RAPPORT DE PROJET DE FIN D'ETUDES

GRENOBLE INP Esisar 2017/2018

Titre du projet
Predicting pictures emotion using physiological signals and Neural Networks

Nom et adresse de l'entreprise
Okayama University Graduate School of Natural Science and Technology Department of Computer Science Division of Industrial Innovation Sciences Address: 3-1-1 Tsushima-naka, Kita-ku, Okayama-shi, 700-8530 JAPAN

Nom et prénom de l'étudiant
Nguyen Hoang

Dates du stage	**01/02/2018 – 15/07/2018**
Spécialité	**ISE**
Tuteur Entreprise	**Dr. Zeynep Yücel**
Tuteur ESISAR	**Dr. David Hely**

Grenoble INP ESISAR

Document : Lettre-kit de départ 20180305

RAPPORT DE PROJET DE FIN D'ETUDES

GRENOBLE INP Esisar 2017/2018

Mots clés :
Interaction homme machine, signaux physiologiques, électroencéphalographie, activité électrodermale, réactivité

Résumé :
Récemment, le nombre d'intelligence artificielles, qui servent ou assistent les humains dans la vie quotidienne, croit rapidement. Tous ces systèmes impliquent un certain niveau d'interaction avec les humains et s'appuient sur les retours utilisateur pour planifier leurs actions futures. Par conséquent, il est important de comprendre le niveau de réactivité des utilisateurs pour adapter les services et améliorer les performances. Le but de cette recherche est d'étudier l'utilisation de signaux physiologiques, à savoir l'électroencéphalographie (EEG) et l'activité électrodermale (AED), pour observer la réactivité humaine. Nous supposons que le niveau d'engagement et l'état émotionnel d'une personne, dérivant de signaux physiologiques, sont de bonnes représentations de la réactivité. Pour tester cette affirmation, deux expériences sont réalisées pour collecter des données EEG et AED de sujets subissant des tâches mentalement / physiquement exigeantes. Bien que l'utilisation de l'EEG pour contrôler le niveau d'engagement n'ait pu être vérifiée, le modèle proposé pour prédire l'évolution de la réactivité par l'AED se révèle performant. À savoir, le modèle a une précision de 53,3%, par opposition à la précision de 43,6% et 62,6% de deux méthodes de référence, respectivement Monte-Carlo (MC) et un réseau de neurones (NN).

Keywords :
Human computer interaction, physiological signals, electroencephalography, electrodermal activity, responsiveness.

Abstract :
Recently, the number of artificially intelligent systems, which serve for and work in conjunction with humans in daily life settings, are rapidly increasing. All such systems involve a certain level of interaction with humans and rely on user's feedback to plan their future actions. Therefore, it is important to understand the level of attention/responsiveness of users to adapt the services and improve performance. The aim of this research is to investigate the use of two physiological signals, namely electroencephalography (EEG) and electrodermal activity (EDA), to observe human responsiveness. We suppose that engagement level and the emotional state of a person, derived from several physiological signals, are good representations of his responsiveness. To test this statement, two experiments are performed to gather EEG and EDA data from subjects undergoing mentally/physically demanding tasks. Although the use of EEG to monitor engagement level could not be verified, the proposed model for predicting the evolution of responsiveness from EDA is found to achieve good performance rates. Namely, the model has 53.3% accuracy as opposed to the accuracy of 43.6% and 62.6% of two reference methods, Monte-Carlo (MC) and a Neural networks (NN), respectively.

Acknowledgements

I would first like to thank my supervisor Dr. Zeynep Yücel for her help throughout my master thesis. Her advice allowed me to achieve my goals in this thesis.

I would also like to thank Dr. Bernard Chenevier, the Research Administrator of Okayama University, for connecting our universities by sending internship offers to Grenoble INP.

Finally, I would like to thank Dr. Akito Monden, the professor in charge of the laboratory of Theory of Programming and Artificial Intelligence for hosting me in his laboratory.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Recently, we encounter more and more artificially intelligent (AI) systems in daily life settings and this number is expected to increase further in parallel with the processing power of computers.

Some of these systems replace human caregivers or teachers, assisting their users in certain dedicated tasks, as in the case of e-learning tutors or robots used for autism or dementia therapy. On the other hand, some systems assist human operators such as surgical robots or driving assistant systems. The goal of all such systems is to interact with humans to provide assistance accounting for the feedback that they receive from them.

It is possible to enhance their service quality by improving the software, e.g introducing more user friendly navigation options, or by improving hardware, e.g. equipping the system with such interfaces as touch screen to provide a more intuitive form of interaction.

In addition, a more fundamental form of improvement involves understanding their surrounding context, and in particular interpretation of user feedback to infer users' cognitive states. One such state can be the level of attention/responsiveness of users, which can be used to trigger the services to invoke attention, in case the user is detected to loose focus.

Our research aims to develop the detection of human responsiveness using spontaneous and involuntary reactions of the human body, called physiological signals (PS), by investigating the attention level and emotional states of human beings. To do so, we observe

sweat secretion and brain activity using electrodermal activity (EDA) and electroencephalography (EEG), respectively.

We pose an argument that engagement level and the emotional state of a person, derived from several physiological signals, are good representations of his responsiveness. In order to test this statement, two experiments are performed to gather EEG and EDA data from subjects undergoing mentally/physically demanding tasks. Although the use of EEG to monitor engagement level could not be verified, the proposed model for predicting the evolution of responsiveness from EDA is found to achieve good performance rates. Namely, the model has 53.3% accuracy as opposed to the accuracy of 43.6% and 62.6% of two reference methods, Monte-Carlo (MC) and a Neural networks (NN), respectively.

This document presents this research, carried out in Okayama University, Japan, in the Graduate School of Natural Science and Technology from February 2018 to Mid-July 2018 under the supervision of Dr. Zeynep Yücel from Okayama University and Dr. David Hely from Grenoble INP - Esisar.

1.1 Okayama University

Okayama University is a national university founded in 1949. The main campus is located in Tsushima-Naka, Okayama in Okayama Prefecture, Japan. The university is composed of 11 faculties and 7 graduate schools, hosting a total of 14 thousand students.

Several partnerships exist between Okayama University and overseas universities in China, Southeast Asia, North and South America and Europe. In particular, the university has a partnership with Grenoble University. 685 international students are studying in Okayama University as of 2018.

My internship took place in the Graduate School of Natural Science and Technology, which is composed of several divisions working on a wide range of research topics/areas.

- Division of Mathematics and Physics
- Division of Earth, Life, and Molecular Sciences
- Division of Industrial Innovation Sciences

- Division of Applied Chemistry
- Division of Medical Bioengineering
- Division of Earth and Planetary Materials Science

I was affiliated with the Division of Industrial Innovation Sciences in the Department of Computer Science. This department focuses on education and research on the basic theory and application of information technology, artificial intelligence and computer technology. It is composed 5 laboratories with different research areas:

- Formal Language Science
- Computer Engineering
- Pattern Information Processing
- Intelligent Design
- Theory of Programming and Artificial Intelligence

I was working in the laboratory of Theory of Programming and Artificial Intelligence. s

1.2 Expected work in this internship

The main goal of our research is to explore the possibility of using PS to observe human responsiveness. In this section, we introduce the different tasks that are scheduled to achieve the research goals.

1.2.1 Scientific literature research

The first task is to perform a scientific literature search to understand the concepts and theory around human responsiveness, as well as the concepts of two certain kinds of PS, namely EEG and EDA. In particular, it is important to understand what can be observed using EEG and EDA, how they can be collected and how they are linked to reactions of the human body. Once the concepts and theory are understood, we investigate on how PS can be processed to observe responsiveness.

Outcomes: Understand the concepts of EEG and EDA. Understand human responsiveness and what can be used to observe it. Make a list of scientific papers that are used as references. Make choices for the processing methods (algorithms, programming language, etc. foe studying the PS.

1.2.2 Explore use of physiological signal sensors

Specific sensors that are used in the experimentation to collect EEG and EDA data were already bought by the laboratory. In order to use them, I am asked to learn their operation methods and associated software.

Outcomes: Knowledge on the sensors and specific software tools.

1.2.3 Experiments design and execution

Once the operation methods of the sensors are understood, experiments are performed to collect EEG and EDA using those sensors. In addition, we design the experiments by defining user tasks and choosing the stimuli.

After the experiment scenario is determined, we recruited several human subjects from members of Okayama University to participate in the required tasks.

Outcomes: Experiment scenario and protocol is determined. EDA and EEG data collected.

1.2.4 EEG and EDA processing

After the data are collected, we processed them using the methods determined as a results of the scientific literature search. In particular, the engagement level was computed using EEG. Concerning EDA, a model to predict the arousal state of a human being was proposed. The performance of this model is compared with several other reference methods.

Outcomes: Development for the processing of EEG and EDA. Experimental results and their discussion.

1.2.5 Write a scientific paper

Once the research is complete, a scientific paper on this research problem is written and submitted. It will be presented at the 36th annual conference of the Robotics Society of Japan (RSJ) 04 07-09-2018.

Outcomes: Scientific paper published.

Estimation of affect scores accounting for user responsiveness
Hoang Nguyen (Grenoble Institute of Technology, France), Serina Koyama (Okayama University), Zeynep Yucel (Okayama University), Akito Monden (Okayama University), Mariko Sasakura (Okayama University)
Scheduled for presentation at the 36th annual conference of the Robotics Society of Japan (RSJ) in the special session of "Robotics x AI" on 04-09-2018.

1.3 Working method

In the laboratory, every student works independently on a different research subject. Thus, I work on my research problem with the help of Dr. Yücel. Weekly meetings are scheduled with her to explain the project progress and discuss about the problems that I encounter. Future goals are also set in these meetings.

To make sure that I keep track of the progress, I wrote daily logs about the task achieved. A second part is composed of a table where the work planned, the work done and the work to do for the following day were written. Thanks to these logs, I am able to keep clearly in mind the tasks that have to be done.

The references gathered during the literature research are also helpful to keep in mind the existing works on the subject.

A monthly report is also written in French to Dr. Hely to explain the project progress. These reports are composed of the following sections:

- Task to achieve: this section explains the tasks that had to be done during the last month.
- Task achieved: this section explains the tasks achieved during the last month.

- Difficulties: this section explains the difficulties encountered during the realization of the tasks.
- Future tasks: this section lists the future tasks that will be done in the future.

Finally, two presentations, a mid-term and a final presentation, carried in front of two laboratories from the Computer Science department and their respective teachers. The purpose of these presentations is to present the progress of the research and receive feedback from the teachers and students.

Chapter 2

Background and related work

In this chapter, the concepts of emotions and PS, used in the research, are explained. Emotions are short-term changes in mental state arising as a reaction of human autonomic nervous system (ANS) to various stimuli. ANS regulates other (mostly unconscious) functions of the human body such as heart beat, digestion etc. Therefore, observation of such psychological responses enables evaluation of changes in the emotional or cognitive state. Here, we provide background to representation of emotions, and their effect on physiological processes.

2.1 Emotions and affect space

In spite of a lack of de-facto categorization of emotions due to the difficulty arising from their subjectivity and cultural dependence, there are basically two fundamental ways of categorizing, namely discrete and dimensional. The most prevailing discrete model belongs to Ekman, who defines six basic emotion categories as Anger, Disgust, Fear, Joy, Sadness and Surprise [6]. Ekman claims these categories are *universal* among all humans, which is highly debated. In addition to the assumption of being universal, another problem of this discrete model is due to the difficulty of comparing/evaluating disagreements. Namely, the emotions are subjective reactions, meaning that two persons could associate the same stimuli with different emotions but using the discrete mode, it is not possible to rank their agreement/disagreement.

As for dimensional approaches, the circumplex model is one of the most popular [7]. Basically, it represents emotions projecting them on the two orthogonal continuous dimensions, *valence* and *arousal* (see Fig. 2.1) [8, 9, 10].

Valence represents the pleasant or unpleasant components of a stimulus. Negative valence corresponds to an unpleasant stimulus such as the picture of a dead animal, whereas positive valence indicates to a pleasant stimulus. Arousal represents the level of interest of a person in a stimulus. Passive arousal corresponds to a stimulus, which is not exciting, e.g. picture of a wall, whereas active arousal involves interest, e.g. picture of an explosion. Arousal does not depend on the pleasant or unpleasant quality of a stimulus. Both image of a cute cat and image of an injured cat can be labeled as high arousal.

Valence-arousal space provides an efficient continuous representation of emotions, which enable the use of quantitative metrics. For this reason, this model is widely used in affective computation and thus is adopted in this study.

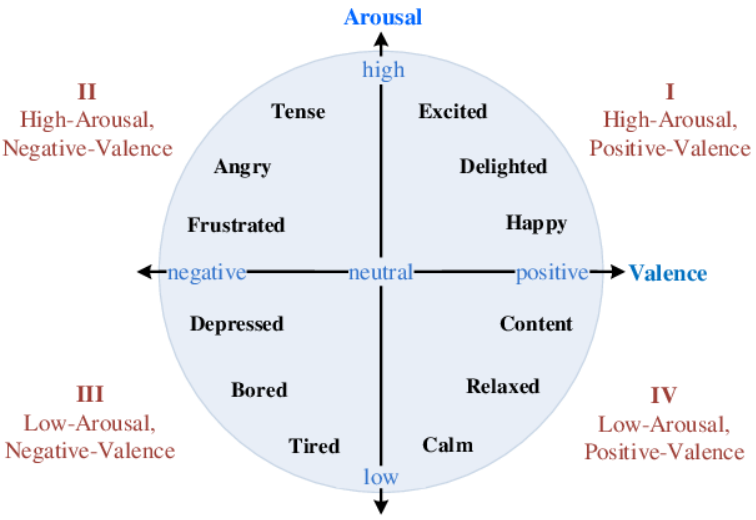

FIGURE 2.1: Two-dimensional affective experience representation. Examples such as depressed, excited, calm and Tired are placed at their approximate Valence and Arousal level [1].

2.2 Physiological signals

PS are the readings taken from bodily processes of human beings, e.g. heart-beat or respiratory rate, skin conductance, or brain electrical activity. Being controlled by ANS, they enable spontaneous observation of involuntary reactions to cognitive stimulation. In addition, they are not sensitive to cultural and social differences [11, 12], which

makes them objective markers leading to an extensive deployment in affect analysis, even though they may be affected by environmental factors (e.g. temperature or noise) or prone to motion artifacts. In this study, we use two PS, namely EEG and EDA.

2.2.1 Electroencephalography

EEG is the study of the electrical activity of the brain measured using electrodes placed on the scalp. EEG is often processed in fequency domain to interpret cognitive processes. In what follows, we give an overview on EEG sensing systems and processing of EEG waveforms.

2.2.1.1 Modelisation of EEG

In EEG studies, the electrode placements on the scalp are ruled by standard locations to make reproductible setups. The most common set of electrode positions is the 10/20 system [13], composed of 21 locations (represented in Fig. 2.2). Different sensors with different electrode placements, corresponding to different parts of the brain, are available in the market [14, 15, 16]. The number of electrodes and their placement on these devices depend on the amount of money invested in the research, the necessary level of precision and the area of the brain under investigation.

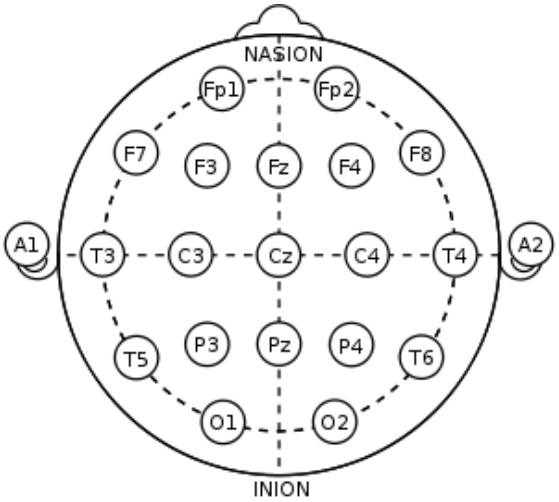

FIGURE 2.2: International 10/20 system electrode placement. [2] Each circle corresponds to an electrode placement location.

In our laboratory, the Muse headset [14], which is a low-cost, limited-precision sensor focusing on the prefontal activity, is used. This headset uses the 10-20 placement with

Modified Combinatorial Nomenclature, a modified version of the 10-20 model, for the placement of the 5 electrodes used. It is a higher resolution system, also known as the 10-10 placement. Muse headset has electrodes located on sites Fpz, AF7, AF8, TP9, and TP10. The electrode Fpz is used as the reference electrode, as shown in Fig. 2.3.

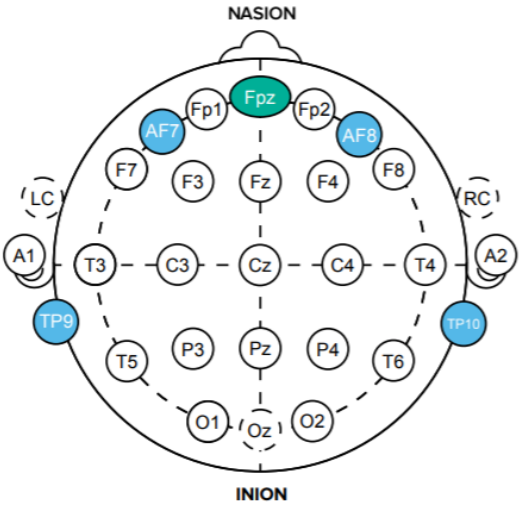

FIGURE 2.3: Electrode placement in Muse headset, using the 10-20 placement with Modified Combinatorial Nomenclature. The channels are shown in blue and the reference in green. [3]

The EEG signals are represented as waveforms on an electroencephalogram as shown in Fig. 2.4. The number of waveforms (i.e. number of channels) corresponds to the number of electrodes except the reference one. For instance, in our case, there are four channels.

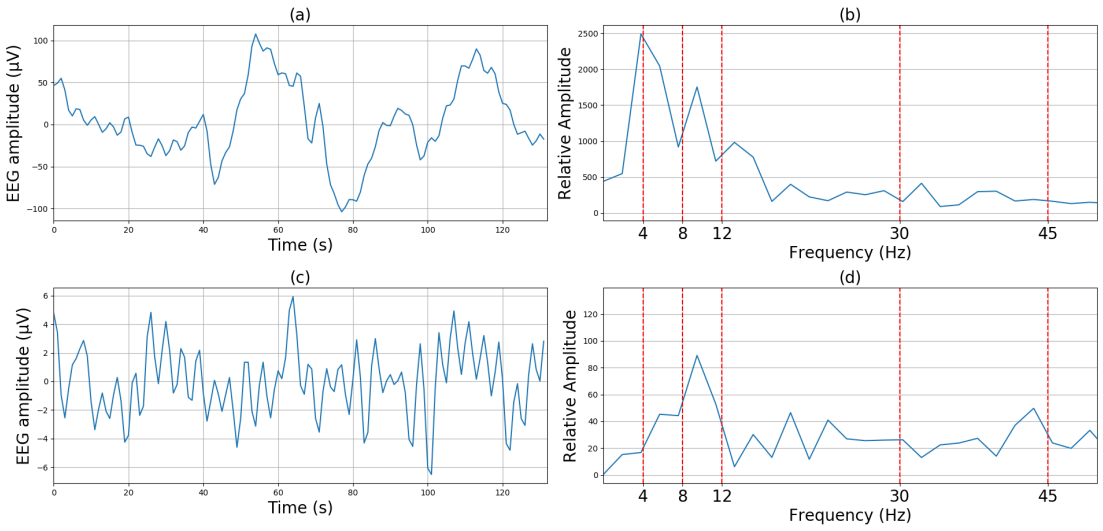

FIGURE 2.4: EEG data (a,c) in time domain, (b,d) in frequency domain.

2.2.1.2 Processing and interpretation of EEG

EEG data are generally processed and interpreted in various frequency ranges, which are called fundamental sub-bands. There are 5 fundamental sub-bands (namely alpha, beta, gamma, delta, theta) corresponding to the frequency ranges shown in Table. 2.1. In order to get the fundamental sub-bands, the raw EEG signal is transformed from time domain to frequency domain using Fast Fourier Transform (FFT). An example of FFT operation is illustrated in Fig. 2.4.

TABLE 2.1: Fundamental sub-bands and their frequency ranges.

	Frequency (Hz)
Delta	$[0, 4]$
Theta	$[4, 8]$
Alpha	$[8, 12]$
Beta	$[12, 30]$
Gamma	$[30, 45]$

The fundamental sub-band amplitudes are used to interpret the EEG data as they are correlated with emotional state [17] and its intensity. For instance, the presence of alpha correlates with how much relaxed the person is. On the other hand, theta band represents drowsiness in adults and teens, as it appears when mental state goes from fully-conscious to drowsy. Beta band reflects active thinking, focus and high alert state, as it appears when the focus is on the outside world or on solving problems. A high-level beta band may also reflect panic state.

2.2.2 Electrodermal activity

EDA is correlated with the activity of eccrine sweat glands, which are found on nearly all skin locations and in highest concentration on hands and fingertips [18, 4, 19]. Eccrine sweat glands are responsible for sweat secretion correlated with emotional response.

2.2.2.1 Modelisation of EDA

From an anatomical point, EDA reflects two basic electrical properties of the skin, namely skin conductance level (SCL), also termed as *tonic level*, and skin conductance

response (SCR), also termed as *phasic response*. An EDA signal y can be decomposed as,

$$y = r + t + \epsilon, \tag{2.1}$$

where r and t stand for phasic response and tonic level, respectively, and ϵ is the noise term. t changes slowly and continuously depending on the skin conductance, whereas r changes faster and is considered as small waves superimposed on t. SCR is sensitive to a wide range of factors such as stimulus novelty, intensity, or affective content. In this study, we focus on intensity of affective content (i.e. arousal).

In our study, we used the Mindfield eSense Skin Response with Ag/AgCl disposable electrodes to measure EDA [20]. It measures the skin conductance in microsiemens (μS) using 2 electrodes.

2.2.2.2 Markers of EDA

Several features can be extracted from the phasic response and tonic level to interpret EDA [4]. Tonic level can be used to investigate the general state of arousal/alertness of a human being. From the phasic response, a number of markers are derived such as latency, amplitude or rise time as listed in Table 2.2, which can be used to study attention and stimulus significance. These markers are represented in Fig. 2.5

Table 2.2: Phasic response markers [4].

Marker	Definition
Amplitude	Increase of skin conductance shortly after a stimulus
Latency	Time interval between stimulus and SCR initiation
Rise time	Time interval between stimulus and SCR Peak
Half recovery time	Time interval between SCR peak and 100% recovery of SCR amplitude
Habituation	Number of stimulus presentations before two or three trials with no response

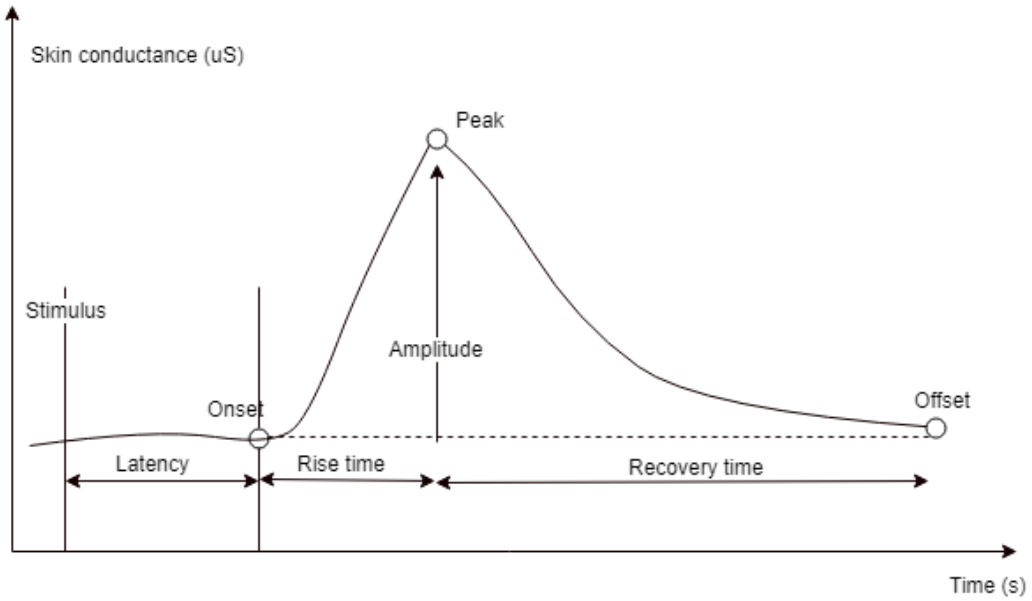

FIGURE 2.5: Components of an SCR [4].

2.3 Related work

Several studies investigate the use of PS (PS) for estimating affect scores of stimuli. PS such as EDA and EEG, are often studied to estimate emotions induced by visual or acoustic stimuli. For instance, Gerdes et al. [21] study brain activations against visual stimulus, where Trochidis et al. [22] study the link between acoustic stimulus and EDA, respiration rate and blood volume pulse.

Other reactions of the human body can be investigated to study fatigue and monitor responsiveness. Some of the recent studies focusing on fatigue to monitor human responsiveness include [23, 24, 25]. Fatigue is defined as a loss of efficiency and disinclination to effort, leading to decreased productivity in the workplace and induces critical errors in the worst cases. To induce fatigue, human subjects are asked to perform long tasks. The task length is variable as it may go from a simple reading task to total sleep deprivation. Cheng et al. [26] use EEG to monitor participants' fatigue. Wascher et al. [27] also use EEG but focused specifically on frontal theta activity. Miro et al. [28] observe the effect of 48h sleep deprivation on body temperature and reaction time. In our research we focus on engagement index and emotional state to understand human body responsiveness but several

Some studies associate mental workload with human responsiveness. Mental workload correlates with a task difficulty. The more difficult the task is, the higher the mental workload is. Ma et al. [29] observe the differences in EEG and EDA signals of participants with varying levels of experience.

PS can be coupled with Neural Network (NN) to predict an emotional state as well. Chanel et al. [30] and McFarland et al. [31] use NN to estimate subjects' emotional states from EEG. Kim et al [32] build a Deep NN (DNN) to predict affective levels of visual stimuli from color, foreground, background features. Peng et al. [33] develop NN to estimate affect levels of also acoustic stimuli.

Multimodal model NN can be created with multiple types of input. Hu et al. [34] use a neural network with two types of input to predict the arousal level of picture. Instead of feeding the NN with only images, they also feed it with the tags of the images.

Other machine learning algorithms can be used instead of NN to predict affective levels of stimuli as well. Gurudath et al. [35] use a K-means clustering algorithm with EEG data as input to detect and predict fatigue of a driver. Shen et al. [36] and Yeo et al. [37] implement a Support Vector Machine for automatic detection of fatigue in EEG.

In addition, PS such as EEG [38] and fNIRS [39] are used in interaction robotics to estimate user's attention and/or stimulate his engagement or estimate several cognitive or affective states. One important application is robot tutors using EEG cues [38] to improve student engagement and motivation. Drachen et al. [40] use EDA to monitor the arousal level of subjects with different experience level for video games. They intend to use the results to improve game development and to keep the player aroused during his gaming session. EDA is also used in [37, 25] to create a driving assistant that can detect the driver drowsiness or the attention level to prevent accidents.

Chapter 3

Datasets and experimentation

The goal of the experiments is to collect data to predict user's state of responsiveness from PS. Therefore, we collect two datasets, where subjects perform a given task over a prolonged duration of time. In one experiment, we collect EEG data as PS, whereas in the other experiment EDA. We use EEG to derive the engagement index of the subjects and EDA to study their arousal levels as a representation of responsiveness.

To gather the data needed, two experiments were set up: one to track EEG of a participant during several tasks to monitor his responsiveness level and the other one to track EDA of a participant over a prolonged exposure to emotional stimuli to infer responsiveness.

In this chapter we will present the details of experimentation and datasets.

3.1 EEG data collection

The goal of this experiment is to gather EEG data for evaluating the subject's responsiveness, while he is performing different tasks. Five human subjects, one male and four female, aged between 21 and 28, took part in the experiment. Each subject performs 3 kinds of tasks. The duration of each task is approximately 3 hours without any breaks. Namely, we gather a total of 45 hours of EEG recordings. We expect the subject responsiveness to decrease in each session as he becomes tired of the tasks.

3.1.1 Protocol

Each experiment involves one of the 3 tasks:

- Passive task: The subject views a slideshow composed of images taken from Judd dataset [41], a saliency benchmark, composed of 1003 images of natural indoor and outdoor scenes.
- Active test: The subject plays Wisconsin card sorting game. In this game, 4 cards a shown to the subject and he is asked to match another card with one of these 4 cards (example in Fig. 3.1). He is not told how to match it but the possibilities are matching by sign, color or number. Once he makes a guess, he is told if his guess is right or not. The rule of the game changes at every 5-10 questions. Therefore, the subject needs to stay focused and discover new rules.

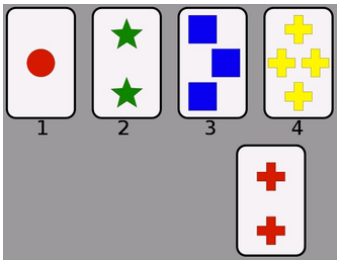

FIGURE 3.1: Example of Wisconsin card game [5].

- Semi-active: The subject listens to a story and, afterwards, is asked to answer questions about the story. Each story is about 3-7 min long, while the question is displayed for 15 seconds. We call this kind of task semi-active, since the listening period is passive and the answer period is active.

3.2 EDA data collection

Four human subjects, one female and three male, aged between 21 and 37, took part in the experiment.

3.2.1 Protocol

We adopt a discrete stimulus paradigm and present emotionally significant visual stimuli in the form of a slide show of 200 images, where each image is displayed for 5 sec followed by a 5 sec reset period as explained in Fig. 3.2.

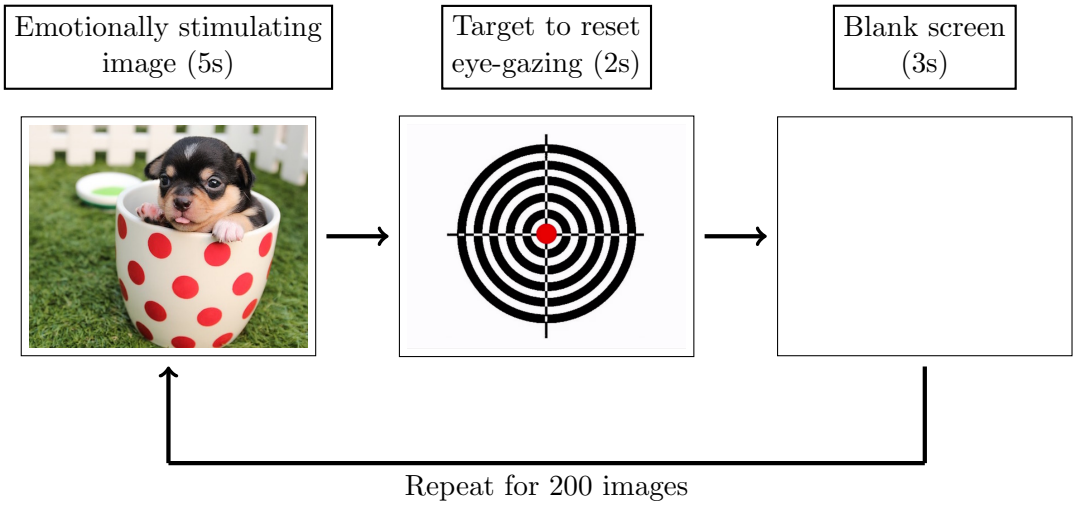

FIGURE 3.2: Slideshow presented to the participant during the experiment.

The length of the experiment is about 33 min. As the subject watches the slide show, EDA is logged by two electrodes on the index and middle fingers' distal phalanges of the non-dominant hand. Before and after the slide show, the participant is asked to fill a consent form and a NASA Task Load Index form.

3.2.2 Selection of visual stimuli

We search for a correlation between PS and sensitivity to emotional stimuli. Thereby, two emotional stimuli datasets with valence and arousal ground truth labels were used: the Geneva affective picture database (GAPED) [42] and the Open Affective Standardized Image Set (OASIS) [43]. Table 3.1 contains insights about the datasets.

TABLE 3.1: GAPED and OASIS original insights.

	Number of images	Valence/Arousal scale	Image size
OASIS	900	[1:7]	500×400
GAPED	730	[0:100]	640×478

The affect score ranges and image resolutions of GAPED and OASIS are not identical (See Table 3.1). In order to have uniform values, we did some preprocessing operations on GAPED. Namely, we scaled the affect scores from $[0, 100]$ to $[1, 7]$ linearly. In addition, the images are scaled to a $[500, 400]$ resolution. Moreover, some specific images (i.e. 160 images of spiders and 133 images of snakes) are discared.

(a) (b) (c)

FIGURE 3.3: Sample images from the OASIS dataset: (a) Low arousal and medium valence, medium arousal and high valence, (c) high arousal and low valence.

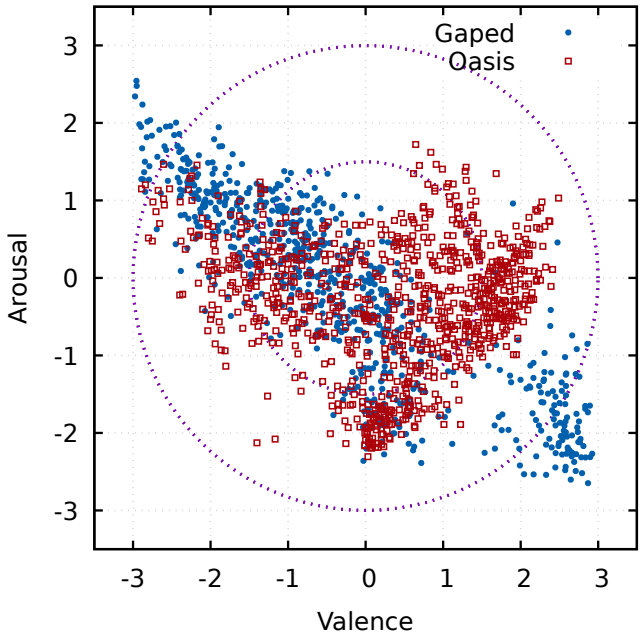

FIGURE 3.4: Distribution of Valence and Arousal labels of OASIS (BLUE), GAPED (RED). Each dot corresponds to one image.

OASIS is employed as the source of visual stimuli to be used in the experiments with human subjects, whereas GAPED is employed in detecting cross-set performance rates. Namely, the performance of our model, which is developed using OASIS, is confirmed by testing it on GAPED.

The stimuli (i.e. images) are selected from OASIS randomly. In particular, 200 images are selected from OASIS, such that the Valence/Arousal space is divided into 4 quadrants

and approximately same number of images are taken from each quadrant (see Fig. 3.5).

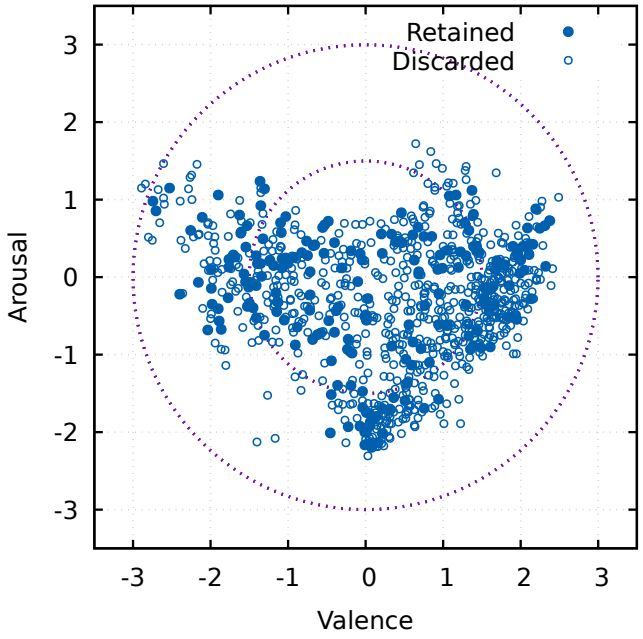

FIGURE 3.5: Distribution of valence and arousal scores of the images in OASIS. Each circle correspond to one image. The 200 images randomly taken for the experiment are represented in plain circle.

Chapter 4

Processing and results

4.1 EEG

We processed the EEG data collected during the experiments by filtering it and applying an FFT to extract the fundamental sub-bands. We first filtered the signal for each channel by applying a 60 Hz Notch filter to remove interference from main power and a 1-50 Hz bandpass filter to suppress the DC offset of the signal due to electrode-scalp interface. Afterwards, we computed FFT with a time window of 500 ms. For each channel C_i, alpha, beta and theta band amplitudes (α_i, β_i, θ_i) are computed. Finally, head wide alpha, beta and theta band amplitudes (α, β, θ) are computed as average of corresponding amplitudes in each channel. For example,

$$\alpha = \frac{1}{N}\sum_{i=1}^{N}\alpha_i, \tag{4.1}$$

where N is the number of channels.

Using the average head wide α, β and θ band amplitudes, we computed an engagement index [38] which represents level of attention/interest of a subject during a task.

$$E = \frac{\beta}{\alpha + \theta}. \tag{4.2}$$

We made the hypothesis that the engagement should decrease over time as the attention level of the subject is expected to decrease since he is likely to be be tired or bored of the experiment as time elapses.

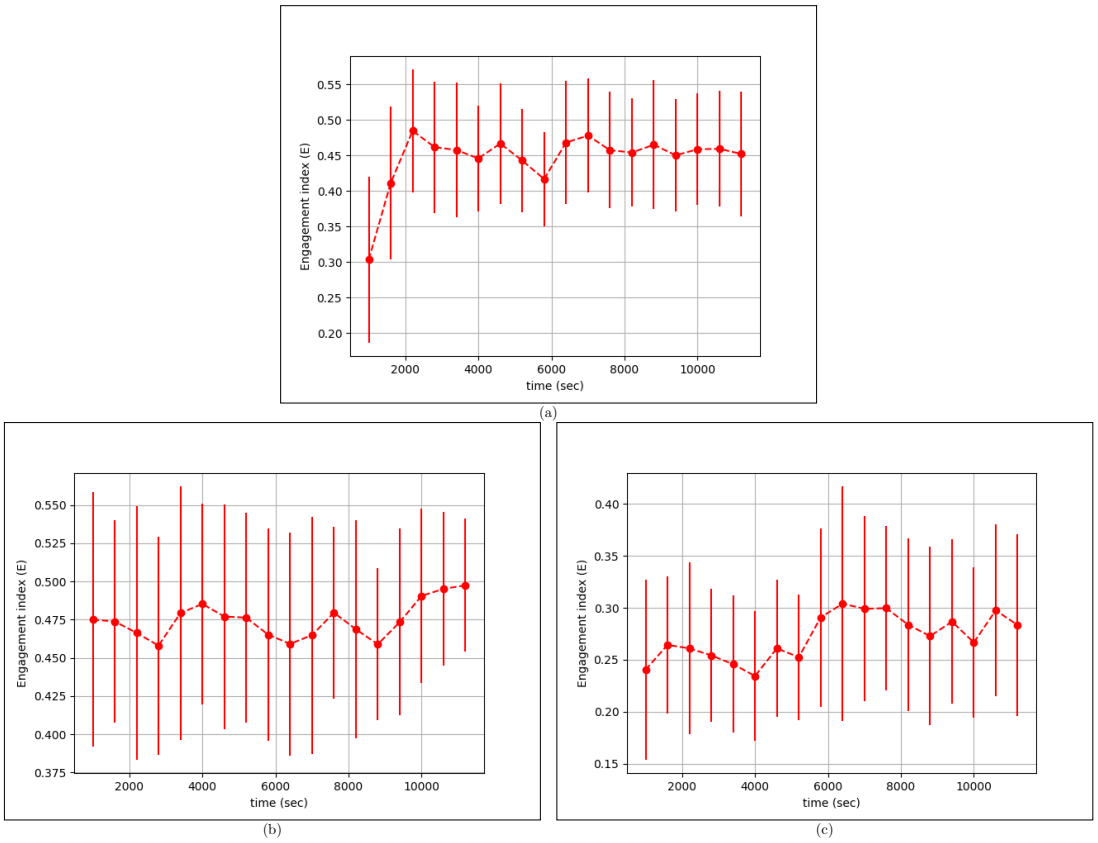

FIGURE 4.1: Engagement index computed for (a) active task, (b) passive task and (c) semi-active task.

The engagement indices for the 3 tasks, each for a different subject, are shown in Fig. 4.1. Based on these results, we cannot prove that the engagement index is decreasing over time for any of the 3 tasks. Therefore, we decided to focus on EDA to monitor the responsiveness of the subject.

4.2 EDA

In this section, we elaborate on the proposed method for modelling the effect of responsiveness on EDA. Namely, we propose an attenuation model to reflect the decrease in responsiveness over time.

We also consider two reference methods for comparing the proposed approach. Since sensation depends to a large extend on personality, it is not uncommon for human coders to have disagreements. Therefore, in order to estimate the level of disagreement, which is inherent in the ground truth labels of OASIS, we carry out a Monte Carlo (MC) simulation as a first reference method in Sec. 4.2.2. In addition, we develop a DNN

based estimator to provide a comparison with a broad-gauge approach (as described in Sec. 4.2.3).

4.2.1 EDA processing

As explained in Sec. 2.2.2.2, a number of markers are used such as latency, amplitude, rise time etc [4] in order to interpret EDA. For estimating the emotional response, the most helpful marker is suggested to be the amplitude of peaks in the phasic component r (See Eq. 2.1) [44].

Therefore, as a first step, it is necessary to decompose the EDA signal and obtain the phasic response r. For decomposition of EDA as in Eq. 2.1, we use cvxEDA, a convex optimization approach proposed by Greco et al. [44]. The tonic component is modeled as a linear combination of cubic spline functions B, together with an offset with a linear term D:

$$t = Bl + Dd,$$

where l and d are the coefficient matrices. On the other hand, r is modeled with a Bateman function,

$$h(\tau) = (\exp(-\tau/\tau_0) - \exp(-\tau/\tau_1))\, u(\tau), \tag{4.3}$$

where u is the unit step. Eq. 4.3 is represented with an ARMA model in frequency domain enabling a convex optimization approach to estimate each component.

After decomposing y into its components, we focus on the amplitude P of the peaks in the phasic response r, which is known to be sensitive to the arousal of stimuli [44]. We model the relation between P and the annotated arousal levels A accounting for the time elapsed from the start of user's task.

Thus, evolution of responsiveness is represented using the attenuation model in Eq. 4.4. Namely, P is modeled as a function of arousal score A and display time τ of the stimulus. P is subject to an exponential rate of decay with τ:

$$P(A, t) = A \exp(a\tau + b) + C, \tag{4.4}$$

where a, b and C are the parameters to be calibrated.

For testing the proposed method, EDA signal is divided into 3 minute long sections, corresponding to batches of 18 images. For each image in a particular batch, the amplitude of peaks P is computed using cvxEDA. A subset of 9 images is randomly chosen and corresponding P values are used to calibrate Eq. 4.4 by minimizing the squared error. Using the resulting $\{a, b, C\}$, the arousal score A' of the remaining subset is computed. This random selection process is repeated 100 times for all batches and all subjects.

In order to test whether the responsiveness in EDA is attenuated according to the model in Eq. 4.4 (due to a possible fatigue etc), we adopt a reverse estimation strategy. That is, we assume that the model in Eq. 4.4 works efficiently and we estimate an arousal level A' for each image, given the amplitude of peaks P and the display time τ,

$$A' = \frac{P - C}{\exp(\alpha\tau + \beta)}.$$

This reverse estimation strategy enables an objective evaluation of performance. In other words, since the arousal level A of the stimulus (i.e. image) is provided as ground truth in OASIS, A' can be compared to A using one of the conventional metrics.

To evaluate the performance we compute two metrics, namely Sign Agreement Metrics (SAGR) and Root Mean Square Error (RMSE).

The SAGR is defined as,

$$SAGR = \frac{1}{n}\sum_{i=1}^{n} \delta(sign(\hat{\Theta}_i), sign(\Theta_i)), \tag{4.5}$$

where δ is the Kronecker delta function,

$$\delta(p, q) = \begin{cases} 1, \; p = q \\ 0, \; p \neq q \end{cases}. \tag{4.6}$$

Namely, SAGR considers a prediction to be correct, if the sign of the ground truth label and the sign of the predicted label are the same.

The RMSE,

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\Theta}_i - \Theta_i\right)^2}, \tag{4.7}$$

measures how concentrated the data is around the line of best fit.

4.2.2 Monte-Carlo Simulation

The first reference method is derived directly from the statistics provided with the ground truth of OASIS. Namely, each image in OASIS, I_i, is evaluated by $N_i > 100$ coders. However, instead of each label $\{A_{ij}, j \in [1, N_i]\}$, OASIS provides the mean, μ_i, and standard deviation, σ_i for each image I_i.

The optimization procedure described in Sec. 4.2.1, considers μ_i as the true of arousal score of I_i. Nevertheless, σ_i indicates that there is a certain amount of disagreement between coders. In order to evaluate this disagreement, we take a MC simulation standpoint. Namely, we compare m pairs of annotations sets, both randomly drawn from the ground truth distribution, under the assumption that the corresponding levels of arousal $\{A_{ij}\}$ for a given image I_i come from a normal distribution such that,

$$A_{ij} \sim \mathcal{N}(\mu_i,\, \sigma_i^2). \tag{4.8}$$

For each pair of annotation sets, the corresponding SAGR as in Eq. 4.5 and average it over the m pairs. In addition, we compute the corresponding RMSE values and provide an insight to the inherent disagreement present in the ground truth.

4.2.3 Convolutional Neural Network

The second reference method uses a CNN architecture, which is a compositional model, building potentially more complex information from primitives in a layered manner. CNNs build feature maps going from lower to higher level features. Although DNNs are used to estimate affective content of a stimulus [21], [45], they are trained with data from a large number of subjects and thus are often not specific to users. For this reason, they can be used as some broad-gauge approach to estimate an anticipated response to emotional stimuli.

4.2.3.1 CNN architecture

Convolutional NNs (CNN) have recently become ubiquitous in many computer vision fields such as object detection and recognition. A convolutional neural network is composed of two components: the feature extraction component and the fully-connected layer [46].

The feature extraction component is composed of 3 parts:

- Convolutional layer extracts a set of features by applying a convolution operation on the image using a 3×3 (px) filter. The filter is slided pixel by pixel from left to right and from the top to the bottom of the image.
- Rectified Linear Unit layer (ReLu) applies a max(0,x) operation on all pixels to introduce non-linearity in the convolutional network.
- Pooling layer reduces the dimension of the input in order to minimize the number of features that will be fed to the fully-connected layer.

The fully-connected layer is responsible for the prediction. It is a basic neural network architecture, where all the neurons from the hidden layer n are connected to all the neurons from the hidden layer $n + 1$. The last layer is composed of a single or multiple activation function based on the architecture. If the output is a discrete label, there are two possibilities:

- The last layer of the FC layer has a neuron for each label, which means we have a sigmoid function for each label, responsible for computing a score in range [0,1]. The label with the highest score will be the predicted label. For instance, if we have a score of 0.82 for label 'dog' and a score of 0.12 for label 'cat', the prediction for this picture would be 'dog'. This method is used if we have 2 or more labels.
- If we have only two possible labels, instead of having two neurons in the last layer of the FC, we can have one neuron with a sigmoid function. For the example mentioned above, if the score is lower than 0.5 the prediction would be 'dog', else the prediction would be 'cat'.

If the output is a continuous label, the input label has to be rescaled in range [0,1]. Then the last layer of the FC layer will be composed of a single neuron that will compute the score. This final score will then be rescaled in the original range to get the prediction.

4.2.3.2 Transfer Learning

Training a CNN from scratch needs millions of images, which is time consuming. Moreover, we do not have a data set big enough to train a CNN (i.e. 900 images). Due to time limitation and the small dataset size, it is advised to use transfer learning, which consists of taking an existing model trained on a large dataset, removing the last fully-connected layer and replacing it with proper layers, depending on the goals.

Since our emotional stimuli are visual, we consider Inception [47] to be a suitable starting point for adapting it to estimate affect scores via transfer learning. It is a model trained with 14 million images from Imagenet database [48] to classify images into thousands classes and is a standard model used for transfer learning. We replace the final classification layer of Inception with a bunch of new layers, which are trained with the images from OASIS to estimate their arousal scores.
The goal of the new model is to predict affective quality of an image. Hypothesis is that features related to this are low level features, like local intensity variations such as a key point or an edge in the image. These low level features can then be combined to make a meaningful classification of the image [49].
In order to extract low level features from the images, we add two convolutional layers, each with a ReLu activation function and a pooling layer. After these layers, a flatten layer is added to turn the multi-dimensional feature table into a (one dimensional) column of features that are fed to a fully connected (FC) layer followed by a dropout layer that will randomly set 50% of the values to 0 so as to prevent over-fitting. Finally, a last FC layer with a one dimensional output is added with a sigmoid activation function. This layer estimates the affective score of an input image with a continuous value between $[0, 1]$. This value is either rescaled to $[1, 7]$ in case of RMSE cost function or considered as positive or negative, in case of SAGR cost function.

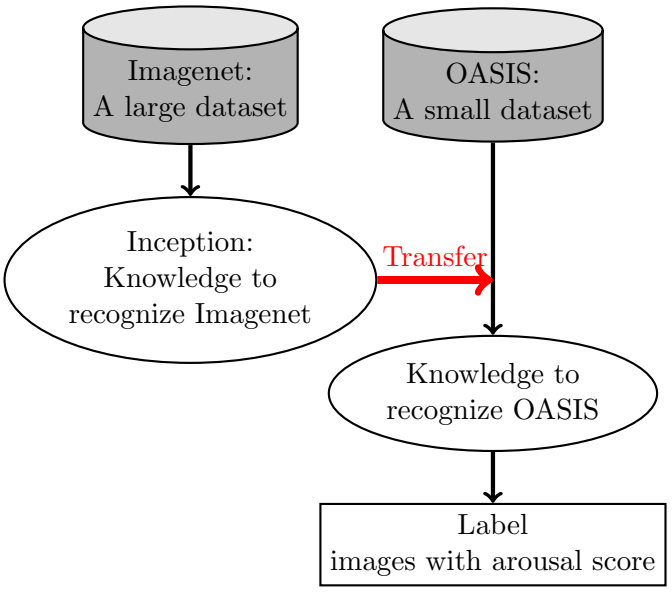

FIGURE 4.2: Transfer learning

TABLE 4.1: Comparison of performance.

	SAGR (%)	RMSE
MC-200	41.7 ± 0.01	2.21 ± 0.01
MC-900	43.6 ± 0.01	2.19 ± 0.01
CNN-200	55.5 ± 0.16	$\mathbf{1.76 \pm 0.10}$
CNN-900	$\mathbf{62.6 \pm 0.12}$	1.98 ± 0.04
CNN-200 (CS)	50.6 ± 0.7	2.21 ± 0.03
CNN-900 (CS)	59.9 ± 0.05	2.15 ± 0.04
Proposed	53.6 ± 0.01	2.5 ± 0.01

4.3 Results

Table 4.1 presents the average performance of the MC, CNN and proposed method. The detailed performances are shown in appendix A. MC and CNN are tested using (i) the 900 images in entire OASIS data set (MC-900 and CNN-900) and (ii) the 200 images which are used as visual stimuli in our experiments (MC-200 and CNN-200), in order to reflect the impact of arousal range of stimuli.

Moreover, CNN-200 and CNN-900 are also tested on GAPED to determine the cross-set (CS) performance. This helps to understand how well the CNN generalizes. As shown in Fig. 3.5, most images used in the experiment have an arousal in $[-2, 2]$, meaning that a small error in A' may change its sign and degrade the performance in terms of SAGR. This is why we chose to compute both SAGR and RMSE to interpret the error.

Table 4.1 shows that the ecological agreement is virtually the same in different arousal ranges (MC-200 and MC-900) in terms of both SAGR and RMSE. This is due to the high standard deviation of the ground truth scores (i.e. some people are extremely sensitive or insensitive to the same stimuli) and translates the difficulty for human beings to agree on the arousal score of the same stimuli.

The effect of arousal range is obvious on CNN-200. Although the mean SAGR (55.5) seems to improve, it has a high standard deviation (0.16). Most tests achieve around 40% and few tests achieve about 90%, which increase the mean performance misleadingly, and prove the over-fitting issue. Therefore, also the low RMSE (1.76) can be explained by the learning of the input range rather than real affective features. Thus, the higher SAGR and lower RMSE values are misleading in interpreting performance of CNN-200.

The proposed method achieves an SAGR of 53.6% with a very small standard deviation, and estimates the *person-specific arousal* better and more stably than CNN-200. The RMSE is quite high since the output range in Eq. 4.4 is not bounded. Therefore, it estimates the sign of arousal score correctly more often but its amount is overestimated. However, since affect analysis considers the sign to be more important than absolute error, we can say that a higher SAGR is desired over a lower RMSE.

In case of CNN-900, SAGR is found to be 62.6 ± 0.01 and RMSE is 1.98 ± 0.04. Obviously and as expected, larger amount of training samples improve the performance of CNN considerably as it helps to solve the over-fitting issue. However, we consider it unfair to compare these values to the proposed method, since our performance is evaluated using a small set (200 images due to time limitation), but is one of the future research directions to diversify the affect range of the stimuli.

We can remark that CNN method generalizes well if trained with enough data. Indeed, the SAGR for the CNN-900 (CS) is close to the SAGR for the CNN-900.

In addition, we searched for studies estimating arousal scores of OASIS but we could not come across any such studies. Actually, most studies use the images as stimuli and analyze their impact on various PS. In this sense, perhaps the most relevant study to our is by Hu et al [34], who achieve around 40% success rate using image features and offer to integrate image tags to improve recognition of affect qualities.

Chapter 5

Conclusion and Future work

5.1 Conclusion

The goal of this study is to search for correlations between human responsiveness and PS, namely EEG and EDA, and propose method for detecting level of responsiveness automatically from the PS.

We started by doing literature search in order to understand the concept of these PSs and how they could be used to monitor responsiveness. We made the hypothesis that Engagement Index and Arousal level can be used as indicators of human responsiveness. Several experiments are carried out to gather data EEG and EDA data for testing this hypothesis.

We started by studying the Engagement Index using EEG. We made the hypothesis that it should decrease as the subject should lose attention or get bored of the task as time elapses. We saw that this hypothesis could not be proved as the engagement indices computed for 3 different tasks and for 5 different subjects were not monotonically decreasing over time.

Therefore, we decided to focus on EDA. We proposed a parametric model for evolution of responsiveness using EDA signal. We consider the arousal level of the subject as a representation of his responsiveness. In order to represent the decline in responsiveness, we proposed an attenuation model. For evaluating performance, we adopted a reverse

strategy. Namely, we assumed the model to work efficiently and we estimated arousal scores of the stimuli accounting for the decreasing sensitivity.

We compared the proposed method performance against two reference methods, namely MC simulation and CNN. The proposed approach is found to give a more accurate estimation of arousal score than CNN-200. In addition, although CNN-900 has better performance rates, a direct comparison is not fair, since the input sets are different.

5.2 Future work

Several modifications can be made on the proposed model to improve its performance and other methods can be implemented to explore responsiveness.

First, we suggest to diversify the affect range of the stimuli. To do so, we can merge different emotional stimuli datasets in order to get a higher number of images to train the NN or to fine-tune the proposed method. Other datasets such as the International Affective Picture System (IAPS) can be use in addition to OASIS and GAPED.

More precise EEG and EDA sensors can be incorporated. Actually, we started to explore the use of the Ultracortex "Mark IV" EEG Headset [50], which contains up to 16 channels against 4 channels of the Muse headset. In addition, Ultracortex "Mark IV" comes with a complete software, whose source code can be modified in Java. As a result, it is a standard equipment used in many EEG studies. A higher number of electrodes can improve data quality, as a wider part of the brain is observed. Moreover, these electrodes are mobile, allowing us to observe specific parts of the brain we want to focus on.

Concerning EEG, other features than the engagement index could be studied. As a matter of fact, we could observe the amplitude of the sub-bands individually. We could also explore if differences exists between the amplitude of the sub-bands from different areas of the brain by using an other EEG device.

Another way to improve our method would be to embed the model of responsiveness we proposed into a NN architecture. The NN model could also be improved, in the manner of Hu et al. [34], by using different types of input. For instance, input such as image tags and PS data can be use. We could also change the features detected in the image, instead of focusing only on low level features. For example, we can use object detection

to estimate the arousal score of an image. For example, if the image contains a space rocket, the picture would potentially be labelled with a high arousal score. The color of the image may also be used, knowing that in the OASIS dataset, the pictures containing fire are rated with mid-high arousal. We can also make the hypothesis that black and white images may be less exciting than colorful images. Other machine learning methods can also be implemented such as Support Vector machine, as presented in the related work section.

Chapter 6

Financial estimation

In this chapter we list the material and resources used during my research at Okayama University.

1. Physical resources:
   - **EEG cap**: 30$ - 25 €
   - **EDA sensors**: 37 000 JPY - 282 €
2. IT resources:
   - **Personal computer**: 1000 €
3. Intellectual resources:
   - **Scientific literature**: Google Scholar and Okayama University online library.
4. Salary:
   - **Monthly income**: 30 000 JPY/Month - 228 €

In this financial estimation I did not take into account the time that Dr. Yucel spent with me which of course counts as human resource.

Appendix A

Results

Detailed performance results on 8 runs for CNN-200, CNN-200 (CS),CNN-900, CNN-900 (CS) using SAGR and RMSE.

Run index	1	2	3	4	5	6	7	8
CNN-200 SAGR	55.0	42.6	90.5	47.5	44.5	42.5	47.5	82.8
CNN-200 (CS) SAGR	37.5	59.6	53.9	57.4	58.3	43.7	43.2	51.7

Table A.1: CNN-200 and CNN-200 (CS) SAGR results for 8 runs

Run index	1	2	3	4	5	6	7	8
CNN-900 SAGR	62.6	62.6	62.3	64.6	61.2	60.6	61.2	61.7
CNN-900 (CS) SAGR	60.2	61.3	59.6	59.6	59.6	59.6	59.6	59.6

Table A.2: CNN-900 and CNN-900 (CS) SAGR results for 8 runs

Run index	1	2	3	4	5	6	7	8
CNN-200 RMSE	1.67	1.74	1.74	1.80	1.76	1.98	1.64	1.75
CNN-200 (CS) RMSE	2.22	2.19	2.17	2.18	2.22	2.24	2.27	2.22

Table A.3: CNN-200 and CNN-200 (CS) RMSE results for 8 runs

Run index	1	2	3	4	5	6	7	8
CNN-900 RMSE	1.96	1.95	1.94	1.92	2.06	1.96	2.04	1.95
CNN-900 (CS) RMSE	2.12	2.18	2.13	2.10	2.13	2.23	2.19	2.16

Table A.4: CNN-900 and CNN-900 (CS) RMSE results for 8 runs

Bibliography

[1] "Two-dimensional valence-arousal space." https://www.researchgate.net/figure/Two-dimensional-valence-arousal-space_fig1_304124018, accessed July, 2018.

[2] "10-20 system (EEG)." https://en.wikipedia.org/wiki/1020_system_(EEG), accessed July, 2018.

[3] "Muse electrode locations." http://developer.choosemuse.com/, accessed June, 2018.

[4] M. E. Dawson, A. M. Schell, and D. L. Filion, "The electrodermal system," *Handbook of psychophysiology*, vol. 2, pp. 200–223, 2007.

[5] "Wisconsin card game example." https://en.wikipedia.org/wiki/Wisconsin_Card_Sorting_Test, accessed July, 2018.

[6] P. Ekman, "An argument for basic emotions," *Cognition & emotion*, vol. 6, no. 3-4, pp. 169–200, 1992.

[7] J. A. Russell, "A circumplex model of affect," *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980.

[8] G. Ball and J. Breese, "Modeling the emotional state of computer users," in *Proceedings of the Workshop on Personality and Emotion in User Modelling*, 1999.

[9] P. J. Lang, M. K. Greenwald, M. M. Bradley, and A. O. Hamm, "Looking at pictures: Affective, facial, visceral, and behavioral reactions," *Psychophysiology*, vol. 30, no. 3, pp. 261–273, 1993.

[10] P. J. Lang, "The emotion probe: Studies of motivation and attention," *American psychologist*, vol. 50, no. 5, p. 372, 1995.

[11] P. Rani, N. Sarkar, and C. Liu, "Maintaining optimal challenge in computer games through real-time physiological feedback," in *Proceedings of the 11th international conference on human computer interaction*, vol. 58, pp. 22–27, 2005.

[12] J. H. Westerink, E. L. Van Den Broek, M. H. Schut, J. Van Herk, and K. Tuinenbreijer, "Computing emotion awareness through galvanic skin response and facial electromyography," in *Probing experience*, pp. 149–162, Springer, 2008.

[13] V. Jurcak, D. Tsuzuki, and I. Dan, "10/20, 10/10, and 10/5 systems revisited: Their validity as relative head-surface-based positioning systems," *NeuroImage*, vol. 34, pp. 1600–1611, 2007.

[14] "Muse headset." http://www.choosemuse.com/, accessed July, 2018.

[15] "Emotiv EPOC+." https://www.emotiv.com, accessed July, 2018.

[16] "OpenBCI." http://openbci.com/, accessed July, 2018.

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

[18] J. Marieke van Dooren, G.-J. de Vries, and J. H. Janssen, "Emotional sweating across the body: Comparing 16 different skin conductance measurement locations," *Physiology & behavior*, vol. 106, no. 2, pp. 298–304, 2012.

[19] B. Figner, R. O. Murphy, *et al.*, "Using skin conductance in judgment and decision making research," *A handbook of process tracing methods for decision research*, pp. 163–184, 2011.

[20] "Mindfield eSense skin response." https://www.mindfield.de, accessed July, 2018.

[21] A. Gerdes, M. J. Wieser, A. Mühlberger, P. Weyers, G. W. Alpers, M. M. Plichta, F. Breuer, and P. Pauli, "Brain activations to emotional pictures are differentially associated with valence and arousal ratings," *Frontiers in human neuroscience*, vol. 4, p. 175, 2010.

[22] K. Trochidis, D. Sears, D.-L. Trn, and S. McAdams, "Psychophysiological measures of emotional response to romantic orchestral music and their musical and acoustic correlates," in *From Sounds to Music and Emotions*, Springer, Jan. 2013.

[23] L. J. Trejo, K. Kubitz, R. Rosipal, R. L. Kochavi, and L. D. Montgomery, "EEG-based estimation and classification of mental fatigue," *Psychology*, vol. 6, no. 05, p. 572, 2015.

[24] B. T. Jap, S. Lal, P. Fischer, and E. Bekiaris, "Using EEG spectral components to assess algorithms for detecting fatigue," *Expert Systems with Applications*, vol. 36, no. 2, pp. 2352–2359, 2009.

[25] S. K. Lal and A. Craig, "A critical review of the psychophysiology of driver fatigue," *Biological psychology*, vol. 55, no. 3, pp. 173–194, 2001.

[26] S.-Y. Cheng and H.-T. Hsu, "Mental fatigue measurement using EEG," in *Risk management trends*, InTech, 2011.

[27] E. Wascher, B. Rasch, J. Snger, S. Hoffmann, D. Schneider, G. Rinkenauer, H. Heuer, and I. Gutberlet, "Frontal theta activity reflects distinct aspects of mental fatigue," *Biological Psychology*, vol. 96, pp. 57–65, 2014.

[28] E. Miro, M. Cano-Lozano, and G. Buela-Casal, "Electrodermal activity during total sleep deprivation and its relationship with other activation and performance measures," *Journal of Sleep Research*, vol. 11, no. 2, pp. 105–112, 2002.

[29] Q. G. Ma, Q. Shang, H. J. Fu, and F. Z. Chen, "Mental workload analysis during the production process: EEG and GSR activity," in *Applied Mechanics and Materials*, vol. 220, pp. 193–197, Trans Tech Publ, 2012.

[30] G. Chanel, J. Kronegg, D. Grandjean, and T. Pun, "Emotion assessment: Arousal evaluation using EEG and peripheral physiological signals," in *MRCS 2006*, pp. 530–537, Springer, 2006.

[31] D. J. McFarland, M. A. Parvaz, W. A. Sarnacki, R. Z. Goldstein, and J. R. Wolpaw, "Prediction of subjective ratings of emotional pictures by EEG features," *Journal of neural engineering*, vol. 14, no. 1, p. 016009, 2016.

[32] H.-R. Kim, Y.-S. Kim, S. J. Kim, and I.-K. Lee, "Building emotional machines: Recognizing image emotions through deep neural networks," *IEEE Transactions on Multimedia*, pp. 1–1, 2018.

[33] S. Peng, L. Zhang, Y. Ban, M. Fang, and S. Winkler, "A deep network for arousal-valence emotion prediction with acoustic-visual cues," *ArXiv e-prints*, May 2018.

[34] A. Hu and S. Flaxman, "Multimodal sentiment analysis to explore the structure of emotions," *arXiv preprint arXiv:1805.10205*, 2018.

[35] N. Gurudath and H. B. Riley, "Drowsy driving detection by EEG analysis using wavelet transform and k-means clustering," *Procedia Computer Science*, vol. 34, pp. 400–409, 2014.

[36] K.-Q. Shen, X.-P. Li, C.-J. Ong, S.-Y. Shao, and E. P. Wilder-Smith, "EEG-based mental fatigue measurement using multi-class support vector machines with confidence estimate," *Clinical Neurophysiology*, vol. 119, no. 7, pp. 1524–1533, 2008.

[37] M. V. Yeo, X. Li, K. Shen, and E. P. Wilder-Smith, "Can SVM be used for automatic EEG detection of drowsiness during car driving?," *Safety Science*, vol. 47, no. 1, pp. 115–124, 2009.

[38] D. Szafir and B. Mutlu, "Pay attention!: designing adaptive agents that monitor and improve user engagement," in *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 11–20, ACM, 2012.

[39] M. Strait and M. Scheutz, "What we can and cannot (yet) do with functional near infrared spectroscopy," *Frontiers in neuroscience*, vol. 8, p. 117, 2014.

[40] A. Drachen, L. E. Nacke, G. Yannakakis, and A. L. Pedersen, "Correlation between heart rate, electrodermal activity and player experience in first-person shooter games," 2010.

[41] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," in *Computer Vision, 2009 IEEE 12th international conference on*, pp. 2106–2113, IEEE, 2009.

[42] E. S. Dan-Glauser and K. R. Scherer, "The geneva affective picture database (GAPED): a new 730-picture database focusing on valence and normative significance," *Behavior research methods*, vol. 43, no. 2, p. 468, 2011.

[43] B. Kurdi, S. Lozano, and M. R. Banaji, "Introducing the open affective standardized image set (OASIS)," *Behavior research methods*, vol. 49, no. 2, pp. 457–470, 2017.

[44] A. Greco, G. Valenza, A. Lanata, E. P. Scilingo, and L. Citi, "cvxeda: A convex optimization approach to electrodermal activity processing," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 4, pp. 797–804, 2016.

[45] P. Tzirakis, G. Trigeorgis, M. A. Nicolaou, B. W. Schuller, and S. Zafeiriou, "End-to-end multimodal emotion recognition using deep neural networks," *IEEE STSP*, vol. 11, no. 8, pp. 1301–1309, 2017.

[46] "Cs231n convolutional neural networks for visual recognition." http://cs231n.github.io/convolutional-networks/, accessed July, 2018.

[47] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2818–2826, 2016.

[48] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pp. 248–255, IEEE, 2009.

[49] V. Gajarla and A. Gupta, "Emotion detection and sentiment analysis of images," *Georgia Institute of Technology*, 2015.

[50] "openBCI ultracortex-mark-iv." https://shop.openbci.com/collections/frontpage/products/ultracortex-mark-iv, accessed July, 2018.