x}; \mathbf{w}) \\ 0 & \text{otherwise.} \end{cases} \tag{5.5}$$

Note that this transformation was also used to decide the output of linear regression classifiers (though, in that case, the output is not guaranteed to be comprised between 0 and 1).

Committee of networks

Instead of using a single network for learning to model the probability density of multiple classes, the combination of neural networks that cooperate to predict the final classification was also investigated. Such group of cooperating networks is called *"committee"* in the following.

A committee C (see figure 5.2) is a set of networks, each of which realizing *one component* f_c of the input-output relationship described in equation 5.2. Each network is an MLP with p inputs and one output. The output $f_c(\mathbf{x}; \mathbf{w})$ of the cth network approximates the probability that a feature vector $\mathbf{x}$ belongs to class c (each network of the committee therefore uses the sigmoid activation function).

Classification decision. Deciding the outcome of the classification is essentially done as in equation 5.5. In general, a committee like the one depicted in figure 5.2 *does not guarantee* that the class probabilities sum to one. If one would be interested in the actual class membership probability, a normalization step should be performed, but it is not necessary in this case, since

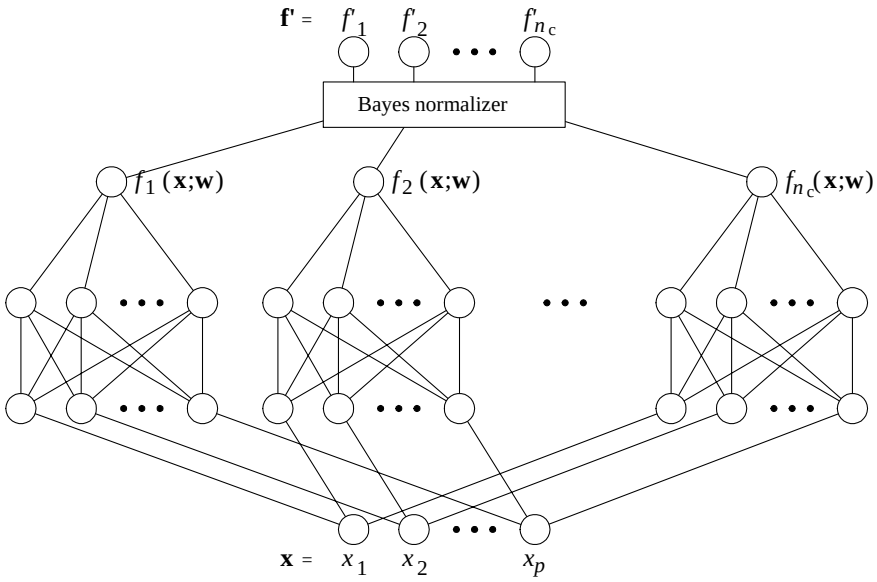


Figure 5.2: Committee of neural networks for classification.

we are interested in the most probable class. Another problem is that the committee C is not guaranteed to respect the prior probability of each class, which was implicitly constrained by using a single vector with the softmax function. In other words, the prior probability of class c should be the sum *over all patterns* of the probability to belong to class c:

$$P(C_c) = \sum_{i=1}^{n} P(C_c|\mathbf{x}_i). \tag{5.6}$$

Bishop [5] proposes a simple solution to this problem, which consists in dividing each f_c by the prior probability of class c resulting from training and multiplying the result by the actual (desired) prior probability of class c. The "Bayesian normalizer" in figure 5.2 performs this scaling and determines the binary output vector $\mathbf{f}'$, according to equation 5.5.

The remaining of this section presents the methods adopted for data pre-processing, network training, and evaluation of trained networks.

5.2.2 Data set generation

Data sets D_S^t were constituted for each subject S and type of stimulus t, containing *patterns* $\mathbf{d}_i =< \mathbf{x}_i, \mathbf{y}_i >$. In practice, the construction of these data sets consisted in three steps: (1) construction of label vectors, based on emotional classification of epochs, (2) selection of p EEG features, and (3) feature vector pre-processing.

Construction of emotional classes of epochs

Essentially the same approach was adopted for determining classes of epochs as in section 4.2.3. That is to say, epochs were classified according to subject emotional ratings collected during the experiment. The main difference with the preceeding chapter is that, for each subject, all available epochs (not corrupted by muscle artifacts) were included.

Three *valence classes* were constructed: happy, unhappy and neutral. The repartition of epochs in these three classes was always symmetrical about the neutral state (5 on the rating scale), but differed between subjects. The different repartitions of epochs according to emotional ratings (varying from 1 to 9) are summarized in table 5.1. The choice between repartition 1, 2 and 3 was made so that the three classes are best balanced.

Table 5.1: Different repartitions of EEG epochs according to emotional valence ratings.

	UNHAPPY	**NEUTRAL**	**HAPPY**
Repartition 1	1, 2	3, 4, 5, 6, 7	8, 9
Repartition 2	1, 2, 3	4, 5, 6	7, 8, 9
Repartition 3	1, 2, 3, 4	5	6, 7, 8, 9

Two *arousal classes* were constructed, excited and calm, by dividing epochs classified according to excited ratings in two sets of comparable size. As a general rule, EEG epochs with the same emotional rating were always included in the same class.

Feature selection

Before deciding *which* features to take, I faced the problem of determining *how many* features should be selected.

The particularity of the problem dealt with in this study is the very small number of patterns available for each subject: as a consequence of experimental choices, only $n = 25{\sim}30$ patterns were finally available. Unlike many other problems, the number of patterns n was hence a limiting factor, which strongly influenced the choice of both the number of features p and the network complexity.

Number of training data vs. input space dimensionality. Before considering any particular classification problem of labeled data, we are interested in determining how many *unlabeled* data are necessary to give sufficient information about the relevance of each input dimension. When $p = 1$, at least two points are needed to describe a single feature. Speculating that the density of patterns should remain constant when the feature space dimensionality p increases, in two dimensions 4 patterns would be required, in three dimensions 8, etc. The exponential grow of needed patterns with the dimension of the input space has been termed the *"curse of dimensionality"* [5, pp. 7–8]. In this application where n is about $25{\sim}30$, requiring a minimum of 2 patterns per dimension of the input space results in limiting the number of features p to 4 or 5 (because $2^4 = 16$ and $2^5 = 32$ samples, respectively). Requiring three patterns per feature results in $p = 3$ ($3^3 = 27$).

This argumentation is not formal but intuitively justifies the arbitrary choice made to limit p to 3, seen the small number of patterns available.

Number of training data vs. network complexity. Assuming now we dispose of n *labeled* patterns of p features, it is interesting to consider the problem of determining how many patterns are necessary to ensure that the classification can be learned successfully by a given network. Seen from another angle, we may want to determine the maximal network complexity allowed for given n and p, in order to guarantee generalization ability. Model selection for the current application is discussed in section 5.3.2.

Attempts have been made to determine bounds on the number of patterns necessary to sufficiently constraint a network of a given size. However, these theories based on the *Vapnik-Chervonenkis dimension* (summarized in [5, 66]) are known (1) to be asymptotic theories (i.e. valid for large size networks and data sets), and (2) to yield worst-case bounds (that is, we may hope in practice that a good generalization would be achievable with fewer patterns than predicted).

Selection of p features. Based on results of the main experiment exposed in section 4.2, 9 features were first selected, the same for all subjects. Selected features were those presenting the highest agreement among subjects (i.e. the highest number of subjects presenting the same significant variation with emotion), restricted to PSD and coherence. Within the nine features, 2 were related to the existence of emotional feeling (emotional vs. neutral epochs), 5 were related to valence sign (happy vs. unhappy epochs) and 2 were related to arousal (excited vs. calm). Selected features are summarized in table 5.2 for the three kinds of stimulus.

Table 5.2: Selected features, for each type of stimulus.

Picture	**Sound**	**Combination**
valence intensity (emotional vs. neutral)		
1. Log PSD (P3 P4) in [15-25]	1. Log PSD (F3 F4) in [30-40] –log PSD (F3 F4) in [8-10]	1. Log PSD (F3 F4) in [15-25] –log PSD (F3 F4) in [11.5-13.5]
2. Log PSD (F3 F4) in [15-20] –log PSD (F3 F4) in [9-11]	2. Log PSD (P3 P4) in [30-35] –log PSD (P3 P4) in [7.5-13.5]	2. Log PSD (P3 P4) in [30-40]
valence sign (happy vs. unhappy)		
3. Log PSD (F7) in [11.5-13.5]	3. Log PSD (F4) in [35-40]	3. Log PSD (F4) in [7.5-9.5]
4. Coherency (F7 F8) in [10-13]	4. Log PSD (P4) in [9-11]	4. Log PSD (F7) in [15-20]
5. Coherency (F3 F8) in [7.5-13.5]	5. Coherency (P3 P4) in [27.5-30]	5. Coherency (F7 F8) in [9-11]
6. Log PSD (P3) in [5-8]	6. Log PSD (P3) in [9-11]	6. Log PSD (F4) in [30-40]
7. Log PSD (P4) in [5-8]	7. Log PSD (P4) in [25-30]	7. Coherency (F3 F8) in [9-11]
arousal (excited vs. calm)		
8. Log PSD (P3 P4) in [25-30]	8. Log PSD (P3) in [9.5-11.5]	8. Log PSD (F3 F4) in [11.5-13.5]
9. Coherency (F7 F8) in [9.5-11.5]	9. Coherency (F4 F8) in [7.5-9.5]	9. Coherency (P4 O1) in [18-22]

In a second selection step, only $p_{val} = 3$ and $p_{aro} = 2$ features were used among the nine.

Feature pre-processing

Features selected in that way were then linearly transformed by subtracting the mean and scaling each feature independently, within each data set D_S^t. By means of this pre-processing, it was ensured that each feature $\mathbf{x}^j = [x_{1j}, \ldots, x_{nj}]^T$ be distributed around $\mu_j = 0$ with variance $\sigma_j^2 = 0.25$ (for $j = 1, \ldots, p$).

5.2.3 Training and testing

Error function

The widely used sum-of-square error is not the most appropriate for classification problems [5], mainly because it is derived from maximum likelihood on the assumption of Gaussian distributed target data, which is certainly not the case for the binary coding adopted here.

Instead, the *cross-entropy* error was adopted:

$$E_{CE} = -\sum_{i=1}^{n}\sum_{c} y_{ic} \ln f_c(\mathbf{x}_i; \mathbf{w}). \tag{5.7}$$

This error function is simply the negative logarithm of the likelihood of the conditional distribution of all patterns:

$$\begin{aligned} E_{CE} &= -\ln \prod_{i=1}^{n} p(\mathbf{y}_i|\mathbf{x}_i), \\ p(\mathbf{y}_i|\mathbf{x}_i) &= \prod_{c} \left(f_c(\mathbf{x}_i; \mathbf{w})\right)^{y_{ic}}. \end{aligned} \tag{5.8}$$

The absolute minimum of this error function with respect to the $f_c(\mathbf{x}_i; \mathbf{w})$ occurs when $f_c(\mathbf{x}_i; \mathbf{w}) = y_{ic}$ for all values of i and c. At the minimum, the function takes the value $E_{CE,\min} = -\sum_{i=1}^{n}\sum_{c} y_{ic} \ln y_{ic}$, which become 0 if all components of the label vectors are either 0 or 1.

In the case of a binary classification, the cross-entropy error (associated with the classical sigmoid activation function of equation 5.4) reduces to [5]:

$$E_{CE} = -\sum_{i=1}^{n}\{y_{ic} \ln f(\mathbf{x}_i; \mathbf{w}) + (1 - y_{ic}) \ln(1 - f(\mathbf{x}_i; \mathbf{w})\}. \tag{5.9}$$

Generalization

A particularly important issue in the current application was to generate a network that does not only learn to classify correctly the examples that are presented, but that also has the ability to classify new, unseen features vectors, that is, the ability to *generalize*.

When very few patterns are used for training, it is easy for even a simple network to learn them all without error. However, it is possible that the network does not learn the underlying model which generates the data, but only the examples it has been shown. Hence, there is often a compromise between the complexity of the mapping realized by the network and its ability to generalize. Especially when there are few patterns and when they are subject to noise (like it is the case here), it is important to take drastic measures against *overfitting*.

Several methods exist to improve the generalization ability of a network, including regularization, early stopping and training with noise (jitter) [5]. Early stopping presents the inconvienients that it relies on a *validation set*, independent of the training and test sets, and that it may overfit the validation set. It was thus rejected, because it could not be afforded to save enough data in order to verify when training should be stopped. Training with noise consists in replacing each training patterns by several version of this pattern that have been perturbed with noise. Problems with this technique include the difficulty to determine the quantity of noise that should be added, and the fact that it often slows down training, requiring a small learning rate [66]. Besides, training with noise is in its effects, very close to a simple weight scaling (smoothing).

In this study, I opted for *weight decay*, a kind of regularization consisting in adding a penalty term to the error function in order to avoid the network weights to become too large:

$$E = E_{CE} + \alpha \frac{1}{2} \sum_{i,j} w_{ij}^2. \tag{5.10}$$

Krogh and Hertz [41] showed why a simple weight decay can improve generalization in feed-forward networks: (1) by suppressing irrelevant components of the weight vector, and (2) by suppressing some of the effects of static noise on the targets. On the other hand, in the context of Bayesian regularization (but this should also apply for a simple weight decay), Penny and Roberts [62] mention the importance to start training with almost no regularization, in order to allow the network to find interesting structure in the data before any regularization takes place. I noticed this effect and propose a two-phase training scheme, as follows:

1. Train the network without any regularization until it reaches a point where the error does not decrease anymore.

2. Retrain the network with a fixed weight decay factor.

Training algorithm

The task of training a neural network can be regarded as an optimization problem, where the objective is to find a weight vector $\mathbf{w}^*$ for which the network error function E is minimum. Several algorithms exist for solving this problem. In this study, I used the scaled conjugate gradient (SCG) algorithm, proposed by Moller [52], which makes use of second-order derivative information in complement to the simple error gradient used, for instance, in back-propagation (BP).

The choice the SCG algorithm was motivated essentially by its rapidity of convergence. To illustrate this choice, results are reported of a comparison between the SCG and other training algorithms.

The evolution of the network error E accross 10,000 training iterations[1] is shown in figure 5.3, for the three algorithms: back-propagation with fixed learning rate 0.0001 (BP), back-propagation with one learning rate parameter for each weight component adapted according to the delta-bar-delta scheme (BP-DBD) (see details in [66, pp. 137–139]), and scaled conjugate gradient (SCG). Training was performed on a training set 156 five-dimensional patterns. Five networks with different initialization were trained three times, once with each algorithm. The average training error for the five networks is represented by the solid line. The error of the "best" of the five networks (in term of the final error) is also shown for each algorithm.

Clearly, the conventional BP algorithm is much slower to converge than the two other algorithms. BP-DBD converges more quickly and the "best network" error is even comparable to the best result obtained with the SCG after 10,000 iterations. The average result is however quite worse than the SCG. Moreover, the SCG shows a quick convergence, yielding an error close to the final error, after only 1000 iterations.

[1] By "iteration", it is meant one evaluation either of the error function (including those in line search for SCG), or of its gradient.

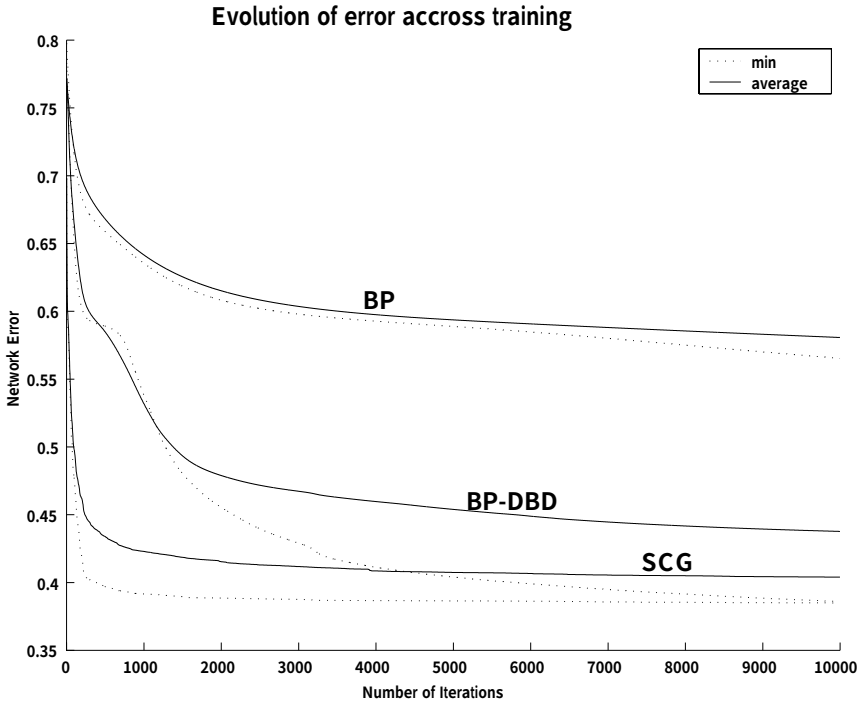


Figure 5.3: Comparison between three training algorithms: back-propagation with fixed learning rate parameter (BP), back-propagation with multiple learning rates adapted according to the "delta-bar-delta" scheme (BP-DBD), and scaled conjugate gradient (SCG).

Testing

The objective of testing is to evaluate the actual performance of a trained network on new, unseen data, that is, to estimate its *generalization ability*.

The most common technique for evaluating trained neural networks consists in saving a portion of available patterns, to use remaining patterns for training, and to compute the error on previously saved data. For this application, however, the number of available patterns being very small, it could not be afforded to put aside, say, one half or one third of the patterns for testing. Even, assuming e.g. 10 patterns (out of 30) were saved for testing, the test error would have been very inaccurate since limited to multiples of 10%. Obviously, the objectives of high quality training and of testing accuracy seem contradictory when few patterns are available, since the former requires many patterns for training and the latter requires many patterns for testing.

Cross-validation. A classical way to tackle this problem is to use *cross-validation* (CV). The basic idea k-fold CV is to divide available data into k subsets of equal size. The network is then trained k times, each time leaving out one of the subsets from training and using the omitted subset to compute the error test [68]. The averaged error over the k networks gives an estimation of the generalization error.

In *"Leave-One-Out" (LOO) CV*, each "fold" contains exactly one sample. Hence, training is performed n times, each time keeping one pattern for testing and training the network on the remaining $n - 1$ patterns. Test error in this study was estimated most of the time with LOO-CV. Some tests were also carried out with what could be called *"Leave-Two-Out" (LTO)*

CV although, strictly speaking, the data set is not partitioned in folds. In LTO CV, 2 patterns are left out each time, allowing to train the network C_2^n times (where C_2^n denotes the number of pairs out of n patterns). This was clearly more time consuming but yielded more precise and reliable generalization error estimate.

Bayesian evidence. The Bayesian evidence framework proposed by MacKay [46] provides a unified theoretical view of neural networks. Estimation of the evidence, central element of this framework, represents the conditional probability of a particular model (i.e. network), given the data. This allows to classify trained networks according to their "evidence", the most probable networks theoretically providing the best generalization. For this particular application, the evidence framework would have allowed to deal with three important issues: regularization, estimation of the generalization error, and model selection.

Unfortunately, though I considered the possibility of applying Bayesian methods in this analysis, I eventually had to renounce to this idea, because of the questionable validity of underlying hypotheses when the sample size is too small. Confirming this doubt, Penny and Roberts [62] showed empirical indication that the evidence framework for model selection becomes meaningful only when sufficient examples are available (about five or ten times the number of network weights).

5.3 Results and discussion

Preliminary remarks

Results in this section will be presented only for a limited number of subjects. The main reason for this selection is that the three arbitrarily selected features did not allow reliable classification for all subjects. Nevertheless, due to the limited time, I decided to concentrate on the training and evaluation of neural network classifiers, rather than on the perfect selection of EEG features, suited for each particular subject.

Most results concern the valence scale, because I found the three-class problem more interesting for the discussion. Clearly, the same techniques can be directly applied to the easier two-class problem on the arousal scale (even with some simplifications).

5.3.1 Training and testing

Training convergence

The general training procedure adopted throughout this analysis is as follows. Networks were trained with the SCG algorithm in two sub-cycles. During the first sub-cycle, no regularization was performed ($\alpha = 0$). The number of training iterations during this sub-cycle was limited to 3000, but the training was sometimes halted before, when both the maximum of ΔE and the maximum of ΔW was below the arbitrary threshold of $\epsilon = 10^{-7}$. During the second sub-cycle, the weight decay parameter α was set to a constant value of 0.05 and training was limited to 1000 iterations but converged more rapidly most of the time.

Figure 5.4 shows the evolution of training error accross training iterations, averaged over 28 MLPs with two hidden neurons (subject "AT", combination stimuli). One error curve is shown for each network output (f_H, f_U and f_N). The first sub-cycle was halted after 3000 iterations. A sudden increase in error happen at the beginning of the second training sub-

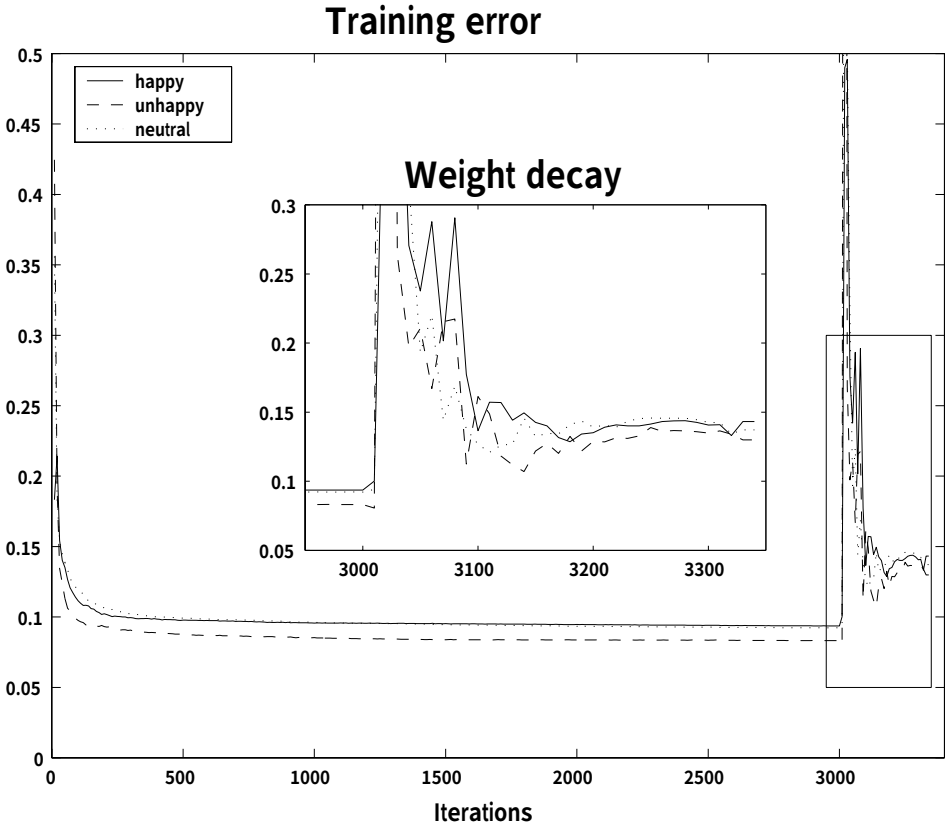


Figure 5.4: Evolution of training error, averaged over 28 networks (subject "TA", combination stimuli). Insert: evolution of training error during weight decay regularization.

cycle, due to the introduction of a non-null regularization factor α. The evolution of error during the second sub-cycle is detailed in the insert.

Effect of regularization

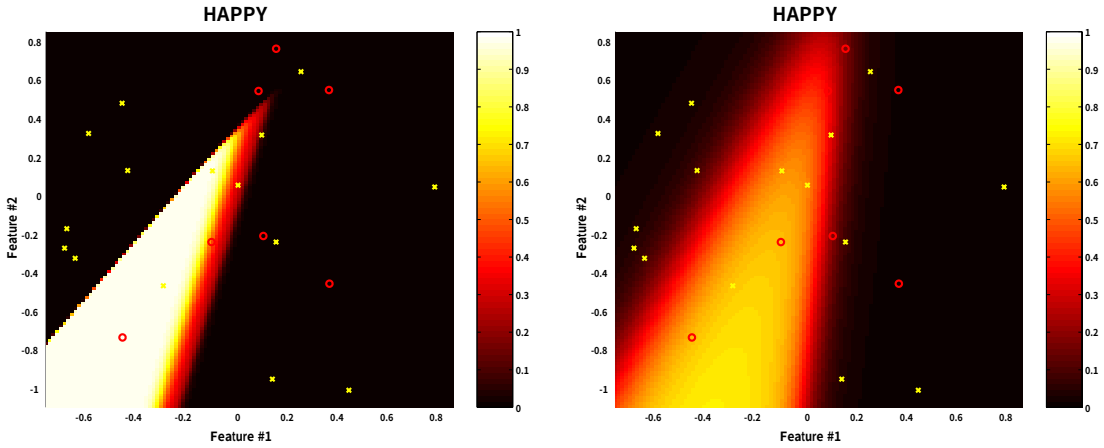
The simple weight decay scheme adopted in this study showed to be efficient for improving generalization. Table 5.3 illustrates this. 22 networks with two hidden neurons (MLP2) were trained for subject "ED" and combination stimuli. The rate of correct answers (estimated by leave-one-out cross-validation) is shown for each network output f_H, f_U and f_N, as well as the total correct classification rate (CCR). The first row was computed after the first training sub-cycle (without weight decay) and the second row at the end of the second sub-cycle (with weight decay).

Table 5.3: Rate of correct outputs and correct classification rate (CCR) with 22 MLP2.

W. decay	f_H	f_U	f_N	**CCR**
WITHOUT	59.09 %	77.27 %	72.73 %	**54.55 %**
WITH	63.64 %	86.36 %	77.27 %	**63.64 %**

The effect of weight decay is essentially equivalent to a re-scaling of the network weights,

resulting in a smoothing of the network outputs. This is illustrated in figure 5.5 which shows the first output f_H of one of the 22 trained networks against the first two features x_1 and x_2 (x_3 set to 0), (a) without weight decay, and (b) with weight decay.



(a) (b)

Figure 5.5: Effect of weight decay on the network output: (a) without weight decay, (b) with weight decay.

Influence of weight initialization

The effect of weight initialization cannot be quantized with respect to the generalization error, since the latter is estimated from several networks trained on slightly different problems. Initialization may however affect the quality of the training of a particular network.

On the one hand, some initial weight values may lead the network to be "trapped" in a local minimum because the optimization performed by the SCG algorithm is not global. On the other hand, if the network happens to be underconstrained, that is, if there are several global minima (other than resulting from weight permutations), different initial weights may lead to different solutions. The latter most probably occurred in this case since networks were sometimes underconstrained due to the small number of patterns. The influence of weight initialization was more observed on MLP with three hidden neurons than with two, because increasing the number of hidden neurons also increases the number of degrees of freedom of the network, and may result in more possible solutions for the weight optimization problem.

In section 5.3.3, I propose to use committees of networks for tackling the problem of different weight initializations leading to different trained networks.

5.3.2 Model selection

Linear regression

For comparison with neural networks, results obtained with classical linear regression technique are shown first. Figure 5.6 shows 3-dimensional feature vectors for subject "AT" with sound stimuli, projected on the first two features.

All labeled patterns are shown in (a). The three other graphs (b)–(d) respectively represent the three components f_H, f_U and f_N of the output of the linear regression classifier. In graph

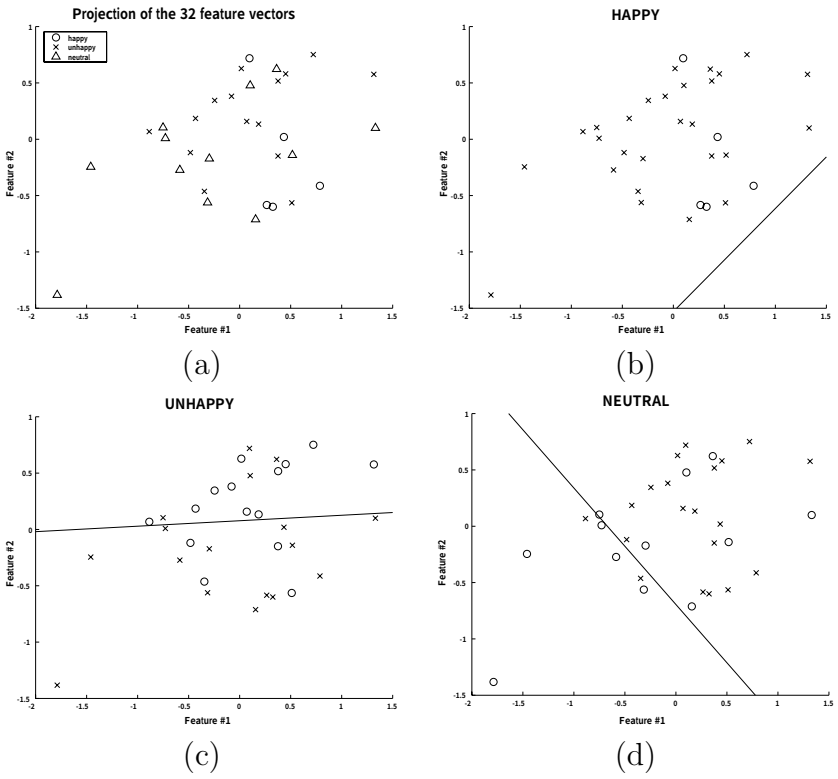


Figure 5.6: Classification with linear regression (LR): projection of labeled samples (a); separating hyperplanes for f_H (b), f_U (c) and f_N (d).

(b), patterns belonging to the happy class are indicated with a '$\circ$' and other patterns with a '$\times$' (in other graphs as well, the patterns belonging to the class of interest are indicated with a '$\circ$').

Obviously, the classification performed with linear regression is not optimal. Linear regression as applied here displays two important drawbacks:

1. Inability to model a non-linear surface for classifying inputs.

2. Unsuitability of the sum-of-squares error (SSE) for classfication.

The latter drawback is illustrated in (b), where the "separating" hyperplane does even not cross the data. In this case, the SSE is minimum when all patterns are classified as "not-happy" ($\times$).

Number of neurons in hidden layer

Networks with two and three hidden neurons have been trained on each data set D_S^t (for all subjects S and for the three types of stimulus t) and generalization error estimated by leave-one-out cross-validation.

Table 5.4 summarizes some results for linear regression (LR), compared to neural networks with 2 (MLP2) and with 3 (MLP3) hidden neurons, respectively. Rate of correct answers is shown for each output, as well as the correct classification rate (CCR).

Table 5.4: Rate of correct answers and CCR with LR, compared to MLP2 and MLP3.

— PICTURES —

subj.		f_H	f_U	f_N	CCR
	LR	70.37 %	55.56 %	62.96 %	**44.44 %**
MI	MLP2	77.78 %	59.26 %	66.67 %	**51.85 %**
	MLP3	70.37 %	55.56 %	62.96 %	**44.44 %**

— SOUNDS —

subj.		f_H	f_U	f_N	CCR
	LR	90.62 %	75.00 %	71.87 %	**68.75 %**
RY	MLP2	90.62 %	53.12 %	56.25 %	**50.00 %**
	MLP3	84.37 %	71.87 %	68.75 %	**62.50 %**
	LR	81.25 %	59.37 %	53.12 %	**46.87 %**
AT	MLP2	84.37 %	56.25 %	53.12 %	**46.87 %**
	MLP3	81.25 %	68.75 %	62.50 %	**56.25 %**
	LR	73.33 %	76.67 %	63.33 %	**56.67 %**
EG	MLP2	66.67 %	66.67 %	60.00 %	**46.67 %**
	MLP3	73.33 %	70.00 %	70.00 %	**56.67 %**

— COMBINATIONS —

subj.		f_H	f_U	f_N	CCR
	LR	79.17 %	66.67 %	45.83 %	**45.83 %**
MI	MLP2	83.33 %	54.17 %	54.17 %	**45.83 %**
	MLP3	79.17 %	66.67 %	70.83 %	**58.33 %**
	LR	59.09 %	77.27 %	63.64 %	**50.00 %**
ED	MLP2	63.64 %	86.36 %	77.27 %	**63.64 %**
	MLP3	59.09 %	72.73 %	77.27 %	**54.55 %**
	LR	50.00 %	64.29 %	85.71 %	**50.00 %**
TA	MLP2	64.29 %	85.71 %	71.43 %	**60.71 %**
	MLP3	71.43 %	78.57 %	78.57 %	**64.29 %**
	LR	61.90 %	66.67 %	66.67 %	**47.62 %**
HE	MLP2	76.19 %	71.43 %	66.67 %	**57.14 %**
	MLP3	71.43 %	66.67 %	57.14 %	**47.62 %**

Several facts can be observed in this table:

- Total correct classification rate (rightmost column) was better for some neural network (MLP2 or MLP3) than for linear regression in 6 cases out of 8.
- Three hidden neurons were sometime necessary to observe an improvement over LR. Consider for example results for subject "MK" where MLP2 did not give better results than LR, but MLP3 did (58% vs. 45% for LR).
- Better performance with two hidden neurons (MLP2) do not imply better performance with three hidden neurons (MLP3). For example, for subject "ED", the correct classification rate decreased from 63% for MLP2 to 54% for MLP3.
- An increase in the number of hidden neurons may be beneficial for some of the classifier outputs but not for others. Results for subject "TA" illustrate this. The rate of correct values increased between MLP2 and MLP3 for both output f_H and f_N but decreased for output f_U.

To summarize, these results show that neural networks (with the appropriate number of hidden neurons) often effectively improve over linear classification.

Superiority of non-linear classification. Figure 5.7 shows ROC[2] curves for the first output f_H of LR, MLP2 and MLP3 for subject "TA" and combination stimuli, computed with leave-two-out CV (378 networks of each type).

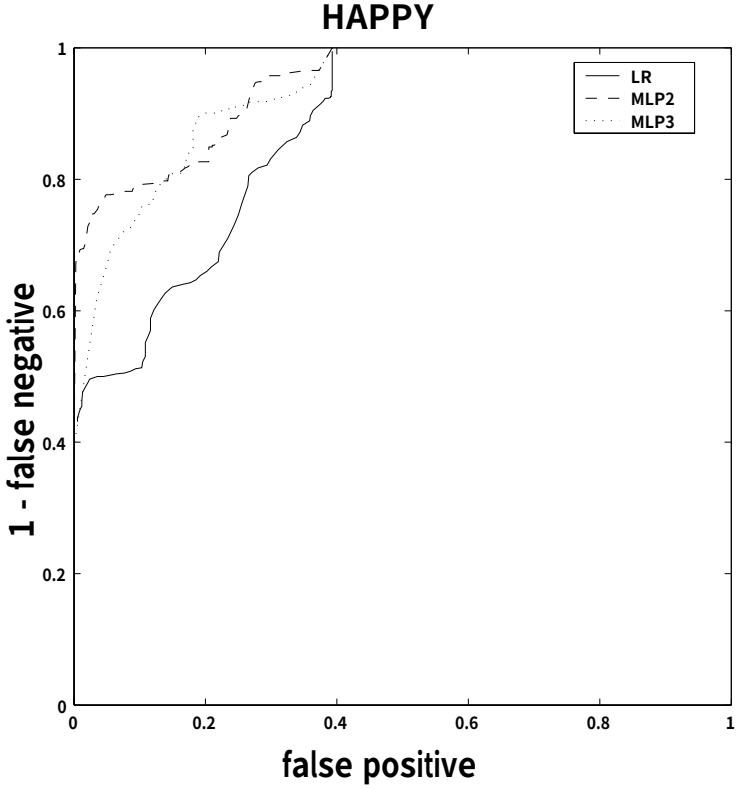


Figure 5.7: ROC curves for the first output f_H of LR, MLP2 and MLP3.

Clearly, both MLP2 and MLP3 yielded smaller misclassification rate than LR. Moreover, the superiority of CCR for both MLP2 and MLP3 over LR was highly significant ($P < 0.005$).

Increasing the number of neurons can improve classification. Loosely speaking, more neurons in the hidden layer allow to model more complex class distributions in the input space. If the *actual* distribution is indeed complex, this ability is necessary and adding hidden neurons will improve classification (provided the network is properly trained). Figure 5.8 shows the first output of an MLP3 for subject "TA" and combination stimuli.

In this case, a more complex network with 3 hidden units was better able to model the class probability densities for the three classes H, U and N than a simpler two hidden neuron network. Note that the generalization error computed with LOO-CV not only depends on the performance of the network network shown here, but also on the performance of the 27 other ones.

[2]Receiver operating characteristic

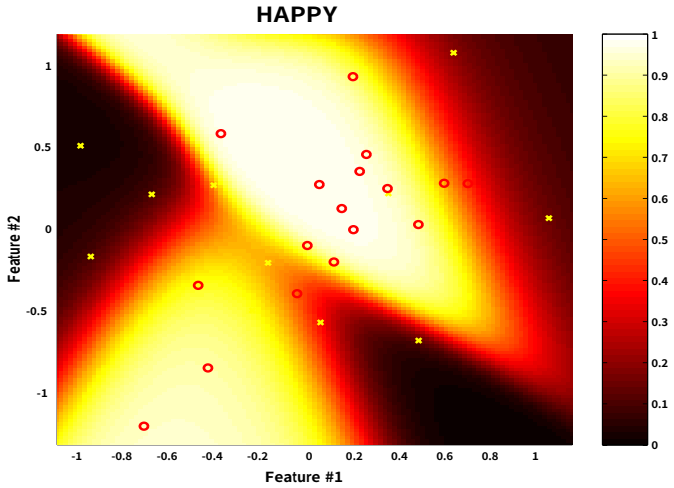


Figure 5.8: "happy" output f_H of MLP3 (subject "TA" combination stimuli).

"Too much non-linearity" is not good. An attempt is now made too explain why MLP3 sometimes showed worse performance than MLP2. A look back at the results for subject "HE" in table 5.4 shows that the correct classification rate for this subject dropped for MLP3, compared to MLP2. Figure 5.9 shows the third output f_N of a two hidden neuron network (a), compared to four networks with three hidden neurons (b)–(e), for subject "HE". The mapping realized by MLP2 (a) is rather simple and the 20 other MLP2 were similar. On the right, f_N is shown for 4 out of the 21 MLP3. It appears that only a slight difference between the training sets (one sample added and another left out) of these networks resulted in an important difference in the resulting mapping. This is because, having more degrees of freedom, the network had more possibilities of finding a minimum of the error function. The convergence to a minimum was rapid, but the mapping did probably not model the actual probability density. The observed generalization error was thus higher.

In summary, when the number of degrees of freedom is too high compared to the constraints applied by the training set, the generalization ability is likely to worsen.

Is one neural network sufficient ? The last fact observed from results of table 5.4 was that the rate of correct values of different network outputs do not vary in the same way when modifying the degree of nonlinearity (number of hidden neurons) of the network. While increasing the number of hidden neurons may improve performance for a given output, it may introduce too many degrees of freedom for another output, resulting in lower performance. Some outputs may have a more complex probability distribution than others, and thus require more neurons than other classes. Of course, the probability densities of the three classes are strongly related, since they are theoretically complementary. In practice, however, the problem of modeling different classes may be seen as separated problems, which are sometimes conflicting since the classes are overlapping.

Based on these considerations, I investigated the possibility to train three separated networks, *one network for each class*. These networks were combined in a committee, as depicted in figure 5.2. Results are presented in the next section.

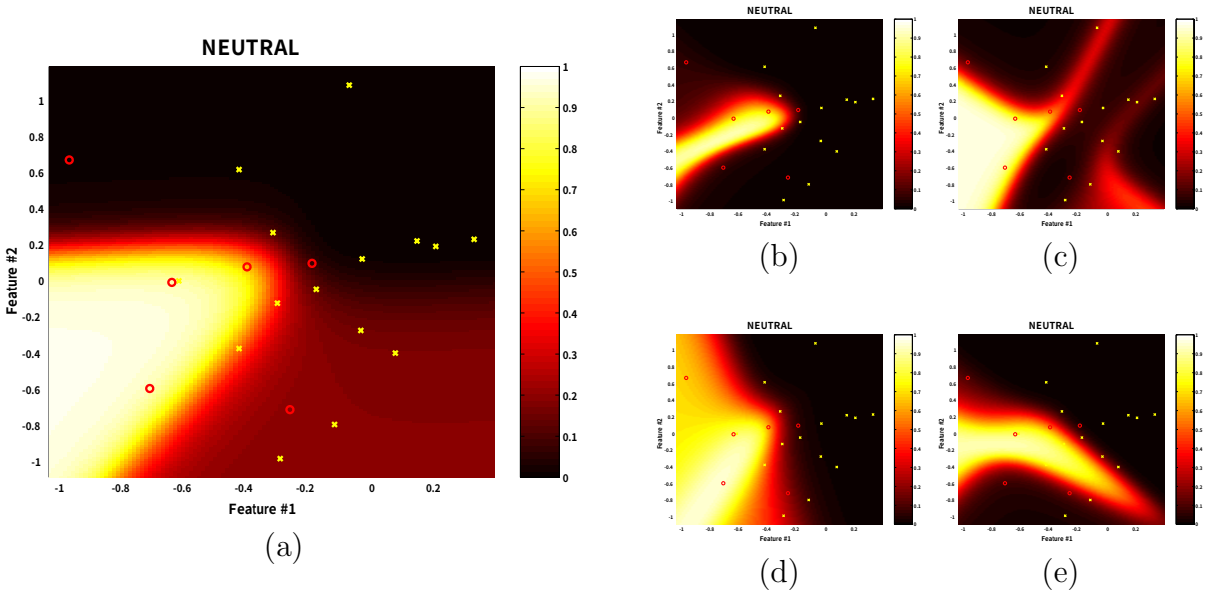


Figure 5.9: "neutral" output f_N of (a) MLP2 and (b)–(e) MLP3 (subject "HE" combination stimuli).

5.3.3 Committees of networks

Two different types of committees of networks have been investigated: (1) committees consisting in one network associated to each class, the outputs of all networks being combined as described in section 5.2.1 to yield an estimate of the membership probability of each class for a given feature vector, and (2) committees consisting in several networks with the same architecture but different initializations and/or with different architectures. The latter are called *average committees*.

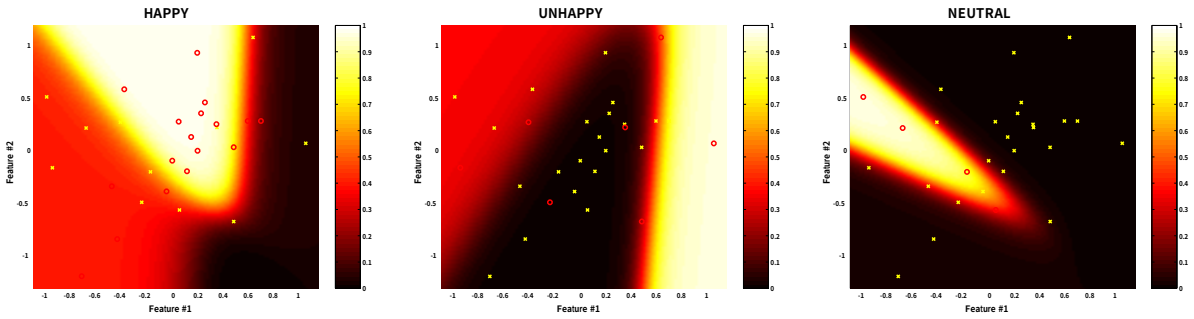
Committees with one network for each class

One advantage of using separate networks for modeling different class membership probabilities is that all the classes are not constrained to be modeled using the same separating hyperplanes. The mapping of each network is simplified, because all each network has to learn is a binary classification, and the number of degrees of freedom to achieve this is increased. However, this can turn into a drawback as increasing too much the number of degrees of freedom eventually yields worse generalization ability.

Committees of networks have shown to be interesting in some cases but not better than a single network in others. Figure 5.10 shows the output of three complementary networks, each with two hidden neurons, trained for subject "TA" and combination stimuli.

These graphs suggest that the three class membership probabilities are more finely estimated by three dedicated networks than by a single network.

This is confirmed by an increase of the correct answer rate for each output and CCR summarized in table 5.5, in comparison with previous results for MLP2 with the same training sets (generalization error was estimated by LOO-CV on 22 committees).



(a) (b) (c)

Figure 5.10: Committee of three MLP2: (a) f_H, (b) f_U, and (c) f_N.

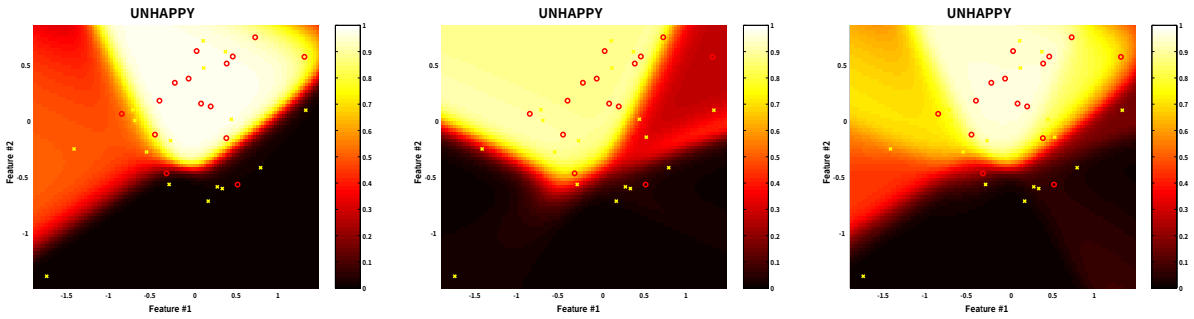
Table 5.5: Rate of correct outputs and correct classification rate (CCR) for MLP2 compared to a committee of 3 MLP2.

	f_H	f_U	f_N	**CCR**
MLP2	64.29 %	85.71 %	71.43 %	**60.71 %**
Committee	75.00 %	75.00 %	78.57 %	**64.29 %**

Average committees

Sensitivity of the performance of trained networks to weight initialization was briefly mentioned in the previous section. Networks that are underconstrained are characterized by an error surface with several global minima. Some global minima may lead to good generalization performance, while others not. However, no criterion exists to choose between these minima and, in practice, the adopted global minimum will depend on a particular weight initialization.

To improve generalization, I propose, instead of arbitrary choosing a single network, to estimate the output of a new feature vector as the *averaged output of several networks with different initializations*. Figure 5.11 illustrates this by an example. (a) and (b) show the second output f_U of two MLP3 trained with different weight initializations (subject "AT", sound stimuli). (c) shows f_U averaged over 10 such networks.



(a) (b) (c)

Figure 5.11: Average committee of ten MLP3: (a)–(b) f_U of two MLP3 with different weight initialization, and (c) average f_U over 10 MLP3.

Committees of networks showed better generalization over single networks in some cases, but not in others. The concept of averaged committee presented here could be improved by weighting the average with an estimate of the quality of each network. Bayesian *evidence* [46] provides such an estimate, defined as the product of the likelihood (goodness of fit) by the Ockham's factor (simplicity of the network) [75]. However, as previously mentioned, this estimate becomes reliable only when sufficient data are available.

Nothwithstanding possible improvements, I see essentially two advantages of averaged committees:

1. Committees of networks have more degrees of freedom than single networks (allowing to better model the actual class probabilities), yet not loosing potential generalization ability thanks to averaging (resulting in smoothed output).

2. Committees of networks prevent poor generalization due to unfortunate weight initialization.

5.4 Conclusion

This chapter showed how neural networks could efficiently be used in an emotion expressing system for mapping EEG features onto emotion.

The main difficulties encountered were (1) the small number of data available for each subject, and (2) the choice of features that were good in general but probably not optimal for each subject.

Results of this analysis could be summarized as follows:

- Training with extremely few patterns, yielding 64% of correct classification for new, unseen patterns in three emotional classes.

- Selection of simple but efficient methods for training: rapid algorithm, simple regularization scheme, involving almost no manual tuning. Estimation of generalization ability.

- Demonstration of the general superiority of neural networks over linear classifier, because of their ability to learn non-linear mappings and generalize from few examples.

- Discussion of model complexity with respect to generalization. When few examples are available, it is important to keep the model simple to ensure good generalization.

- Investigation of the use of committees to improve generalization.

I believe that the current results could be substantially improved by using different EEG features, adapted to each subject. Further research could include the investigation of other neural networks than MLP.

Chapter 6

Adaptive Emotion Expressing System

Based on neurophysiological substrates of emotion presented in chapter 2 and using digital signal processing techniques developed in chapter 3, EEG features correlated with emotion were selected in chapter 4. Using these features, chapter 5 discusses the training of neural networks to non-linearly classify EEG epochs into emotional classes.

This chapter explains how these findings can be integrated into an online adaptive system allowing disabled individuals to express their emotion to the outer world. Several issues are discussed and the future development of a hybrid BCI is proposed which would integrate emotion expression and verbal communication.

6.1 Online emotion expression

An online system based on techniques presented in the preceding chapters can be designed for the patient suffering from ALS to express emotion to his/her surrounding. This system, depicted in figure 6.1, continuously record the EEG of the patient and estimates spectral features every 10 seconds. EEG features are fed to a neural network trained for that particular person and almost instantaneously provides an estimate of the emotional state of the patient in terms of valence and arousal.

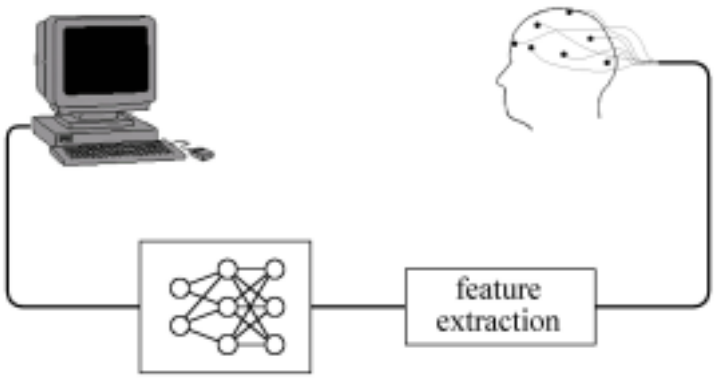


Figure 6.1: Online system for emotion expression.

As previously, the emotional state is assumed to be represented by one of three valence conditions: happy, neutral or sad, and by one of two arousal conditions: excited or calm. In principle, assuming it has been trained correctly, the neural network is able to estimate

the class membership probabilities for a new feature vector given as output. However, in practice, some epochs may be rejected because of noise in the EEG. Besides, even when the network actually yields an output, it may be hazardous to give an answer in some cases, if the probabilities are not clear enough or contradictory (for instance, it would be dangerous to say the subject is happy if the class membership probabilities are 0.72 for the happy class, but also 0.69 for the unhappy class).

The system should therefore be allowed for an *"unknown"* answer, when the signal is noisy or when the probabilities are not clear.

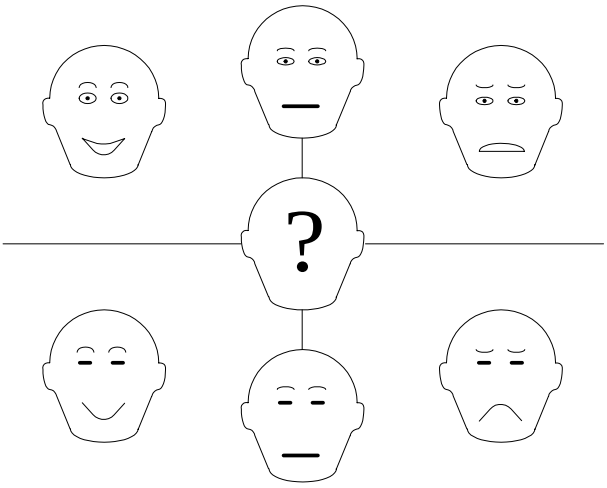


Figure 6.2: Graphical emotion expression.

In practice, emotion can be expressed in various ways. For instance, a graphical representation can be implemented by modifying facial expression of a manikin, according to the estimated emotional state. An example of such representation, derived from the SAM system [43] is shown in figure 6.2. Alternative representations include vocal output, by modifying the prosody of speech[1], and of course a combination of both. Ideally, the system's emotion expression should be as close as possible to human expression.

6.2 Learning with feedback from the user

The system described in section 6.1 lacks an essential element: the ability to continuously *adapt* to the user. An erroneous estimate is always possible and should be corrected so that the system does not commit twice the same error. On the other hand, the brainwaves of the patient may evolve with the malady, requiring the system to adapt to these changes.

For the system to be able to adapt, it must be provided with a *feedback from the user*. Different types of feedback can be envisioned for the user to correct the system. The simplest feedback would be a YES-NO answer, which gives no more information to the system than the notice of success or failure. A more elaborate type of feedback could specify which of valence and arousal is correct or erroneous. A complete feedback signal would involve a detailed description of the experienced emotion. Hence, the feedback provided by the user ranges from purely evaluative feedback to purely instructive feedback [73].

[1]This would be particularly interesting in combination with an EEG communication system with vocal output.

When the emotion actually experienced by the patient is known, the system can be trained as described in chapter 5 using supervised learning. However, as the feedback becomes evaluative rather than instructive the problem turns to a reinforcement learning problem, thus involving different techniques than for classical neural network training. In practice, we may think of adding an *instructor* that collects evaluative feedback signals from the user and uses this information to teach the network how to adapt its estimation. An emotion expressing system that learns with feedback from the user is depicted in figure 6.3.

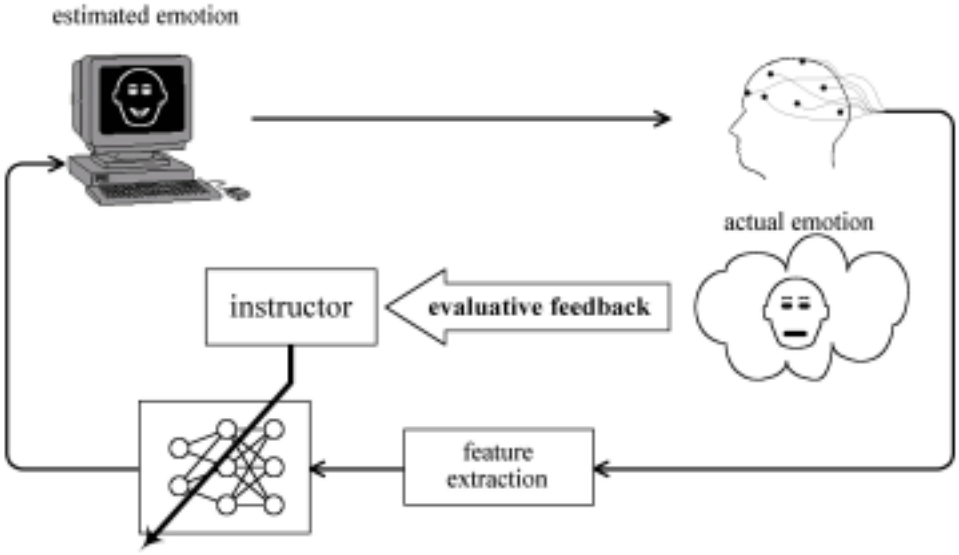


Figure 6.3: Emotion expressing system with feedback learning.

How can the patient give any feedback to the system ? Let us assume the system to be brought to the patient at early stage of the disease, that is, *before* communication becomes completely impaired. In a first phase, when the patient still can speak or use a pencil to communicate his/her emotion to the people around, the system can be trained in a supervised way by providing a rather detailed feedback. In a second phase, when communication becomes more difficult, the feedback will become more succint and evaluative. The third phase is when the patient reaches a complete "locked-in" state and no other way exist to communicate than the EEG. At that stage, feedback from the user is particularly important to verify emotion estimated by the system, but cannot be collected as before. In the next section, a hybrid system is proposed which uses controlled and uncontrolled EEG for solving this problem.

6.3 A hybrid BCI using neuro-feedback

As mentioned in the introduction, much work has been done in the field of brain-computer interfaces (BCI) for allowing communication by means of controlled EEG. The proposed hybrid BCI combines techniques developed in the previous chapters for extracting emotional information from the *uncontrolled* EEG with techniques for communicating with *controlled* EEG.

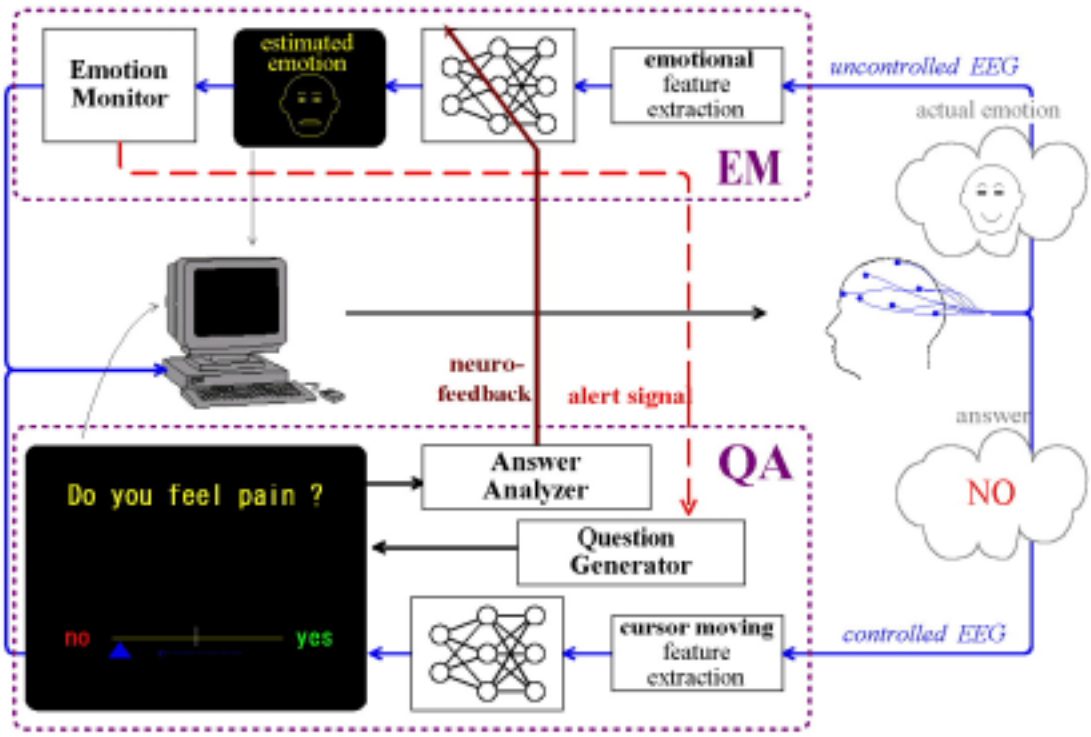


Figure 6.4: Hybrid BCI using neuro-feedback.

The hybrid BCI (shown in figure 6.4) consists of two units:

1. An *emotion monitoring unit* (EM), which is basically the system depicted in figure 6.3. This element continuously estimates the emotional state of the patient from uncontrolled EEG. Emotion is monitored and compared to a "baseline state" in order to detect extreme states (e.g. pronounced unhappiness).

2. A *question-answer unit* (QA), which is triggered when an extreme emotional state is detected. This element asks simple questions to the user in order to confirm and diagnose the emotion estimated by the EM. The patient answers the questions by controlling his/her EEG as in classical BCIs.

In this system, user feedback is collected trought the EEG and is therefore called *neuro-feedback*[2].

In practice, neuro-feedback can be implemented as follows. The QA unit alerted by the EM of an extreme emotional state asks a simple (binary) question to the user (e.g. "Do you feel pain ?"). A cursor is displayed on the screen at half-way between two targets (say, YES and NO). The user then moves the cursor towards one of the targets to answer the question, by controlling his/her EEG with a particular mental task. Examples of such mental task are imagination of movement or control of mu-rhythm. It is also possible to increase the robustness of the QA unit by providing bilateral auditory stimuli and asking the subject to concentrate on either the left or the right stimulus (this method called *attention switching* requires stimuli but makes the discrimination more robust).

[2]In this context, the term *neuro-feedback* does not relate to EEG biofeedback (therapy based on voluntary EEG regulation).

When the cursor reaches one of the two answers, the QA unit analyzes the answer and relays neuro-feedback to the EM unit so that it can adapt emotion recognition if necessary. In addition, other questions can be asked to the user in that way in order to refine the diagnostic and collect necessary information to improve the situtation of the patient.

Practical problems in the development of the hybrid BCI described above include:

- Dispose of a sufficiently robust BCI for the QA unit (with a very low rate of false negative answers).
- Devise a suitable reinforcement learning scheme for allowing the system adapting to the user with evaluative feedback only.
- Having a system capable of adapting very fast to a new user (requiring little training).

Nothwithstanding these open issues, the hybrid BCI combining emotional expression combined with verbal communication may become realizable in the near future.

Chapter 7

Conclusion

Summary

This research focused on the development of a communication interface based on electroencephalogram (EEG) analysis, for allowing patient suffering from a motor-neuron disease (such as amyotrophic lateral sclerosis, ALS) to express their emotions to their surroundings (family, doctors).

Emotion was studied in chapter 2, from both neurophysiological and neuropsychological points of view. Substrates of emotional experience in the brain were analyzed at three levels: limbic system (responsible for rough emotional feeling), paralimbic cortices (where emotion is refined and associated with high-level sensory information) and neocortex (highest level of emotional experience; seat of motivation and personality). These facts were summarized into a neurophysiological model of emotion. Psychological views of emotion were also considered in order to better describe and classify emotion. A simple representation was adopted, in which emotion is expressed in two dimensions: valence and arousal.

Two experiments involving EEG recording and analysis were carried out. Chapter 3 presented EEG recording and signal processing techniques used in these experiments. Suited techniques were developed for minimizing artifacts (due to eye movements and to muscle activity). Spectral analysis techniques based on the fast Fourier transform (FFT) were applied to estimate spectral features on short EEG segments: power spectral density, inter-electrode coherence, auto- and cross-bicoherence.

Results of the two experiments were reported and discussed in chapter 4. In a preliminary experiment, the effect of emotional stress was assessed by analyzing the EEG of nine subjects when playing video games. Single features were shown to reflect individual characteristics, while integrated features allowed to highlighting inter-subject agreement. In a second experiment, the effect of emotional valence (happy vs. unhappy) and arousal (excited vs. calm) on the EEG was observed in twenty subjects. According to a computer-driven procedure, short EEG epochs were recorded while subjects were presented visual, auditory and audio-visual emotional stimuli. Emotional ratings were collected after each epoch. EEG features significantly correlated with particular emotional experience were isolated by comparing among subjects EEG epochs associated with extreme situations: emotional vs. neutral, happy vs. unhappy and excited vs. calm.

In chapter 5, artificial neural networks were used to learn to nonlinearly classify EEG epochs of each subject into several emotional classes. Previously selected spectral features

were fed as input of multi-layer perceptron (MLP) classifiers. Issues of training and testing networks with few examples, model selection, and generalization were addressed in particular. Training was carried out with the scaled-conjugate gradient (SCG) algorithm and with a simple weight decay scheme. MLP with two or three hidden units were shown to outperform linear classifiers. The use of committees of networks was proposed to improve generalization.

The use of previously developed techniques in an online adaptive emotion expression system was discussed in chapter 6. The development of a hybrid brain-computer interface was proposod, combining emotion expression with verbal communication.

Future issues

Possible extensions of this research include:

- Investigation of other EEG features. In particular, further research could be done about the description of emotion with higher-order spectra and with parametric (AR, ARMA) features.
- Replication of experiments with standard emotional stimuli in order to increase the amount of data. The availability of more examples for the same subject would indeed allow to increase the number of features, and thereby, the overall performance of the system.
- Study of other neural network architectures for comparison.

Bibliography

[1] L. I. Aftanas, V. I. Koshkarov, V. L. Pokrovskaja, and Y. N. Mordvintsev. Dimensional analysis of human EEG and anxiety coping styles. *Journal of Psychophysiology*, 10:49–60, 1996.

[2] L. I. Aftanas, N. V. Lotova, V. I. Koshkarov, V. L. Pokrovskaja, S. A. Popov, and V. P. Makhnev. Non-linear analysis of emotion EEG : calculation of Kolmogorov entropy and the principal lyapunov exponent. *Neuroscience Letters*, 226:13–16, 1997.

[3] C. W. Barlow, E. R. Soicher, J. B. Barlow, B. M. Friedman, and D. P. Myburgh. Post-exercise time-course analysis of ST segment and T wave changes: an important contribution to the role of stress electrocardiography in aircrew. *Aviation Space & Environmental Medicine*, 62(2):165–71, Sept. 1991.

[4] N. Birbaumer, N. Ghanayim, T. Hinterberger, I. Iversen, B. Kotchoubey, A. Kübler, J. Perelmouter, E. Taub, and H. Flor. A spelling device for the paralysed. *Nature*, 398: 297–298, Mar. 1999.

[5] C. M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.

[6] J. C. Borod. Interhemispheric and intrahemispheric control of emotion: a focus on unilateral brain damage. *Journal of Consulting and Clinical Psychology*, 60(3):339–348, 1992.

[7] J. C. Borod, B. A. Cicero, L. K. Obler, J. Welkowitz, H. M. Erhan, C. Santschi, I. S. Grunwald, R. M. Agosti, and J. R. Whalen. Right hemisphere emotional perception: evidence across multiple channels. *Neuropsychology*, 12(3):446–458, 1998.

[8] M. M. Bradley and P. J. Lang. International affective digitized sounds (IADS): stimuli, instruction manual and affective ratings. Technical Report B-2, The Center for Research in Psychophysiology, University of Florida, 1999.

[9] T. Canli, J. E. Desmond, Z. Zhao, G. Glover, and J. D. E. Gabrieli. Hemispheric asymmetry for emotional stimuli detected with fMRI. *NeuroReport*, 9(14):3233–3239, Oct. 1998.

[10] R. Cooper, A. L. Winter, H. J. Crow, and W. Grey Walter. Comparison of subcortical, cortical and scalp activity using chronically indwelling electrodes in man. *Electroencephalography and Clinical Neurophysiology*, 18:217–228, 1965.

[11] H. J. Crawford, S. W. Clarke, and M. Kitner-Triolo. Self-generated happy and sad emotions in low and highly hypnotizable persons during waking and hypnosis : laterality and regional EEG activity differences. *International Journal of Psychophysiology*, 24: 239–266, 1996.

[12] F. L. da Silva. Neural mechanisms underlying brain waves: from neural membranes to networks [review]. *Electroencephalography and Clinical Neurophysiology*, 79:81–93, 1991.

[13] R. J. Davidson. Anterior cerebral asymmetry and the nature of emotion. *Brain and Cognition*, 20:125–151, 1992.

[14] R. J. Davidson. Emotion and affective style: hemispheric substrates. *Psychological Science*, 3(1):39–43, 1992.

[15] R. J. Davidson and W. Irwin. The functional neuroanatomy of emotion and affective style. *Trends in Cognitive Sciences*, 3(1):11–21, Jan. 1999.

[16] G. Dawson. Frontal electroencephalographic correlates of individual differences in emotion expression in infants: a brain systems perspective on emotion. *Monographs of the Society for Research in Child Development*, 59(2–3):135–51, 1994.

[17] V. De Pascalis, W. J. Ray, I. Tranquillo, and D. D'Amico. EEG activity and heart rate during recall of emotional events in hypnosis: relationships with hypnotizability and suggestibility. *International Journal of Psychophysiology*, 29:255–275, 1998.

[18] J. del R. Millán, J. Mouriño, M. G. Marciani, F. Babiloni, F. Topani, I. Canale, J. Heikkonen, and K. Kaski. Adaptive brain interfaces for physically-disabled people. In *20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Oct.-Nov. 1998.

[19] D. Denny-Brown and R. A. Chambers. The parietal lobe and behavior. *Association for Research in Nervous and Mental Disease*, 36:35–117, 1958.

[20] D. Derryberry and D. M. Tucker. Neural mechanisms of emotion. *Journal of Consulting and Clinical Psychology*, 60(3):329–338, 1992.

[21] P. Ekman. Are there basic emotions ? *Psychological Review*, 99(3):550–553, 1992.

[22] S. Francis, E. T. Rolls, R. Bowtell, F. McGlone, J. O'Doherty, A. Browning, S. Clare, and E. Smith. The representation of pleasant touch in the brain and its relationship with taste and olfactory areas. *NeuroReport*, 10:453–459, 1999.

[23] T. Gasser, P. Bächer, and J. Möcks. Transformations towards the normal distribution of broad band spectral parameters of the EEG. *Electroencephalography and Clinical Neurophysiology*, 53:119–124, 1982.

[24] K. M. Heilman. The neurobiology of emotional experience. *The Journal of Neuropsychiatry and Clinical Neurosciences*, 9(3):439–448, Summer 1997.

[25] K. M. Heilman and R. L. Gilmore. Cortical influences in emotion. *Journal of Clinical Neurophysiology*, 15(5):409–423, 1998.

[26] H. Hinrichs and W. Machleidt. Basic emotions reflected in EEG-coherences. *International Journal of Psychophysiology*, 13:225–232, 1992.

[27] K. Hirota. [The concept of vagueness and entropy analysis for questionnaire inquiries]. *Operations Research*, pages 38–44, Jan. 1981. (In Japanese).

[28] J. Hurri, H. Gävert, J. Särelä, and A. Hyvärinen. FastICA MATLAB^{TM} package [online]. URL `http://www.cis.hut.fi/projects/ica/fastica`.

[29] A. Hyvärinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9:1483–1492, 1997.

[30] I. E. Ifeachor, B. W. Jervis, E. L. Morris, E. M. Allen, and N. R. Hudson. New online method for removing ocular artefacts from EEG signals. *Medical & Biological Engineering & Computing*, 24:356–364, July 1986.

[31] C. E. Izard. Basic emotions, relations among emotions, and emotion-cognition relations. *Psychological Review*, 99(3):561–565, 1992.

[32] B. W. Jervis, E. C. Ifeachor, and E. M. Allen. The removal of ocular artefacts from the electroencephalogram : a review. *Medical & Biological Engineering & Computing*, 26: 2–12, 1988.

[33] T. Johnstone. Emotional speech elicited using computer games. In *International Conference on Spoken Language Processing, ICSLP Proceedings*, volume 3, pages 1985–1988. IEEE, 1996.

[34] R. Joseph. The limbic system: emotion, laterality, and unconscious mind. *Psychoanalytic Review*, 79(3):405–456, Fall 1992.

[35] E. R. Kandel, J. H. Schwartz, and T. M. Jessell, editors. *Principles of Neural Science*. Appleton & Lange, third edition, 1991.

[36] Z. A. Keirn and J. I. Aunon. A new mode of communication between man and his surroundings. *IEEE Transactions on Biomedical Engineering*, 37(12):1209–1214, Dec. 1990.

[37] W. Keynes. Medical response to mental stress. [review]. *Journal of the Royal Society of Medicine*, 87(9):536–9, Sept. 1994.

[38] E.-S. Kim, D.-Y. Cho, Y.-J. Lee, and C.-S. Ryu. An estimation of the bispectrum for the EEG in emotional states. In *The Fifth International Conference on Neural Information Processing, ICONIP'98*, pages 438–441, Kitakyushu, Japan, 1998.

[39] A. Kostov and M. Polak. Brain-computer interface: development of experimental setup. In *RESNA 1997*, pages 54–56, Pittsburgh, 1997.

[40] M. B. Kostyunina and M. A. Kulikov. Frequency characteristics of EEG spectra in the emotions. *Neuroscience and Behavioral Physiology*, 26(4):340–343, 1996.

[41] A. Krogh and J. A. Hertz. A simple weight decay can improve generalization. In J. E. Moody, S. J. Hanson, and R. P. Lippmann, editors, *Advances in Neural Information Processing Systems 4*, pages 950–957. Morgan Kauffmann Publishers, San Mateo, CA, 1995. (available at: `http://genome.cbs.dtu.dk/krogh/papers/weightdecay.ps.gz`).

[42] A. Kübler, B. Kotchoubey, T. Hinterberger, N. Ghanayim, J. Perelmouter, M. Schauer, C. Fritsch, E. Taub, and N. Birbaumer. The thought translation device: a neurophysiological approach to communication in total motor paralysis. *Experimental Brain Research*, 124:223–232, 1999.

[43] P. J. Lang. The emotion probe. Studies of motivation and attention. *American Psychologist*, 50(5):372–385, May 1995.

[44] P. J. Lang, M. M. Bradley, and B. N. Cuthbert. International affective picture system (IAPS): stimuli, instruction manual and affective ratings. Technical Report A-4, The Center for Research in Psychophysiology, University of Florida, 1999.

[45] P. J. Lang, M. M. Bradley, J. R. Fitzsimmons, B. N. Cuthbert, J. D. Scott, B. Moulder, and V. Nangia. Emotion arousal and activation of the visual cortex: an fMRI analysis. *Psychophysiology*, 35:199–210, 1998.

[46] D. J. C. MacKay. The evidence framework applied to classification networks. *Neural Computation*, 4(5):698–714, 1992.

[47] P. D. MacLean. Psychosomatic disease and the "visceral brain". Recent developments bearing on the Papez theory of emotion. *Psychosomatic Medicine*, XI(6):338–353, Nov.-Dec. 1949.

[48] R. J. Maddock. The retrosplenial cortex and emotion: new insights from functional neuroimaging of the human brain. *Trends in Neuroscience*, 22:310–316, 1999.

[49] I. Maremmani, E. Bonanni, F. Pieraccini, G. C. Santerini, L. Murri, and P. Castrogiovanni. Emotivity, personality, and task-dependent EEG asymmetry. *Physiology and Behavior*, 51:1111–1115, 1992.

[50] D. J. McFarland, A. T. Lefkowicz, and J. R. Wolpaw. Design and operation of an EEG-based brain-computer interface with digital signal processing technology. *Behavior Research Methods, Instruments and Computers*, 29(3):337–345, 1997.

[51] J. M. Mendel. Tutorial on higher-order statistics (spectra) in signal processing and system theory : theoretical results and some applications. *Proceedings of the IEEE*, 79(3):278–305, Mar. 1991.

[52] M. F. Moller. A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks*, 6:525–533, 1993.

[53] J. Murata, K. Matsukawa, J. Shimizu, M. Matsumoto, T. Wada, and I. Ninomiya. Effects of mental stress on cardiac and motor rhythms. *Journal of the Autonomic Nervous System*, 75(1):32–7, Jan. 1999.

[54] Muscular Dystrophy Association. Facts about amyotrophic lateral sclerosis (ALS). Available online at: `http://www.mdausa.org/publications/fa-als.html`, .

[55] Muscular Dystrophy Association. When a loved one has ALS: a caregiver's guide. Available online at: `http://www.mdausa.org/publications/alscare/index.html`, .

[56] T. Musha, Y. Terasaki, H. A. Haque, and G. A. Ivanitsky. Feature extraction from EEGs associated with emotions. *Artificial Life Robotics*, 1:15–19, 1997.

[57] E. Nierdmeyer. The normal EEG of the waking adult. In E. Niedermeyer and F. L. da Silva, editors, *Electroencephalography: basic principles, clinical applications and related fields*. Williams and Wilkins, 1993.

[58] P. L. Nunez, R. Srinivasan, A. F. Westdorp, R. S. Wijesinghe, D. M. Tucker, R. B. Silberstein, and P. J. Cadusch. EEG coherency I: statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales. *Electroencephalography and Clinical Neurophysiology*, 103:499–515, 1997.

[59] C. E. Osgood, G. J. Suci, and P. H. Tannenbaum. *The measurement of meaning*. University of Illinois Press, 1957.

[60] L. P. Panych, J. A. Wada, and M. P. Beddoes. Practical digital filters for reducing EMG artefact in EEG seizure recordings. *Electroencephalography and Clinical Neurophysiology*, 72:268–276, 1989.

[61] J. W. Papez. A proposed mechanism of emotion. *Archives of Neurology and Psychiatry*, 38:725–743, 1937.

[62] W. D. Penny and S. J. Roberts. Bayesian neural networks for classification: how useful is the evidence framework ? *Neural Networks*, 12:877–892, 1999.

[63] B. O. Peters, G. Pfurtscheller, and H. Flyvbjerg. Mining multi-channel EEG for its information content: an ANN-based method for brain-computer interface. *Neural Networks*, 11:1429–1433, 1998.

[64] H. Pihan, E. Altenmüller, and H. Ackermann. The cortical processing of perceived emotions : a DC-potential study on affective speech prosody. *Neuroreport*, 8(3):623–627, 1997.

[65] M. B. Priestley. *Spectral Analysis and Time Series*. Academic Press, 1981.

[66] R. D. Reed and R. J. Marks, II. *Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks*. The MIT Press, Cambridge, 1999.

[67] S. J. Roberts and W. D. Penny. Real-time brain-computer interfacing: a preliminary study using Bayesian learning. *Medical & Biological & and Computing*, 38(1):56–61, 1999.

[68] W. S. Sarle. Neural Networks FAQ. `ftp://ftp.sas.com/pub/neural/FAQ.html`, 1997.

[69] D. Schellberg, C. Besthorn, T. Klos, and T. Gasser. EEG power and coherence while male adults watch emotional video films. *International Journal of Psychophysiology*, 9: 279–291, 1990.

[70] G. Stenberg. Personality and the EEG: arousal and emotional arousability. *Personal and Individual Differences*, 13:1097–1113, 1992.

[71] D. T. Stuss, C. A. Gow, and C. R. Hetherington. "No longer Gage": frontal lobe dysfunction and emotional changes. *Journal of Consulting and Clinical Psychology*, 60(3): 349–359, 1992.

[72] K. Suenaga. *[EEG Artifact Atlas]*. NEC San-Ei, Aoyama Hospital. (in Japanese).

[73] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA, 1998.

[74] S. K. Sutton, R. J. Davidson, B. Donzella, W. Irwin, and D. A. Dottl. Manipulating affective state using extended picture presentations. *Psychophysiology*, 34:217–226, 1997.

[75] H. H. Thodberg. A review of Bayesian neural networks with application to near infrared spectroscopy. *IEEE Transactions on Neural Networks*, 7(1):56–72, Jan. 1996.

[76] T. J. Turner and A. Ortony. Basic emotions : can conflicting criteria converge ? *Psychological Review*, 99(3):566–571, 1992.

[77] M. van de Velde, G. van Erp, and P. J. M. Cluitmans. Detection of muscle artefact in the normal human awake EEG. *Electroencephalography and Clinical Neurophysiology*, 107: 149–158, 1998.

[78] R. Vigário, J. Särelä, V. Jousmäki, M. Hämäläinen, and E. Oja. Independent component approach to the analysis of EEG and MEG recordings. *IEEE Transactions on Biomedical Engineering*, 47(5):589–593, May 2000.

[79] R. N. Vigário. Extraction of ocular artefacts from EEG using independent component analysis. *Electroencephalography and Clinical Neurophysiology*, 103:395–404, 1997.

[80] P. D. Welch. The use of the fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions of Audio and Electroacoustics*, AU-15:70–73, June 1967.

[81] R. E. Wheeler, R. J. Davidson, and A. J. Tomarken. Frontal brain asymmetry and emotional reactivity: a biological substrate of affective style. *Psychophysiology*, 30:82–89, 1993.

Appendix

Table A: The three sets of stimuli used in the experiment.

pictures	
IAPS#	*description*
2271	woman
7195	teeth
6836	police
9040	starv. child
2518	quilting
4220	erotic wom.
1999	mickey
9250	war victim
7490	window
6350	attack
8220	runners
5621	sky-divers
9620	shipwreck
5760	nature
3053	burn victim
1812	elephants
3160	eye disease
5940	lava
2205	hospital
9921	fire
8120	athlete
1270	roach
8160	rock climber
2385	girl
8500	gold
9001	cemetery
2220	male face
2530	couple
1750	bunnies
7175	lamp
1450	gannet
2440	neutral girl

sounds	
IADS#	*description*
802	natives
215	erotic couple
116	wasp
113	cows
816	guitar
320	office
352	sports crowd
810	beethoven
721	beer
278	child abuse
353	baseball
109	carousel
704	touch tone
725	soda fizz
706	war
709	alarm clock
600	bike wreck
220	boy laugh
291	prowler
380	jack hammer
708	clock tick
815	rock 'n roll
290	fight
292	male scream
370	court sport
730	glass break
712	buzzer
254	video game
262	yawn
723	radio
410	helicopter
285	attack

combinations		
IAPS#	IADS#	*description*
9230	626	fire
2340	221	father
9911	424	car accident
7002	700	bathroom
2383	319	office
2600	702	beer
2870	820	teenager
7020	701	fan
2050	110	happy baby
2702	724	binge eating
9622	501	jet crash
8350	351	tennis
2485	401	musician
6312	277	abduction
2660	206	baby bath
1321	133	bear
2480	722	elderly man
1390	115	bees
6360	279	attack
7502	310	crowd
4660	201	erotic couple
2700	280	funeral
9210	602	rain
6313	276	attack
5750	151	nature
5450	415	lift off
8490	360	roller coaster
3230	287	dying man
8116	352	am. football
1301	106	dog

Acknowledgements

I would like to express all my gratitude to Professor Dr. Yukio Kosugi, my supervisor, who made this research possible. His continuous help, his patience and his numerous advices and discussions were invaluable. Thanks are due to Dr. Toshimitsu Musha and to the members of the Brain Functions Laboratory, Inc. who introduced me the field of EEG measurement and analysis. I thank Takao Tanabe of NFCorp, Inc. for his technical support and answer to my numerous questions. I am particularly grateful to my 25 subjects. The achievement of this thesis would never have been possible without their voluntary participation to one or both of my experiments. I wish to thank Ryoko Ikeda for the translation in Japanese of documents and instructions used in both experiments, as well as for her warm encouragements. Finally, I thank Dr. Keisuke Kameyama and Dr. Kuniaki Uto for their useful advices and discussions, and all the members of Kosugi laboratory for their disponibility and cooperation.