論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Emotion recognition from EEG and individual characteristics in HCI scenarios
著者(和文)	MARTINEZ TEJADA Laura Alejandra
Author(English)	Laura Alejandra Martinez Tejada
出典(和文)	学位:博士(学術), 学位授与機関:東京工業大学, 報告番号:甲第12138号, 授与年月日:2021年9月24日, 学位の種別:課程博士, 審査員:小池 康晴,金子 寛彦,八木 透,吉村 奈津江,小尾 高史
Citation(English)	Degree:Doctor (Academic), Conferring organization: Tokyo Institute of Technology, Report number:甲第12138号, Conferred date:2021/9/24, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

Tokyo Institute of Technology
School of Engineering
Department of Information and Communication Engineering
Graduate Major in Human Centered Science and Biomedical Engineering

Emotion recognition from EEG and individual characteristics in HCI scenarios

Laura Alejandra Martinez Tejada

Submitted in part fulfillment of the requirements for the degree of

Doctor of Philosophy (Ph.D.)

Laura Alejandra Martinez Tejada: *Emotion recognition from EEG and individual characteristics in HCI scenarios*, Doctor of Philosophy (Ph.D), © July 2021

Supervisors:
Professor Koike Yasuharu
Professor Yoshimura Natsue

Location:
Tokyo, Japan

Abstract

Emotion Recognition is a process that uses affective models and different measurement methods to identify behavioral states in the human body and label them as emotions or affective states. Brain Computer Interphases (BCIs) have adopted the use of emotions recognition under the domain of affective BCIs, which attempt to create systems able to detect affective states from neurophysiological signals, to enhance the human-computer interaction (HCI).

The purpose of this research is to Identify emotional reactions using both electroencephalography (EEG) signals and participants' individual characteristics under HCI scenarios. EEG traits and individual characteristics were analyzed to identify which ones are useful in the emotion classification process (represented as arousal and valence labels) from public emotion recognition dataset. Then, a videogame to elicit different emotions, while recording EEG signals, self-assessment emotional answers, and behavioral cues was designed, to analyze the correlation, and classification between EEG traits, arousal-valence responses and videogame time events as behavioral cues related with emotional reactions.

Acknowledgements

I would like to express my gratitude to family: my mother, my sister, my father, and my grandmother, who have always given me all the support and encouragement throughout the course of my studies. Without their words and care I would have not reached this stage in my life.

I would like to show greatest appreciation to my advisor Professor Dr. Yasuharu Koike, and professor Dr. Natsue Yoshimura, for giving me the opportunity to become part of their laboratory, for their support and guidance during my doctoral process.

I would like to thank Koike & Yoshimura laboratory students and the staff members, for teaching me and helping me during this challenging process.

I would like to extend my deepest gratitude to the Japanese government for giving me the opportunity to come to Japan and prove myself in a different environment, this experience allowed me to grow as a student, as a professional, and as a human being.

I am greatly in debt with Alex Puertas, professor and friend, for his support, his comments, his occurrences, and his amazing attitude, without him, this work could not have been completed.

I am grateful for the companion of my Colombian, International and Japanese friends for their smiles and their helpful hand in times of need, for showing me that you can always encounter good people along the life's journey.

Dedication

Dedicated to family,

my mother, my sister, my father, and my grandmother.

Dedicated to all my past, present and future students,

they are my motivation to pursue new horizons in my academic journey

”””””””””””””””””””’

Cada logro comienza con la decisión de intentarlo. (Gail Devers)

Contents

List of Figures

Chapter 1

Introduction

Human computer interaction is the study field of interaction between users and computers and the design, evaluation and implementation of computer interfaces that are receptive to the user's needs and habits [1]. One of the key aspects on those interactions with technology are human emotions due to their involvement in multiple cognitive processes such as attention, perception, imagination, thinking, learning, memory, decision making, and problem-solving [2]. When interacting with technology, emotions can shape these cognitive processes, altering the experience and the interaction performance with the machine, this is why it is important to build technological systems that can assess emotion to enhance the user experience. With potential use in fields as perceptual information, human health, arts and entertainment, companion technology and assistive computing [3], emotion recognition becomes a potential tool that can benefit the development of better computer interfaces.

Emotion recognition using electroencephalographic signals is a recent and wider field of study where the researchers try to predict emotions using information derived from physiological signals. However, one of the many challenges faced by these kinds of researches is the selection of tools and data sources in the human body that allow to assess the emotions without being intrusive on the process they try to analyze and obtain higher accuracies on their emotion prediction.

1.1. Human computer interaction

Human computer interaction (HCI) is the field of computer science that concerned about "the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them" [4]. HCI groups several disciplines with different emphases: computer science (application design and engineering), psychology (cognitive processes and the empirical analysis of user behavior,) sociology and anthropology (interactions between technology, work, and organization), and industrial design (interactive products development). HCI focus on the study of the joint performance, and the communication structure between users and machines, together with the interface's usability, reliability and functionality [5].

1.2. Emotion Recognition

Emotion recognition is the process that uses affective models and measurement methods to identify human body states and categorize them as affective states. It aims to model the emotional interactions between a human and a computer by measuring the user behavioral cues related with emotions. Emotion recognition is a two-step process where: first, data is gathered from behavioral cues related to emotional states, like body movement or electrophysiological signals, using sensing devices (cameras, microphones, or physiological sensors). Second, the data is processed to analyze and identify changes and patterns related to the emotional states, applying machine learning, deep learning and data mining algorithms [6]. The initial goal of emotion recognition the design computer systems that can recognizes emotions and act according to this information [3].

1.3. Affective brain computer interfaces

Brain computer interfaces (BCIs) are technological communication systems that allow users to interact with external devices using brain activity [7]. A BCI system can monitory and preprocess brain signals, to extract information and translate it into commands to execute a particular action. The result can be perceived by the user who can modulate his brain activity to accomplish his intents [8]. BCIs have adopted the use of emotions recognition under the domain of affective brain computer interfaces (aBCIs), which attempt to create devices able to detect affective states from neurophysiological signals to enhance the human-computer interactions [9]. Neurophysiological signals are closer to the origin of the affective states (activation of the sympathetic branch of the autonomic nervous system), the signals are less dependent on observable behavior, and less susceptible to deception. However, they require more intrusive techniques and sensors for measurement, and present difficulties while recording in real-world settings and interpreting them in a participant-independent manner [9],[10].

1.4. Electroencephalography (EEG) signal

One of the most used neurophysiological signals for emotion recognition is electroencephalography (EEG), which measure the oscillations of local field potentials of neuronal masses detected at the scalp using electrodes. The neurophysiological measurement of emotions using EEG is complicated due of the low spatial resolution and the number electrodes to be placed on the participant’s head (around 8 to 128 depending on the experiment and the robustness of the equipment used). Furthermore, since most of the core affective structures are located in the ventral part of the brain, a direct assessment of the emotional activity by EEG, which primarily records signals from superficial neocortical regions, is difficult [9]. However, EEG provides great time resolution, allowing researchers to study phase changes in response to emotional stimuli and how it relates with other cognitive tasks [11]. A few methods for analyzing EEG data that correlates with emotion are: event-related potential (ERP), event related desynchronization/synchronization, steady-state visual evoked potentials, frequency-domain analysis, and frontal EEG asymmetry.

Chapter 2

Backgrounds and Aims

This chapter will provide the detail about previous work and assumption that are used in this thesis. This chapter talks about emotion recognition, EEG signals, individual characteristics, affective models, emotional stimuli tools, limitations and the aim of this research.

2.1. Emotion recognition

Emotion is a psychological state that is accompanied by physiological changes that can lead to the modification of a person's expressions, which are observable and measurable manifestations and can be perceived and evaluated by others as evidence of an emotional state [12]. For the identification of emotional states by HCI systems, varied approaches are grouped under the term emotion recognition, which uses affective models, and measurement methods to identify individuals' behavioral states, related to emotion or affective states. Emotion recognition can be examined by pattern extraction through machine learning techniques from signals like speech, body movement, and facial expressions, or physiological signals that describe individuals' behavior [13]. Figure 1 shows the basic pipeline for emotion recognition systems where the basic blocks are: input information (e.g., EEG, individual characteristics), affective model, emotional stimuli, feature extraction/feature selection, and classification/regression.

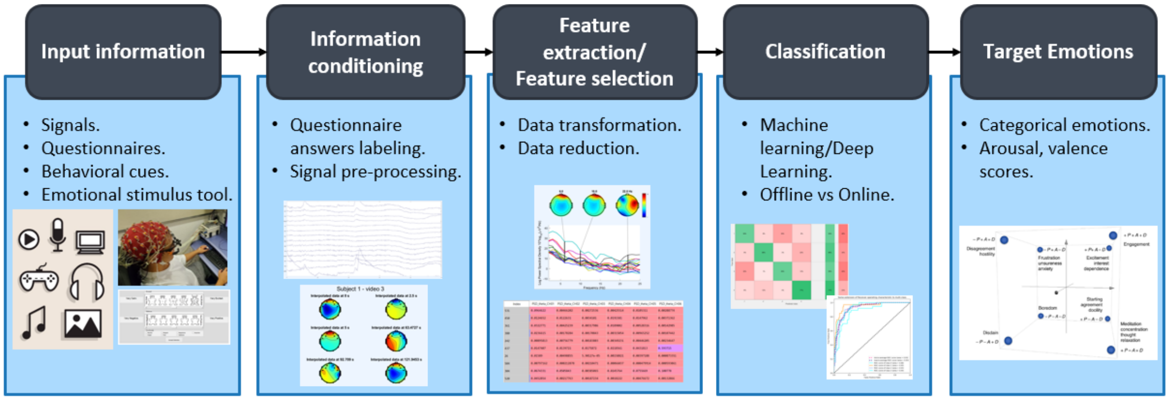


Figure 1. Emotion recognition pipeline.

2.2. EEG signals and emotion recognition

The major difficulty with physiological signals approaches relays on the effective acquisition of these signals in real world and in unsupervised scenarios due to the lack of robust equipment designed to use outside laboratory; however, one of the advantages of using physiological signals for emotion recognition is the difficulty to disguise an individual's affective states, this is the main reason why they are becoming so widely use in affective computing solutions. The limbic system is known for controlling basic motivations, including emotions, also, affect-related processing in the human brain is distributed across the brainstem, limbic, paralimbic, and neocortical regions [2].

EEG has gained increasing attention owing to its promise of potential applications in brain–computer interface (BCIs) for assistive technological solutions to overcome physical and speech disabilities. Emotion recognition using EEG signals focuses its development on two main application fields: first, medical applications designed to provide assistance, enhancement, monitoring, assessment, and diagnosis of human psychiatric and neurological diseases; and, second, non-medical applications designed to entertain, educate, and monitor emotional states in a commercial or personal context [14], [15].

EEG records the oscillations of local field potentials of neuronal masses detected at the scalp using electrodes. The EEG activity recorded at the scalp surface consists mainly of the summed postsynaptic potentials of many neurons (called a neuronal mass) aligned in the same direction and firing synchronously.

EEG signals are a powerful method for studying the brain's responses to emotional stimuli because its measurement equipment is noninvasive, fast, and inexpensive. EEG data lacks spatial resolution and requires several electrodes (around 8 to 128 depending on the experiment and the robustness of the equipment used) to be placed on the participant's head; however, it provides great time resolution, allowing researchers to study phase changes in response to emotional stimuli [11].Generally, EEG is measure using the International 10–20 system, that is an internationally recognized method to describe and apply the location of scalp electrodes (figure 2).

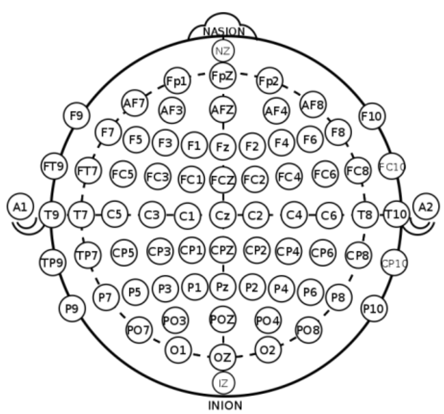


Figure 2. International 10-20 system for EEG.

2.2.1. EEG technique analysis and features

Features are characteristics of the signal that describe the behavior according to different analysis domains, on this case, 2 types of features that have proved to be related with emotions were considered [16]:

- Time domain features: are statistical parameters of the physiological signal time series, over a relatively a long-time window.
- Frequency domain features: considers a frequency spectrum and different frequency bands related to signal activation produce by a specific stimulus.

a. Time domain features:

- Picard parameters [17], [18]: mean, standard deviations of the physiological signal, max/min ratio of the EEG signals.

- Higher order statistics [19]: skewness measures the degree of asymmetry of a distribution around the signal's mean. Kurtosis is the measure of relative heaviness of the tail of a distribution with respect to the normal distribution.

- Hjorth variables [20] [21]: activity represents the signal power by the variance of a time function. Mobility represents the mean frequency or the proportion of standard deviation of the power spectrum. Complexity represents the change in frequency comparing the signal's similarity to a pure sine wave, the value converges to 1 if the signals are similar.

- Event-related potential (ERP): ERPs are signal patterns of the recorded EEG trace in response to a specific stimulus event. In the ERP methodology, a subject is repeatedly presented with a stimuli and stimulus presentation times are flagged in the raw EEG data and segmented to be averaged together. After averaging, a waveform is left that has positive and negative peaks and troughs of varying latencies called components. Each component in the waveform is either positive or negative and can be labeled according to polarity, order, or latency in ms [2].

b. Frequency domain features:

Electrical oscillations detected on the human scalp using EEG have a frequency range that has been divided into frequency bands: delta (0.05–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (12–30 Hz), and gamma (>30 Hz).

- Power spectral densities (PSD): can be obtained by Welch's method, over a specific time window, and averaged over a frequency bands range.

- Power spectral asymmetry (PSA): is calculated with the PSD ratio between asymmetric pairs of electrodes in the frequency bands correlated with emotion. These pair of electrodes comprised two electrodes located in the same scalp region, but on the opposite side of the head.

- Differential entropy (DE): can be defined as the entropy of continuous random variables and is used to measure its complexity. DE is equivalent to the logarithm of the energy spectrum (ES) in a certain frequency band for a fixed length EEG sequence [22] [23].

- Differential asymmetry (DASM) and rational asymmetry (RASM): are calculated as the differences and ratios between the DE of the asymmetric pairs of electrodes [24].

2.3. Individual characteristics and emotion recognition

In any study, variation exists between individuals and insight into these differences can help explain and predict behavior. Almost every brain is composed of largely the same components, but different features have been modulated to varying degrees to produce individual differences, for example, emotional reaction like anxiety and depression can vary from person to person [2]. Individual characteristics are person-specific attributes such as demographic variables, psychological factors, and physiological and behavioral responses, that allow differences between individuals to be distinguished.

Demographic characteristics describe an individual, based on inherit attributes that are categorized in a cultural level. A person inherits from the culture a set of lexically coded categories into which emotions are divided, and rules on how emotions are express and under what circumstances [2]. Examples of demographic characteristics include age [25], sex[26], ethnicity, and education [27].

Psychological factors are represented with personality traits, defined as continuous and consistent characteristic responses when a person acts under different circumstances. The Big Five model [28] is used to describe those individual responses using 5 traits:

- Neuroticism: The degree of stimulus needed to arouse negative emotions in an individual, and reflected by emotional stability.
- Extraversion: The degree to which a person is comfortable with relationships with others, and is reflected by degree of social and interpersonal interactions, need for stimulation, self-confidence, and amount and degree of competitiveness.
- Openness to experience: The degree to which an individual can accept unfamiliar and novel things, and is reflected in active searching for new experiences in response to certain causes.
- Agreeableness: The ability to get along, communicate, and cooperate with others, and is reflected by kindness to others.
- Conscientiousness: The degree to which a person focuses on goals and displays concentration; it is reflected in organizational ability, self-restraint, persistence, and goal-oriented behavior.

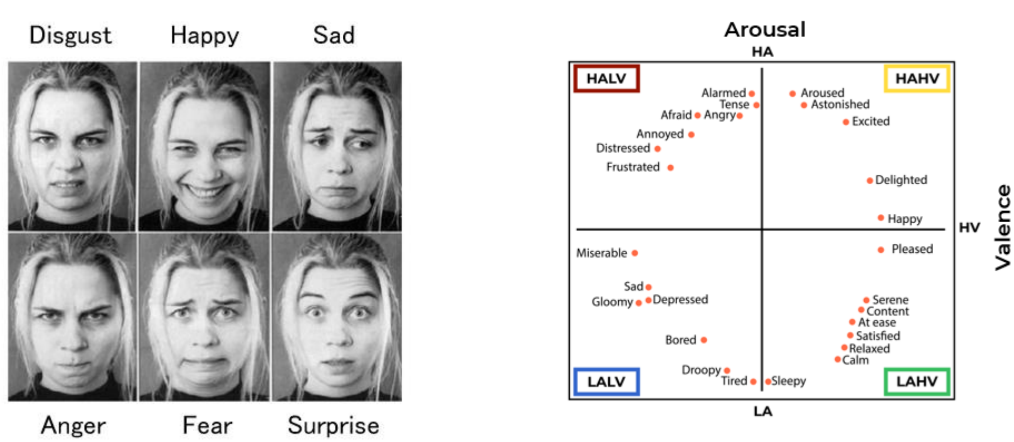
And, finally Behavioral characteristics are based on behavior of the person within a specific context, it describes physiological changes due to an external trigger, or behavioral patterns under a specific situation, for example: voice, gaits, and keystroke dynamics [29].

2.4. Emotional labels and affective models

Affective models define the values and the emotional labels to be classified, using two approaches: the discrete model and the continuous dimensional model.

The discrete model (figure 03. a.) assumes that emotions are discrete values with only a finite number of possible values. The model focuses on strong basic emotions (such as disgust, sadness, happiness, fear, anger, and surprise) and it cannot accommodate a variety of closely related emotions or combinations of emotions [30].

The continuous dimensional model (figure 03. b.) defines emotions as continuous values in a two-dimensional space. Russell circumplex model of emotion is a two-dimensional model that links arousal with valence. Arousal measures intensity, or how energized an individual feels, ranging from calmness to excitement. Valence measures how pleasant or unpleasant an individual feels, ranging from positive to negative [31].



a. b.

Figure 3. Emotional labels and affective models [32].

To measure arousal and valence scores in emotion recognition experiments, questionnaires are used to assess the emotional values from the participants. The self-assessment manikins (SAM) [33] uses a scale from 1 to 9 to assign an arousal and valence score, where 1 is "very calm" and 9 is "very excited" for arousal, and 1 is "very negative" and 9 is "very positive" for valence.

2.5. Emotional stimuli

In traditional EEG emotion recognition, researchers often need to induce users' emotional states using sounds, pictures and movies. These tools are well-known and used in different experiments and datasets related to the study of human emotions, for example: the International Affective Picture System (IAPS) [34], which is a series of standardized, emotionally-evocative photographs, and the International Affective Digital Sounds (IADS) database that uses sounds as acoustic emotional stimuli [35].

Video approach are used by showing clips, for example, clips from specific movie scenes or clips from music videos. Some well-known emotion recognition databases using videos as emotional stimuli and physiological signals, including EEG, are shown in table I. From the table is important to notice the different variables used in each research. The first one is the number of subjects, second is the number of EEG channels, and the third one is the emotion labels used to classify affective state in the participants in this case, only one work used four discrete basic emotion labels (happy, sad, fear and neutral) [36], against the arousal – valence labels used in the other works.

TABLE I. EMOTION RECOGNITION WORKS USING EEG AND VIDEOS AS EMOTIONAL STIMULI [37]

DATASET	Participants	EEG channels	Type of video	Video length	Elicit emotions
MAHNOB-HCI [38]	27	32	Commercially produced movie clips	34.9 - 117 s	Arousal/ Valence
DEAP [39]	32	32	Music videos	60 s	Arousal/ Valence
SEED [36]	15	6	Chinese movie clips	120 s	Discrete emotions
ASCERTAIN [40]	58	8	Movie clip	51-127 s	Arousal/ Valence
AMIGOS [24]	40	14	Music videos	~250 s	Arousal/ Valence

However, there are some authors that assure that these tools and approaches evoked emotions in an indirect and passive way [41]–[43]. The ability to dynamically detect users' emotional states is crucial to design adaptive emotion recognition system. One potential field of study are videogames as emotional elicit tool. Videogames are as digital interactive tools with the ability of elicit emotions in players, using resources as storytelling, game mechanics, aesthetics and digital media (as music, sounds, pictures, videos). Videogame play is cognitively demanding, and emotionally arousing and engaging, eliciting both positive and negative emotions on the players [44]. With this potential, some emotion recognition works have used videogames as an emotional stimulus, mainly under game development scenarios to target specific emotions as boredom, stress, fear or engagement [42], [43], [45]–[47]. The study of EEG signal on gaming scenarios can give an approximation on how emotions manifest on an HCI scenario level, not only for game related events, but also for cognition, training, and BCI applications [48]–[52]. In table II, some works where videogames are used as an emotional stimulus in emotion recognition experiments, along with EEG signals to identify and assess participants' emotional reaction are summarized.

TABLE II. EMOTION RECOGNITION WORKS USING EEG AND VIDEOS AS EMOTIONAL STIMULI [53]

Article	Participants	EEG Channels	Video Game and Measured Emotions	Game Play time length	Elicit emotions
[54]	28	14	Train Sim World, Unravel, Slender—The Arrival, and Goat Simulator.	5 min	Arousal/Valence
[55]	38	9	Four architectural environments designed based on Kazuyo Sejima's "Villa in the Forest" modifying illumination, color, and geometry.	1.5 min	Arousal/Valence
[56]	35	24	Candy Crush and Stickman Archers.	10 min	Discrete emotions
[57]	14	19	Tetris: medium, easy, and hard levels	5 min	Arousal/Valence

2.6. Limitations of the overall emotion recognition pipeline

The general pipeline of emotion recognition needs 2 basic sources of information to be able to classify emotions from the input data: the data describing the behavioral reactions related to the emotional experience from the participant, and the participant answers related to what they felt during the exposure to the emotional stimuli. In previous works, the experimental time window for emotion recognition goes from 35 seconds to 10 min approximately, all the works asked the participants to report how they felt after the experimental time window was over, even though, the questionnaires used to assess the emotional experience are presented to the participant immediately after the exposure to the emotional stimuli, long time windows do not allow to identify specific moments where the emotion takes place, due to the lack of labeling, this do not allow the study of dynamic changes of emotion inside the experimental time window. However, asking the participants their emotional assessment during the exposure to the stimuli can disrupt their perception against the stimuli and deliver wrong and not accurate assessment from the participants.

Also, to predict more accurate emotional states it is important to identify which kind of input information is more relevant for the whole emotion recognition pipeline, and more specially, under HCI scenarios. Works like [24], [58], [59], [60], [61], [62] have focused on identify if individual characteristics are useful to increase the emotion recognition accuracy. Also, a numerous amount of works has reported to use different EEG characteristics to improve the emotion recognition process [16]. However, it is still unclear what characteristics perform better under HCI scenarios due to the bast amount of information that can be use and the interactive nature of the experimental task.

2.7. Aims

This study aims to propose an emotional elicit tool that allows to study the emotional reactions under HCI scenarios, both inside and after the experimental time window where the participant is expose to the emotional stimuli, using EEG and individual characteristics as input data.

The main goal is to explore:

a. Which input information is suitable to increase the accuracy of emotion recognition compare with an existing public dataset, from EEG and individual characteristics.
b. How to build an emotional elicit tool that allows the study of emotional reaction within and after the exposure time window.
c. Which EEG characteristics are correlated with self-assessment emotional responses, and emotional stimuli time events.
d. The performance/regression classification of self-assessment emotional responses and emotional stimuli time events using EEG and individual characteristics information.

Chapter 3

Arousal and valence labels classification, using different EEG features, alongside age, sex, and personality traits

3.1. Aim of this section

The main objective of this chapter is to identify which EEG features and individual characteristics (age, sex, and personality traits) improve the performance of arousal and valence labels classification. The chapter explain the process and the results of testing the hypothesis that age, sex, and personality traits, can improve the classification accuracies for arousal and valence labels, when they are used alongside EEG data for emotion recognition processes by machine learning algorithms, using a public emotion recognition dataset.

3.2. Methodology

The methodology used to test the hypothesis focus on:

- Select the dataset AMIGOS and identify the information related to participants, experiment protocol, emotional answers, individual characteristics and EEG signals.

- Identify which EEG features were calculated in the original study and which other features could be calculated to analyze: their importance and the performance in the classification process of arousal and valence labels.

- Analyze the classification performance of arousal and valence labels (emotional stimuli tool labels, and participants self-assessment answers labels), using basic classification models with the selected features.

3.2.1. Dataset AMIGOS

AMIGOS is a public emotion recognition dataset to study the relationship between affect, personality, and mood [24]. The dataset consists of multimodal recordings of participants and their responses to music video clips. 40 participants (male = 27, female = 13, aged 21–40 years, mean age = 28.3 years) watched 16 videos (duration < 250 s) —four from each high and low arousal–valence emotional levels combination (figure 4. a.): high arousal and high valence (HAHV), high arousal and low valence (HALV), low arousal and high valence (LAHV), and low arousal and low valence (LALV). The experiment (figure 4. b.) consisted of an initial self-assessment session for arousal, and valence scores that participants felt before any stimuli were shown. Next, the 16 videos were presented in a random in random order with a fixed cross screen for 5 seconds at the beginning of each video, and at the end, each participant rated the video using arousal and valance scales from the SAM questionnaire (figure 4. c.). After the 16 trials, the recording session was ended.

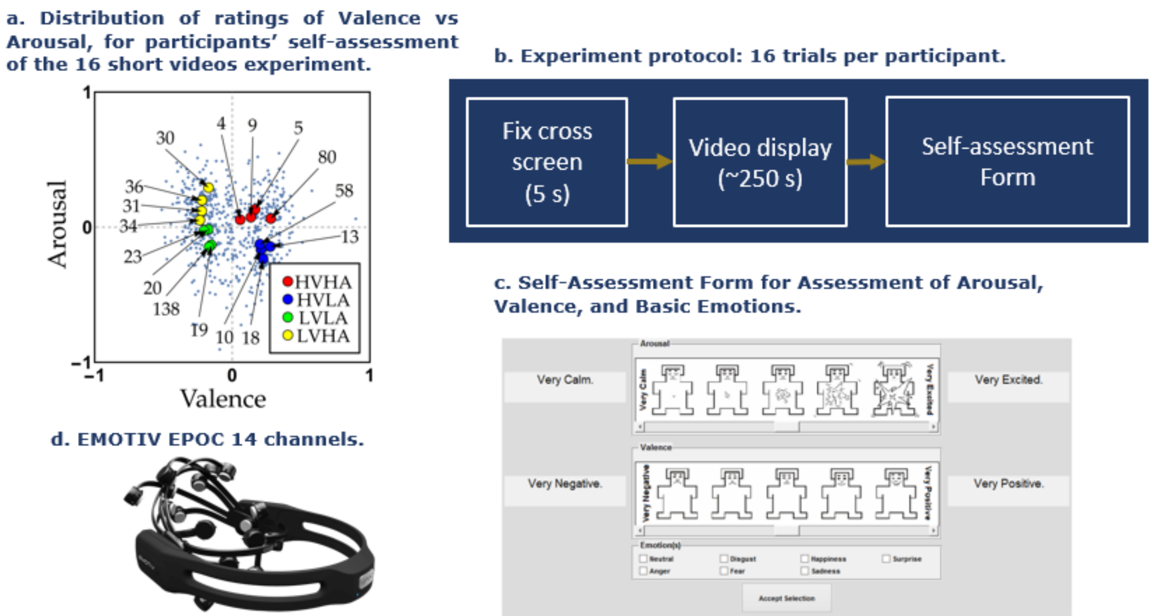


Figure 4. AMIGOS dataset characteristics

In AMIGOS dataset, 14 EEG signals from Emotiv EPOC Neuroheadset (figure 4. d.) were recorded at 128-Hz sample rate and 14-bit resolution from electrodes AF3, AF4, F3, F4, F7, F8, FC5, FC6, T7, T8, P7, P8, O1, and O2.

3.2.2. EEG features and individual characteristics data

From AMIGOS dataset, 112 features were used in this study: demographic characteristics (2 features: age and sex), personality traits (5 features: neuroticism, extraversion, openness to experience, agreeableness, conscientiousness) which were acquired before the experiment using an online form, and EEG features (105) calculated from the preprocessed EEG signals recorded in the study. The EEG signals were averaged to the common reference, filtered with a band-pass frequency filter from 4.0Hz to 45 Hz, EOG removal was applied and then segmentation was performed.

The 105 EEG features correspond to PSD and PSA between pairs of electrodes. PSD was calculated in five frequency bands: theta (3–7 Hz), slow alpha (8–10 Hz), alpha (8–13 Hz), beta (14–29 Hz), and gamma (30–47 Hz) for each electrode (70 features). PSD was obtained by Welch's method (time window = 128 samples corresponding to 1 second) between 3 and 47 Hz and averaged over the frequency bands. PSA was calculated between each of the seven pairs of electrodes in the five frequency bands (35 features). These pair of electrodes comprised two electrodes located in the same scalp region, but on the opposite side of the head: AF3/AF4, F3/F4, F7/F8, FC5/FC6, T7/T8, P7/P8, and O1/O2.

Also, EEG features that were not include in the original study were calculated to analyze their performance in the classification process (154 features): fractal dimension (FD) and the differential entropy (DE) for each one of the EEG channels in the five frequency bands, the rational asymmetry (RASM) and differential asymmetry (DASM) for each of the seven pairs of electrodes in the five frequency bands were calculated (70 features). These EEG features are related to participants' emotional [16], [20], [63], [64].

A total of. 266 features were used in this study. Feature selection methods [63] were applied to analyze how the different features are related with the classification labels and to obtain a reduced set of features (from the total 266 features), to analyze the improvement in the classification performance. Univariate selection and a recursive feature elimination with cross validation were applied to select the features that improve the classification rates and to build a second set of features [65].

3.2.3. Classifiers

The classification process was performed for 2 different scenarios: first, the arousal (HA and LA) and valence (HV and LV) labels assigned to the emotional stimuli tools, and second the arousal (HA and LA) and valence (HV and LV) labels obtained by the participants self-assessment responses. To transform the arousal and valence responses into classification labels, we use a threshold of 5.0 to convert the response values into binary labels to obtain categorical data. Both scenarios were analyzed with 2 set of input features: first the complete EEG, demographic characteristics, and personality traits features set (266), and EEG, demographic characteristics, and personality traits reduced feature set.

The classifiers were chosen to test and compare the emotion recognition accuracy using simple machine learning models and 640 observations (16 videos × 40 participants), observations that had missing personality and EEG data were excluded. The classification was performed using PANDAS framework under python language. For the classification scenarios 5 classifiers were implemented: SVM with linear (C = 100) and RBF kernel (C = 100, gamma = 0.1), Naïve Bayes, Random Forest (estimators = 2000, max_depth = 300), and an artificial neural network (ANN) with 134 hidden units, one “relu” activation function hidden layer, and, for the output layer we used a “sigmoid” activation function (optimizer = “rmsprop”, batch size = 32, epochs = 100). Parameters were tuned using grid search with cross-validation. To evaluate the classifier accuracy, the mean accuracy, mean F1, and, mean area under the curve (AUC) scores were obtained using a 10-fold cross-validation approach over the training set of features (75% of all the dataset) [65].

3.3. Results

3.3.1. Feature selection for arousal – valence labels

Using the univariate feature selection algorithm propose by [66], we obtained the best features based on an analysis of variance, F-test, and p-value of the features related to the two arousal and valence labels scenarios, selecting 10% of significant features [67]. For the arousal (HA and LA) and valence (HV and LV) labels assigned to the emotional stimuli tools no features were selected by the univariate selection algorithms. Performing RFE with personal and EEG traits, 15 features for arousal label, and 1 feature for valence were selected. In this case, no demographic characteristics nor personality traits were selected by the algorithm. For arousal label, PSD from slow alpha (AF3), alpha (AF3), and gamma (FC5) bands; F3/F4 and F7/F8 PSA index in the theta band; FD of channel P8; DE in the theta

(AF3, O1) and gamma (T8) bands; RASM in theta (T7/T8), slow alpha (F3/F4, O1/O2), beta (FC5/FC6), and gamma (F7/F8) bands were selected by the RFE. For valence label only, DE from F4 channel in the beta band was selected.

For the labels obtained by the participants self-assessment responses, the features selected by the univariate selection algorithms were: openness, PSD in the theta (O2, P8), slow alpha (O2, T8), and alpha (O2, T8); PSA index for FC5/FC6 and T7/T8 in the theta, slow alpha, and alpha bands, for O1/O2 in the beta band, and for P7/P8 and O1/O2 in the gamma band; DE in theta (O2) and gamma (CH14); and DASM in the theta, slow alpha, and alpha bands for FC5/FC6, in beta for O1/O2, and in gamma band for P7/P8 and O1/O2. For valence label, important EEG features selected were: DE for AF3 and F7 in the theta band. Performing RFE with personal and EEG traits we obtained 16 features for arousal label: PSD in slow alpha (AF3, T8) and gamma band (FC6), PSA index in the theta (FC5/FC6), alpha (T7/T8), and gamma (FC5/FC6) bands; and DE in the theta (F3, T7, O1, O2, F4), slow alpha (P8, AF4), alpha (T7), beta (FC6), and gamma (AF3) bands. Also, We obtained 40 features for valence label: PSD in theta (P7, T8, AF4), slow alpha (AF3, T8), alpha (O1, T8), beta (T8, FC6), and gamma (T8, F8, AF4) bands, PSA index in the theta (F7/F8), slow alpha (AF3/AF4, F7/F8, T7/T8, O1/O2), alpha (FC5/FC6), and beta (F7/F8, O1/O2) band; FD in FC5, T7, O2 channels; DE in theta (F7, F3, F4), beta (F3, FC5, P8, F4, AF4), and gamma (AF3) bands; DASM for theta (AF3/AF4), alpha (P7/P8), and beta (F7/F8) bands; and RASM for beta (AF3/AF4, P7/P8), theta (AF3/AF4), slow alpha (O1/O2), and alpha (F7/F8) bands [65].

3.3.2. Classification of arousal and valence labels

We tested the different machine learning classification models with a 10-fold cross-validation for the two features sets. For the arousal (HA and LA) and valence (HV and LV) labels assigned to the emotional stimuli tools, the best classifiers were SVM with linear kernel (accuracy 0.52, F1 0.49, AUC 0.51); and ANN (accuracy 0.51, F1 0.67, AUC 0.56) respectively (table 3. a.). We used receiver operating characteristic (ROC) curves to describe the performance of the best classifiers obtained from each scenario. In Figure 5. a., shows the 10-fold cross-validation ROC curves for the arousal label scenario and the valence label scenario with the best accuracies scores. The curves show the best classification performance for arousal label was obtained using EEG traits with feature reduction and SVM with linear kernel classifier (0.52 accuracy score when AUC score is higher than chance). For the valence scenario, the second set of features and the ANN classifier had the best accuracy; in this case, the curve shows that the classification process was slightly higher than chance [65].

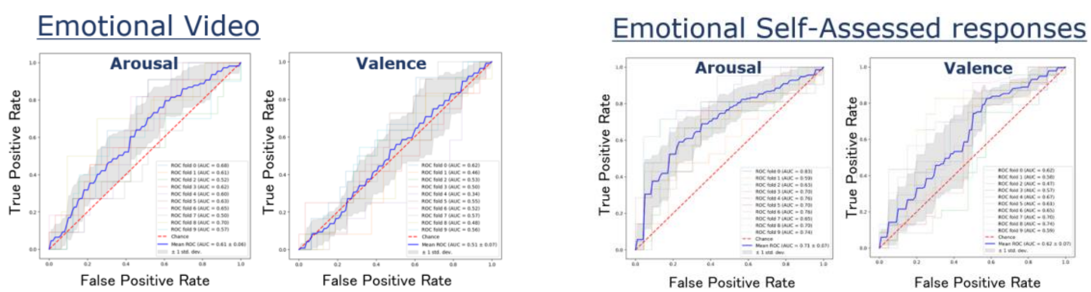


Figure 5. Best receiver operating characteristic curves with 10-fold cross-validation for arousal and valence labels.

TABLE III. CLASSIFIERS PERFORMANCE FOR EACH OF THE SCENARIOS WITH THE DEFINED SET OF TRAITS ACCURACIES, F1, AND AREA UNDER THE CURVE (AUC) SCORES

a.

Scenario	**Classifiers Label**	**Dataset reduction (selected features)**		
		Mean accuracy	**Mean F1**	**Mean AUC**
Arousal	SVM linear	0.52	0.49	0.51
Valence	ANN	0.51	0.67	0.56

b.

Scenario	**Classifiers**	**AMIGOS**	**Full dataset**			**Dataset reduction (selected features)**		
		F1	**Mean accuracy**	**Mean F1**	**Mean AUC**	**Mean accuracy**	**Mean F1**	**Mean AUC**
Arousal	SVM linear	0.592	0.63	0.60	0.66	0.62	0.58	0.65
	SVM RBF		0.68	0.67	0.71	0.64	0.63	0.67
Valence	SVM linear	0.576	0.53	0.56	0.47	0.61	0.65	0.62

Table 3. b. shows the results for the arousal (HA and LA) and valence (HV and LV) labels obtained by the participants self-assessment responses. For the arousal scenario, the first set of features (EEG data, sex, age, and personality traits without reduction) performed better when SVM with RBF kernel (accuracy 0.68, F1 0.67, AUC 0.71) was used. For the valence scenario, the second set of features (EEG data, sex, age, and personality traits with reduction) performed better when SVM with linear kernel (accuracy 0.61, F1 0.65, AUC 0.62) was used. In these cases, no demographic characteristics nor personality traits were selected in the reduced set of features; i.e., the improvement in the classification accuracies was owing to the EEG traits selected. ROC curves were used to describe the performance of the best classifiers obtained. Figure 5. b. shows the 10-fold cross-validation ROC curves for arousal and valence labels. The best accuracy scores were obtained for the arousal with 0.68 and valence with 0.61. For the discrete emotions, we decided not to show the ROC curves owing to the low F1 scores obtained in each case [65].

3.4. Discussion

The results obtained in this work revealed that none of the current age, sex, and personality values had a correlation with arousal and valence labels from the emotional stimuli. However, compared with self-assessed emotional labels, some demographic characteristics and personality traits were chosen by the feature selection for arousal, this might be because the self-assessed responses relied on participants' subjective emotion assessment. If so, demographic characteristics and personality traits would correlate more with the self-assessed emotion responses than with the emotional labels from the stimuli videos. Feature selection showed only an improvement in the classification scores for the valence label; neither demographic characteristics nor personality traits were selected by the feature selection process, which shows that age, sex, and personality traits did not foster classification performance improvement for the selected labels.

It is known form previous works that sex and age can be correlated with these emotional labels and can improve emotion recognition process [58]; however, it is still unclear how personality can be used to obtain better emotion recognition models. One of the possible reasons why sex, age, and personality were not chosen by the feature selection algorithms was because the nature of the data. If we adjust the values to a categorical and binary codification, the feature selection algorithms could select these kind of features (as age was selected in the Rukavina and colleagues' work [58]). We decided to work with the continuous data owing to the real description of the population. Other possible limitation is related to the distribution of personalities in the participants, because the sample is relative small to obtain a vast distribution in the five personality traits assessed, and the reported scores are close to each other, implying that the participants exhibit the same type of personality among the group [24]. There is also a need for a behavioral metric that can identify differences between how people perceive and manifest emotions. Behavior changes and emotional reaction can vary from person to person owing to past experiences, memories, and context.

Perceived emotions may be owing to exposure to the emotional stimuli (video in this case); however, the chosen dataset did not have information about arousal–valence scores related to the video time traces. In the scope of this analysis, we did not try to trace the changes in emotional response related to the emotional stimuli over time; instead, we wanted to determine EEG data, age, sex, and personality traits performance to classify emotions compared with the AMIGOS dataset results, in which the classification was made by averaging the time window.

Chapter 4

Interactive emotion stimulus tool design and correlation between EEG signal features with arousal and valence information and stimulus tool time-events

After the analysis of arousal and valence labels classification with EEG and individual characteristic features, using a public emotion recognition dataset, some conclusions were found:

- For small number of participants, demographic and personality features do not have major contribution in the emotion classification.
- The measurement of emotional reactions cannot be done only relaying in self-assessment questionnaires applied after the experimental time window, but instead, should be complemented with the acquisition of dynamic information inside the experimental time window that describe key events about the emotional behavior of the participants.
- To obtain information about emotional behavior, an interactive emotional stimuli tool should be use, where the participants immerse themselves in the emotional task.

According to these conclusions, a videogame was designed as an emotional stimulus tool to obtain and measure emotional reactions from the participants and to study the correlation between EEG signal features with: arousal and valence information (from self-assessment participants responses), and stimulus tool's time-events reactions (as emotional behavioral information).

4.1. Aim of this section

The main objective of this chapter is to explore which EEG signal characteristics, induced by the interaction with the emotional stimulus tool, correlated with: the arousal-valence scores, and the emotional stimulus time-events. The chapter describes the videogame design to induce different emotional states, the EEG signals features calculation and correlation with: arousal-valence scores and stimulus time-events, and the regression/classification performance of arousal-valence scores and stimulus time-events using EEG signal features.

The chapter explains the process and the results of testing the hypothesis that it is useful to consider stimulus time events (related to videogame tasks or game mechanics), self-assessment answers, and EEG features, to analyze and understand how the participants reacts emotionally to the stimulus.

4.2. Methodology

The methodology used to test the hypothesis focus on:

- Design of a videogame, as an emotional stimulus tool, to induce different emotional states.
- Acquisition of EEG signals, self-assessment emotional responses, and stimulus time-event information, while interacting with the emotional stimulus tool.

- Correlation analysis of EEG signals features with: arousal-valence scores and stimulus time-events.
- Regression performance of arousal – valence scores using EEG features.
- Classification performance of stimulus time events using EEG features.

4.2.1. Emotional stimulus tool

The videogame design was based in two videogame development frameworks (the Mechanics-Dynamics-Aesthetics (MDA) framework, and the 6–11 framework), and the concept of flow in games. MDA framework [68], breaks the games into their distinct components (rules, system and fun) and its design counterparts (mechanics, dynamics and aesthetics), where aesthetics become the main element to create emotions in the participants using the game mechanics and dynamics. The 6–11 framework [69] defines six basic emotions and eleven instincts that interact with each other to elicit “fun” using game events that affects and influence the player’s experience. Finally, flow [70] refers to the player’s ability to choose and control actions inside a videogame environment. To be on flow, the game must achieve a balance between challenges and players’ abilities: if the game is too challenging and the player’s abilities are not that high, the game experience becomes anxious, in contrast, if the game is not that challenging and the player’s abilities are too high, the experience becomes boring.

A 2D space theme platform game (figure 6) was designed with the following characteristics: the participant controls a spaceship to collect positive tokens (astronauts and coins) and avoid negative tokens (asteroids and aliens) across different game levels, to achieve the best possible score in each of the game level. The spaceship (avatar) is located in the left part of the screen and, the movement is limited to the y axis only with 3 controls: moving the spaceship to the upper part of the screen with the up-arrow key, moving the spaceship to the lower part of the screen with the down-arrow key, and increase the tokens’ scroll speed with the right-arrow key. All of these functions are controlled with one hand [53].

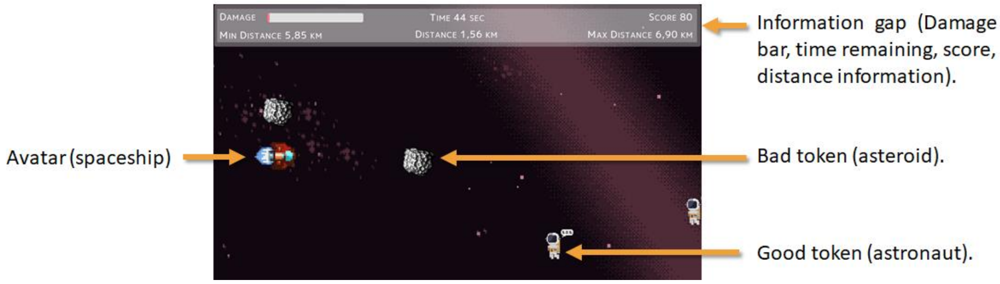


Figure 6. Game interface with the player's avatar and different tokens deployed. The information gap is located in the upper part of the screen containing information about the spaceship's damage, time remaining, the score, the distance traveled and an upper and a lower distance threshold (game mechanics)

4.2.1.1. Videogame levels

The videogame has 8 base levels that target 4 different emotional states according to Russell's circumflex model of emotion [31] (2 levels per each dimensional emotion that we are trying to elicit). The emotional states are (figure 7): frustrated (high arousal and low valence—HALV) elicit by: H—hard, OA—only asteroids; excited (high arousal and high valence—HAHV) elicit by: N—normal, SU—speed up; calm (low arousal and high valence—LAHV) elicit by: E—easy, WS—without speed; and bored (low arousal and low valence—LALV) elicit by: SD—speed down, WT—without tokens. Each game level has its own music tracks, and each token has a distinctive sound to enhance the targeted emotion according to each level [71]–[74]. Characteristics of each game level are shown in table IV.

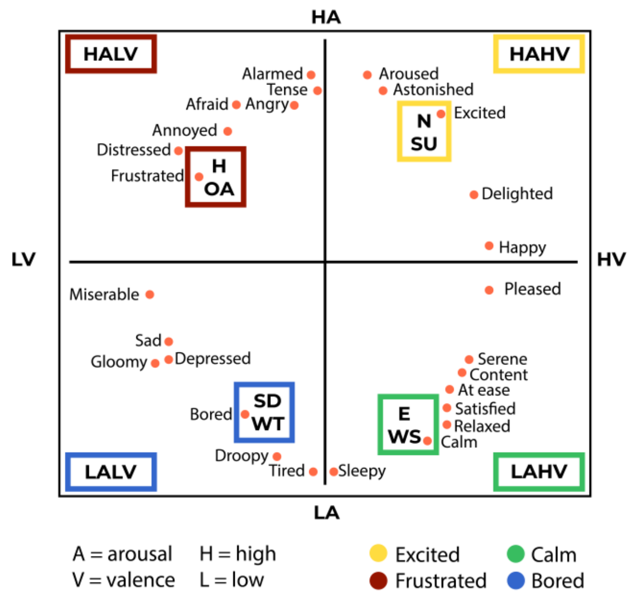


Figure 7. Russell's circumflex model of emotions for the game level designed, where the emotions selected are highlighted in the different quadrants of the two-dimensional plane.

TABLE IV. LEVEL CHARACTERISTICS

Elicited Emotion	Game scenarios' names	Controls	Bad tokens	Avatar's speed	Token's speed	Additional game mechanics
HAHV - Excitement	N SU	Normal	Normal	Normal	Normal Speed increases slow and gradually	
HALV - Frustrated	H OA	Inverted Normal	Bigger size	Decrease	Normal Speed increases fast and gradually	If collision, the size of the bad tokens increases If collision, the size of FOV decrease
LAHV - Calm	E WS	Normal	Fewer No tokens	Normal	Decrease No negative tokens	In WS Only Good tokens appear on the screen
LALV - Bored	SD WT	Normal	Fewer or none	Greater decrease	Greater decrease No tokens	In WT no tokens will appear on the screen

Each game level, is composed of: a cross fixation screen displayed for 5 seconds, the game level in which the participant interacts directly with the virtual environment for 60 seconds, the SAM arousal and valence questionnaires [33], and a game score feedback screen where a brief summary of the level performance is showed (figure 8). With this structure, the participants evaluate their emotional experience immediately after they stop playing each level, and will allow them to assess their performance without worrying about the emotional report.

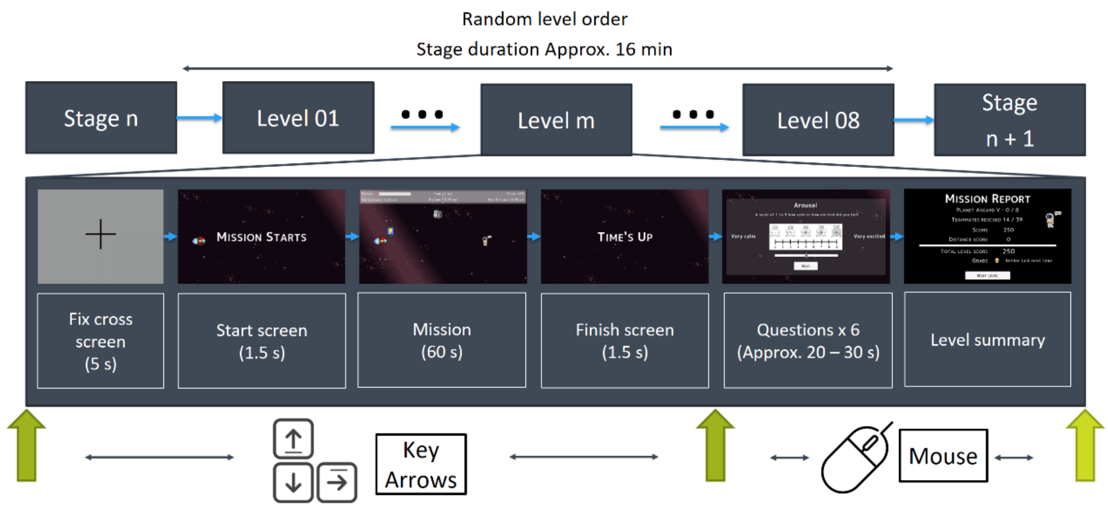


Figure 8. Game level's structure [53].

4.2.1.2. Videogame structure

Three stages were designed: Stage 1 is the basic stage that will include the basic videogame tokens (astronauts, asteroids and coins). For Stage 2 and Stage 3, the aesthetics of the background environment was changed and a distinctive new game element was added, represented as a new negative token.

The game is divided in 4 phases (figure 9), for the first, second and third phase, the stage 1, 2 and 3 were presented respectively. For the fourth and final phase, a final level was presented to the participant. This final level does not have a specific emotional label, but instead it works as the final videogame goal with two outcomes: win or lose, its structure is different from the previous levels, because it is intended to be a level that is modified according to the participant performance across the 3 first phases. In stages 1, 2, and 3, the eight game levels are presented in a random order, each stage takes approximately 16 minutes to be completed, and the whole videogame takes about 1 h.

Figure 9. Game structure composed by 4 phases containing the 3 designed stages [53].

4.2.1.3. Emotional stimulus tool evaluation

A first version of the videogame was implemented, and the emotional results were tested from a control group of participants [75]. After analyzing the result, a second version of the videogame was designed, and its performance was analyzed with two participants group: a control group that tested the first version of the videogame (17 students, females—4, males—13, ages ranging between 21 and 25, mean 22.70 ± 1.31), and a second group with participants experiencing the videogame for the first time (13 students, females—2, males—11, ages ranging between 20 and 25, mean 22.0 ± 1.78). Both participant groups performed one session of gameplay in the computational laboratories from the Pedagogical and Technological University of Colombia (figure 10), in a simultaneous scheme using personal computers and earphones. At the end of the sessions, the researchers collected a log file from each of the participant's gameplay, containing emotional questionnaires' answers (arousal and valence scores), behavioral, and performance data for each game level. All participants agreed to participate in the experiment by signing an informed consent [75].

Figure 10. Participants performing the experiment in the evaluation stage.

The arousal and valence dispersion among the 2 groups was analyzed. In figure 11, the arousal and valence scores from all the participants are represented in a two-dimensional plane. According to the arousal–valence dispersion, the participant reported different emotional states within each level, the obtained means are related to a generalized emotional experience among the participants, and they work as a reference on how each level is eliciting emotional states. The dispersion shows that it is possible to have a wide range of arousal–valence scores, and each of the designed game levels can induce the four emotional targets selected in the videogame design stage.

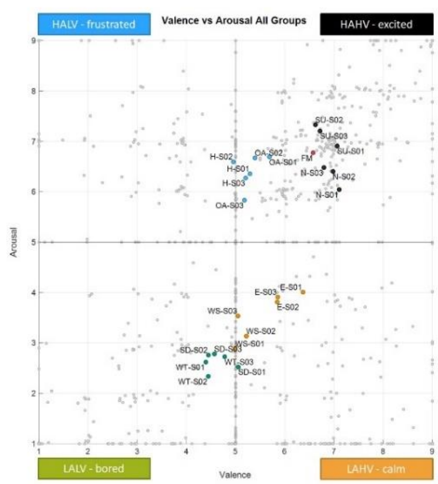


Figure 11. Arousal–valence dispersion. Levels names: N—normal, SU—speed up, H—hard, OA—only asteroids, E—easy, WS—without speed, SD—speed down, WT—without tokens. Stage names: S01—Stage 01, S02—Stage 02, S03—Stage 03. The colors are related to the emotional quadrants that we intent to induce: HAHV—black, HALV—blue, LAHV—orange, LALV—green.

For “normal” and “speed up” levels (N, SU, black dots), HAHV scores were reported throughout the three stages. It is possible to see that the mean values are similar and the majority of the distribution is located in the HAHV quadrant, leading to infer that the excited emotion was achieved over the six levels. For “hard” and “only asteroids” levels (H, OA, blue dots), the arousal scores aimed to represent HA responses. However, valence scores distribution is located between HV and LV quadrants, making the mean reference to be located close to the neutral valance value, which lead to infer that the six levels produced two kind of emotions, excited and frustrated.

For easy and without speed levels (E, WS, orange dots), LA responses were reported by the participants. For easy levels, HV scores were higher than the without speed levels’ valence scores. Without speed levels’ mean reference values are located close to the neutral valence value. In general, the easy levels’ mean reference values showed that calm emotion was induced by the six levels, and the without speed levels’ mean reference values showed that bored and calm emotion were induced in the participants. For speed down and the without tokens levels, LA responses were reported by the participants. Valence responses distribution is located between the two valence quadrants making the mean reference values to be located in the LV quadrant but close to the neutral valence value. In general, bored emotion was achieved over the six levels. Finally, for the final mission level, most of the answer distribution was located in the HAHV quadrant, leading to infer that excitement is the emotion felt by the participants while playing this last level.

A repeated analysis of variance (RANOVA) was performed, followed by a Tukey H SD (HSD—honestly significantly different) post hoc test to identify the stages with a significant difference ($p < 0.05$) within each group. After identifying the stages, a two tailed t- test was performed between the arousal and valence scores of the levels’ stages that had significant differences. The results showed that there is no significant difference between the 2 test groups concerning their arousal and valence responses between game levels.

4.2.2. Type of Acquire Information

4.2.2.1. Emotional self-assesses questionnaires

To record the emotional experience from the participants after each game level, 2 videogame screens with emotional questionnaires were designed. two variables from the self-assessment manikins (SAM) [33] were used: arousal (1 being “very calm” and 9 being “very excited”), and valence (1 being “very negative” and 9 being “very positive”). These questionnaires are manipulated using a mouse, virtual sliders, and buttons, the scales are represented with a guiding bar, numbers, and a graphic guide. The answers gave by the participants correspond to the game-level assessment after the 60 second time-window [53].

4.2.2.2. Emotional stimulus time events

Inside each videogame level, the participant will interact with 2 kinds of events: positive and negative. We defined and gathered time information of the videogame events: first positive events, which happened when the participant’s avatar collided with a token that adds points to the overall score on the game level (is an event which aims to be pleasant for the participant), examples of positive tokens are: astronauts (as an emphatic element), coins, and power ups (as an identifiable element of reward). Second, negative events, which happened when the participant’s avatar collided with a token that takes points from the overall score on the game level (is an event which aims to be unpleasant for the participant), examples of negative tokens are: asteroids and enemies (as elements of danger and damage). Inside each game the time when the event happened and the type of event was recorded, and after each game level, the total amount of positive and negative events was recorded. Each event is also presented with an audio cue related to a sound produced when the participant collides with a token.

4.2.2.3. Electroencephalography signals

The EEG signals were acquired from 66 channels (64 EEG channels and 2 earlobe references) using a Biosemi Active Two amplifier system with active sensors (Biosemi, Amsterdam, Netherlands), at a sampling rate of 2048 Hz. The EEG electrodes were positioned on the head according to the International 10–20 system. ActiView was used to monitor and setup the signals prior recording, and the LSL's Biosemi application software was used to record the signals.

4.2.3. Experiment Setup

4.2.3.1. Experiment protocol

12 participants (female 4, male 8) with ages ranging from 25 to 43 (mean 32.10 ± 5.40) from the following countries: Japan, Macedonia, Greece, Canada, Filipins, and Vietnam, participated in the study. The experiment was conducted in an individual scheme, at the Laboratory for Future Interdisciplinary Research of Science and Technology (FIRST) from Tokyo Institute of Technology. The experiment was approved by the ethics committee of the Tokyo Institute of Technology (Approval No. A20039) and conducted in accordance with the Declaration of Helsinki. The procedures were explained to each participant prior the experiment, and they were allowed to rehearse using 3 practice levels corresponding to the normal level from the first stage of the stimulus. The experiment took about 2 hours and a half to be completed for each participant, 1 hour and a half for the experimental setting and 1 hour for playing the videogame. Participants were positioned sitting in a reclining chair in a sound-attenuated chamber, and instructed to look at the monitor positioned approximately 70 cm away from their eyes during the experiment (figure 12).

Figure 12. Participant performing the experiment.

The videogame was presented to the participants using a 24” monitor, and the participants manipulated the videogame using a keyboard and a mouse. The chamber light was kept on during the experiment and participants were allowed to take breaks of unlimited time between each game level. To record the data, we used one computer with two sets of screens, keyboards and mice: the first set was used to show the stimulus and allow the participant to control the experiment task. The second set was used to monitor the signal acquisition and the participant’s performance during the task. The participants’ signals were recorded simultaneously with different software, and synchronized with LabRecorder-1.12 (Lab Stream Layer (LSL) Swartz Center for Computational Neuroscience, University of California, San Diego, CA, USA) which gathered the EEG signals and the trigger events from the videogame. After signal inspections, information from 2 participants were rejected due to the high artifact contamination .

4.2.3.2. EEG preprocessing

Lab stream layer (LSL, Swartz Center for Computational Neuroscience, University of California, San Diego, CA, USA) recorded the data’s streams that contained the actual sample data related to the signal values, and event markers that came from the videogame, together with the timestamp for each sample that is read from a local high-resolution clock of the computer. To extract the signal related to each level, we identified the start level and finish level markers events of each level, generated from the game environment, and extracted the signals portion related to these time windows.

Data were processed using MATLAB R2019b (The MathWorks, Inc., Natick, MA, USA), we reference the 64 channels' signals to the ears and applied a finite impulse response (FIR) notch filter at 50 Hz, then we applied a FIR high pass filter at 1 Hz and a FIR low pass filter at 40 Hz, then down-sampled the signal at from 512 HZ to reduce the computational cost. Then we inspected and reject the noisy channels and referenced the signals to the average of the channels. Finally, we applied independent component analysis (ICA) [76]–[78], and inspected each of the components to manually reject the ones related with noise and artifacts (eye movement, blinks and muscular activity) [16], [79], [80].

4.2.3.3. EEG signal features calculation

8 time-domain features (mean, standard deviations, ratio max/min, skewness, kurtosis, activity, mobility, complexity) and 10 frequency-domain features (power spectral density—PDS, differential entropy—DE for 5 frequency bands: theta, alpha, beta, gamma bands, and full frequency spectrum) for each EEG channel (18 × 64 = 1152) were calculated; in addition, 15 frequency-domain features (Power spectral asymmetry — PS-ASM, differential asymmetry — DASM, rational asymmetry — RASM) for each of the 27 pair of electrodes (15 × 27 = 405) were also calculated. In total, 1557 EEG features were calculated and used for correlation analysis and predictions with emotional labels. For arousal and valence scores a 60 seconds time window (related to the full gameplay of each of the 25-videogame levels) was considered, and for game time-events, a 500 ms time window (from the occurrence of all game events in each of the 24 levels from phases 1, 2, 3) was considered.

4.3. Results

4.3.1. Arousal–Valence Dispersion

In figure 13, the arousal and valence scores from all the participants are represented in a two-dimensional plane. The participants reported different emotional states within each game level, where the obtained means are related to a generalized emotional experience, they work as a reference on how each game level elicited the emotional states. In table 3, we summarized the mean and standard deviation values for each of the game levels.

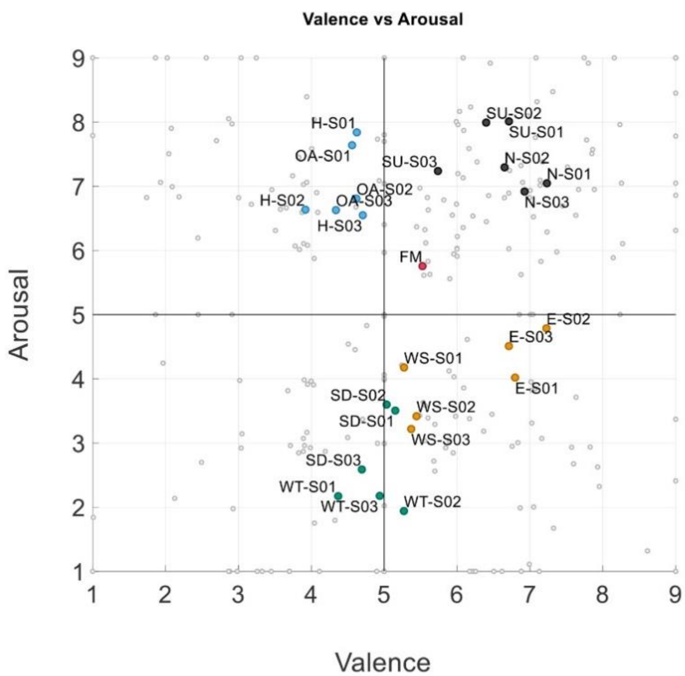


Figure 13. Arousal–valence dispersion. Levels names: N—normal, S01—Stage 01, S02—Stage 02, S03—Stage 03. The colors are related to the emotional quadrants that we intent to induce: high arousal and high valence (HAHV)—black, high arousal and low valence (HALV)—blue, low arousal and high valence (LAHV)—orange, low arousal and low valence (LALV)—green.

For normal and speed up levels (N, SU, black dots), HAHV scores were reported throughout the three stages. It is possible to see that the mean values are similar, and the majority of the distribution is located in the HAHV quadrant, leading to infer that the excited emotion was achieved over the six levels. For hard and only asteroids levels (H, OA, blue dots), the arousal scores aimed to represent HALV responses, leading to infer that the frustrated emotion was achieved over the six levels.

For easy and without speed levels (E, WS, orange dots), LAHV responses were reported by the participants. For easy levels, HV scores were higher than the without speed levels' valence scores. Without speed levels' mean reference values are located close to the neutral valence value. In general, the easy levels' mean reference values showed that calm emotion was induced by the six levels, and the without speed levels' mean reference values showed that bored and calm emotion were induced in the participants. For speed down and the without tokens levels, LA responses were reported by the participants. Valence responses distribution is located between the two valence quadrants making the mean reference values to be located in the LV quadrant but close to the neutral valence value. In general, bored emotion was achieved over the six levels. Finally, for the final mission level, most of the answer distribution was located in the HAHV quadrant, leading to infer that excitement is the emotion felt by the participants while playing this last level.

4.3.2. Game Events

The number of positive and negative events obtained by each of the participants were inspected. In figure 14, an example of different participant's performance according to the number of positive and negative events in each level is showed, along the arousal and valence responses reported. Participant 1 had a lower number of negative events across the game levels, in contrast, Participant 9 had a higher number of negative events in levels aimed to induce HALV and in the final game level. It is possible to see that the number of positive events is higher than the number of negative events, this is understandable because of the objectives proposed by the virtual environment of collecting positive tokens and avoid collision with negative tokens, however, there is still a high number of negative events that allowed the study of emotional reaction.

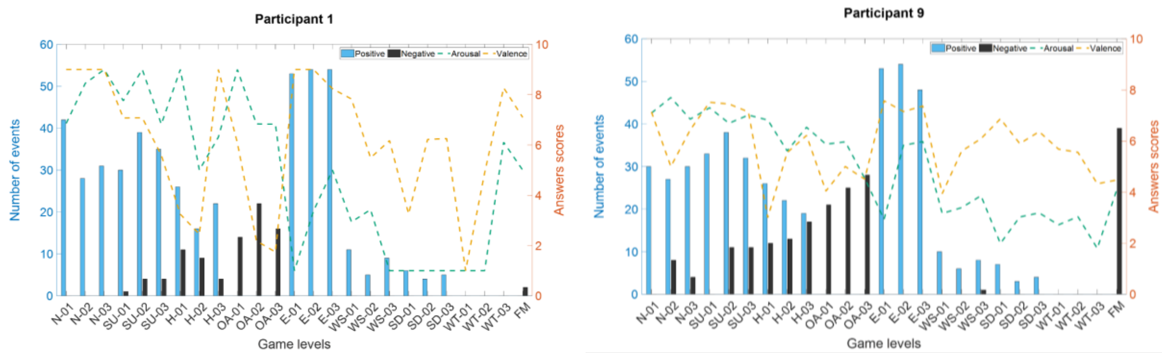


Figure 14. Total amount of positive and negative events obtained by each participant across all the game levels along the arousal and valence responses.

4.3.3. Spearman's correlation between EEG features and arousal - valence scores

To study the relation between the self-assessment responses and the EEG features, Spearman's correlation was performed, the EEG features with a strong (equal or higher than |0.5|) correlation's score and significant p-value ($p < 0.005$) are reported. The Spearman's correlation was performed between the 1557 EEG features and the arousal-valence scores given by the participants after completed each of the videogame levels (60 seconds time window).

On table V. a., the number of features correlated for each participant is showed. The lower amount of EEG features for arousal was 106 (6.8% of the total number of calculated features) and the higher amount was 287 (18.43%). For valence, the lower amount of EEG features was 7 (0.45%) and the higher amount was 323 (20.74%). On average, the number of features correlated with arousal was higher than the number of features correlated with valence (only for participant 5 the number of correlated features for valence was higher than for arousal).

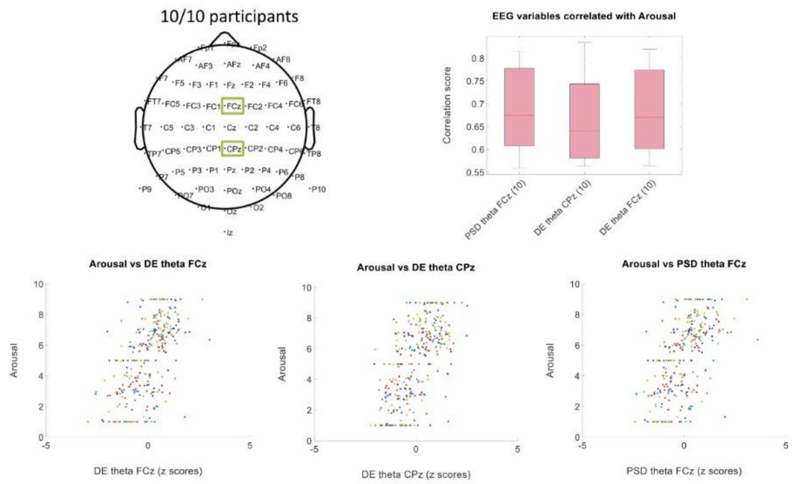
In contrast, when common correlated EEG features were considered, few common features had a correlation for the participants' majority and arousal answers, it is possible to see from table V. b., that when we consider more participants the number of common features decrease. From the 1557 EEG features calculated, only 3 were common correlated for all the 10 participants, and 461 EEG features were unique among the 10 participants (461 traits had only one occurrence when the correlated EEG features of the 10 participants were inspected). For valence answers, fewer common features were found (the higher number was 4 participants with 1 common trait), and 480 features had only one occurrence when the correlated EEG features of the 10 participants were inspected.

TABLE V. NUMBER OF ELECTROENCEPHALOGRAPHY (EEG) TRAITS CORRELATED WITH AROUSAL AND VALENCE SCORE: (A) FOR EACH PARTICIPANT, (B) TRAITS COMMON AMONG PARTICIPANTS [53]

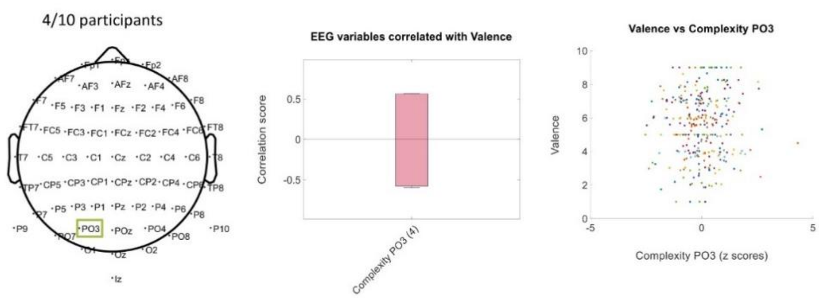
a. Individual traits correlated for each participant			
Participant	**Gender**	**Arousal Num. of features**	**Valence Num. of features**
1	Male	223	200
2	Male	245	9
3	Male	207	10
4	Female	265	8
5	Male	287	323
6	Male	272	16
9	Male	146	13
10	Female	254	60
11	Male	140	7
12	Female	106	2
Total		2145	648

b. Number of traits correlated common among participants		
Number of participants	**Arousal Num. of features**	**Valence Num. of features**
1/10	461	480
2/10	260	82
3/10	155	0
4/10	79	1
5/10	35	0
6/10	10	0
7/10	4	0
8/10	9	0
9/10	2	0
10/10	3	0
Total	2145	648

The correlation scores of the common features among participants was explored. PSD and DE traits on the theta band for channels in the frontal (F), central (C), and parietal (P) regions: FCz, CPz, Cz, FC1, FC2, C1, CP1, CP2 had positive correlation for the majority of the participants. In figure 15. a., the features correlated with arousal for all the participants and the dispersion of each trait is showed. In contrast, in figure 15. b., the only common correlated feature with valence for 4 participants is shown (complexity of PO3 channel had both positive and negative correlations and lower rho scores compare with the features correlated with arousal), from the dispersion is possible to identify that is more difficult to identify a clear pattern. There is a higher number of features that correlated with arousal scores than valence scores across all participants. Arousal scores are related with features in the theta frequency band and electrodes located in the frontal central, central, and central parietal regions.



(a)



(b)

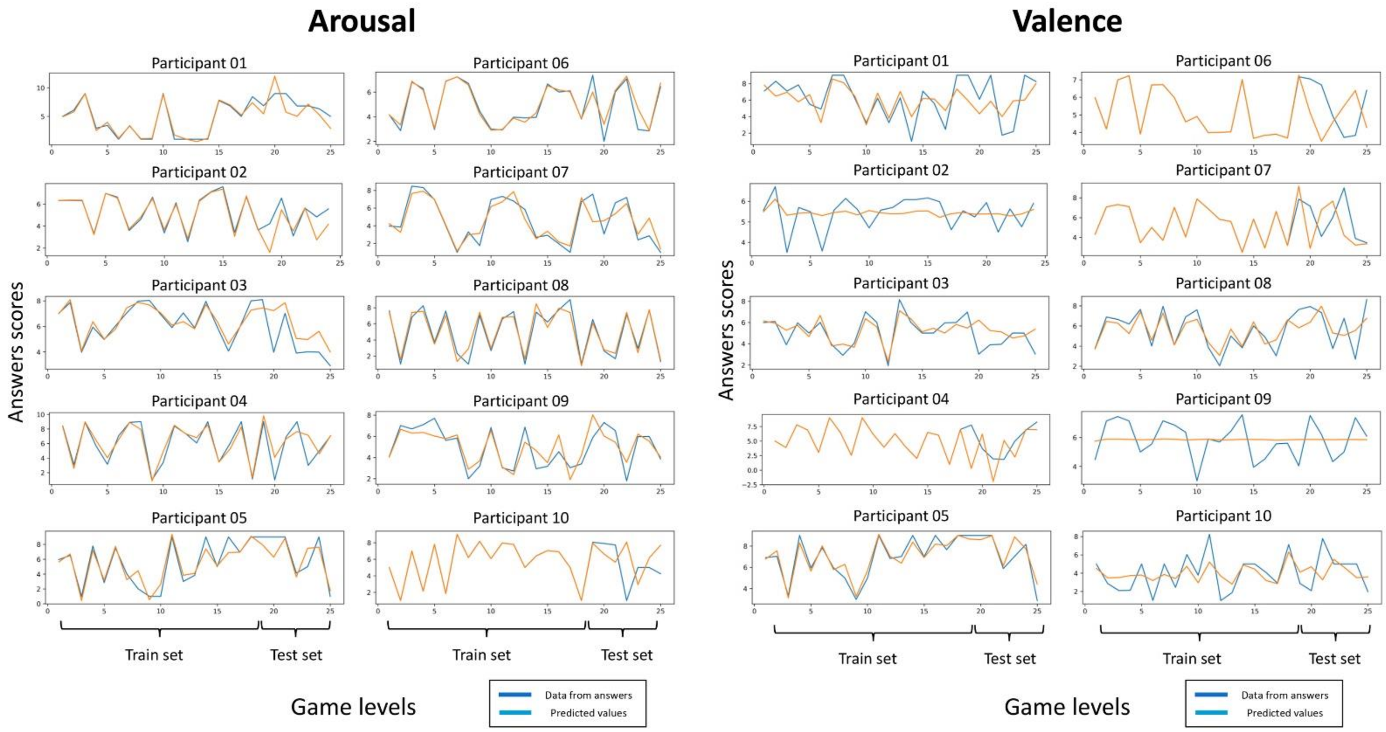
Figure 15. Spearman correlation score's dispersion of features common for the participants. (a) Features correlated with arousal scores. The reported traits have positive rho scores with a mean value above 0.6. (b) Features correlated with valence scores, for 4 participants only one trait had a strong correlation, in this case, the correlations had positive and negative rho scores among participant whit scores no higher than 0.6 [53].

The feature correlations with arousal scores vary between participants due the individual difference on behavior and reaction to the emotional content, and how each individual approached the experimental task. In addition, as can be related with the time window to extract the different traits, during to each level, different physiological activations can occur depending on the events and challenges that each game level presents. However, despite the difference between participants, it is possible to see that arousal scores can be describe by different EEG signal features. There are other features that correlated in an individual approach for both arousal and valence scores, for this reason, we decided to implement a regression algorithm to predict these scores and evaluate the performance on an individual level.

4.3.4. Arousal and Valence Prediction using Bayesian Ridge Regression Model

To identify the prediction performance of arousal and valence scores using the EEG traits, a Bayesian ridge regression model was implemented, to make predictions using 25 observations per participant and EEG features selected form the 1557 calculated. Bayesian regression techniques can be used to include regularization parameters in the estimation procedure, this is done by introducing uninformative priors over the hyper parameters of the model. For Bayesian ridge regression the loss function is augmented in such a way that not only minimize the sum of squared residuals but also penalize the size of parameter estimates [66].

The z scores of each feature to normalize the values was calculated, then, the dataset was split into train set (75%) and test set (25%), and feature selection for regression approaches was performed, using mutual information [81] and grid search with repeated cross validation (splits = 10, repeats = 3, random state = 1) over the train set, with mean absolute error scoring. Features selected for arousal and valence score are related with time domain (standard deviation, complexity, mobility, kurtosis, skewness), and PSD and DE from theta, alpha, beta, gamma, and all EEG frequency spectrum. Then, the models were trained over the train set using the features selected, with a repeated cross validation scheme (splits = 10, repeats = 3, random state = 1), and the mean absolute errors (MAE) and the mean square errors (MSE) were calculated as performance scores.



(a) (b)

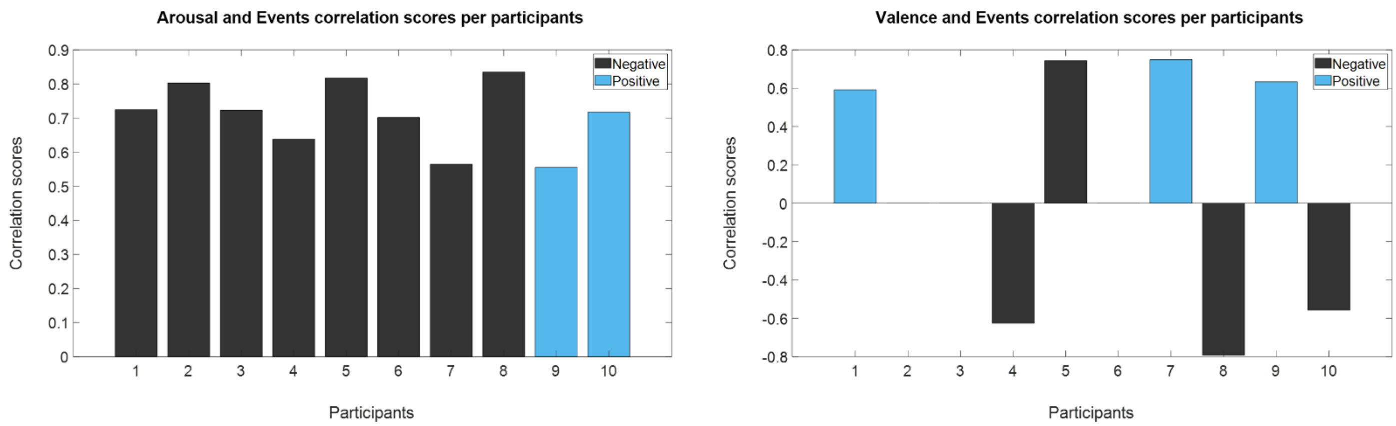
Figure 16. Performance of Bayesian ridge regression predictions for arousal and valence scores. (a) Arousal score values and predictions over the train and the test set. (b) Valence score values and predictions over the train and the test set [53].

Finally, the obtained model was tested over the test set. Figure 16. a. shows the regression result. For arousal, MAE had an average of 0.973 ± 0.316 on train set, and 1.199 ± 0.321 on test set, MSE had an average of 1.786 ± 1.122 on train set and on test set 3.186 ± 2.876, the best result was achieved for Participant 6 in the train set (MAE: 0.406 ± 0.299, MSE: 0.328 ± 0.406). For valence, the MAE had an average of 1.199 ± 0.321 on train set, and 1.670 ± 0.784 on test set, the MSE had an average of 2.504 ± 1.2112 on train set and on test set 4.680 ± 4.032, the best result was achieved for participant 5 in the train set (MAE: 0.723 ± 0.317, MSE: 0.806 ± 0.655). From the prediction models' implementation, better result for arousal scores were obtained than valence scores in both the train set and the test set (only participant 5 got better results for valence than for arousal scores in both train and test set). In figure 16, it is possible to see the regression's prediction of arousal and valence values across all videogame levels (25 game levels or observations) for each of the participants. In general, is easier to find EEG features that correlated and helps to describe the arousal answers given by the participants than valence answers.

4.3.5. Analysis from time related events and EEG features

Although, the time events were labeled as positive or negative according to the context of the videogame and the event that represented, the correlation allowed to determine if those time events have the same appraisal for each of the participants. To perform the correlation, the amount of both events (positive and negative) in each of the game levels per participants were counted, then Spearman's correlation was calculated to identify which event types correlated with the answers gave by the participants. On figure 17 is showed: (a) the correlation score between arousal and number of events, and (b) the correlation score between valence and number of events.

For arousal scores, the number of negative events had a positive correlation across the majority of the participants, only for participants 9 and 10, positive events correlated positively. For valence score, strong correlations were found for negative events (participants 4, 5, 8 and 10) and for positive events (participants 1, 7, 9). Presence of negative events can induce higher levels of arousals responses; this can be due to the nature of each of the designed game levels. With the results, it is possible to identify emotional reactions thought the identifications of game time events.



a. b.

Figure 17. Correlation between arousal–valence scores and number of events per game level. (a) Arousal correlations. (b) Valence correlations [53].

4.3.6. Spearman's correlation between EEG features and game time events

To study the relation between the game time events and the EEG traits, Spearman's correlation was performed, the EEG features with a strong (equal or higher than |0.5|) correlation's score and significant p-value ($p < 0.005$) are reported. First, the game time events that were too close from another event were excluded (if an event had another event inside the 500ms time window, that event was excluded from the analysis). For all 10 participants, the excluded events were less than 15% of the total positive events, and less than 30% of the total negative events. The final number of events used in our analysis is showed in table VI.

TABLE VI. NUMBER OF POSITIVE AND NEGATIVE EVENTS PER PARTICIPANTS [53]

	Participants									
Events	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Positive	470	428	437	409	465	478	467	445	450	320
Negative	85	121	94	130	114	91	117	114	151	171

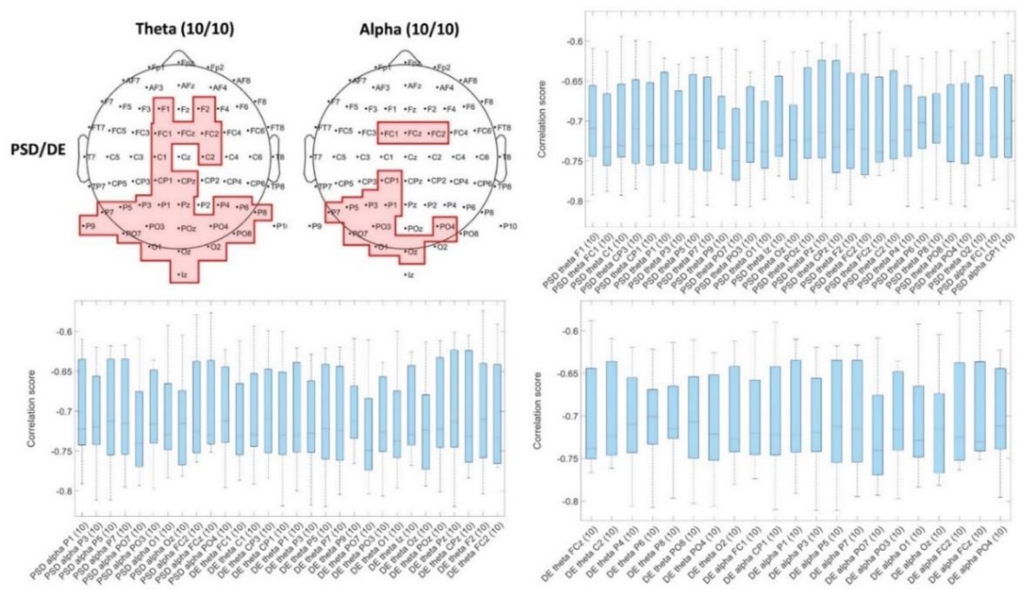


Figure 18. EEG features correlated with game events. Theta band's PSD and DE from electrodes on the occipital and central brain region and, alpha band's PSD and DE from electrodes on the frontal-central and the occipital brain regions. The EEG features had a negative correlation with positive events and positive correlation with negative events [53].

The Spearman's correlation was performed between the 1557 EEG features and the game time events inside each of the game levels. The EEG features were calculated, on a signal portion extracted from a time window of 500 ms (0 s at event onset). There is a high number of common features for all the participants, PSD and DE features of theta and alpha bands for all the EEG channels except of P2 correlated negative with positive events and positively with negative events. With lower rho scores, PSD features of beta band for all EEG channels except of: Fp1, AF7, F7, F5, FT7, FC5, Fpz, AF8, F8, and P2; and DE features of beta band for all EEG channels except of: Fp1, AF7, F7, F5, FT7, FC5, Fpz, AF8, F8, P2, Fp2, AF8, correlated negative with positive events and positively with negative events. As, example in figure 18, the traits and the topographical plots of channels with a rho scores stronger than 0.7 (rho > = |0.7|) are showed, the PSD and DE features related with the theta (channels: F1, F2, FC1, FC2, FCz, C1, C2, CP1, CPz, P1, P3, P4, P5, P6, P7, P8, P9, Pz, POz, PO3, PO4, PO7, PO8, O1, O2, Oz, Iz) and the alpha band (channels: FC1, FC2, FCz, CP1, P1, P3, P5, P7, PO3, PO7, PO4, O1, Oz) had a negative correlation with positive events (positive correlation with negative events).

4.3.7. Game events classification using ensembling methods

The ratio between the total amount of positive and negative events acquired during the full gameplay is not even, the number of positive events was higher than the number of negative events (figure 14, and table VI), because of this, it is necessary to implement a classification model for imbalance data on the classes. In ensemble classifiers, bagging methods build several estimators on different randomly selected subset of data. Ensemble methods use multiple learning algorithms to obtain better performance, bagging methods work by building multiple estimators on a different randomly selected subset of data, allowing to train the classifier that will handle the imbalance without to under-sample or oversample manually before training [82].

First, some of the observations from the majority class (undersample the positive events) are randomly deleted, to obtain the same number of observations from the minority class (negative events). Then, feature selection is performed with the under-sampled dataset using recursive feature elimination with cross validation and a support vector classification algorithm (linear kernel, regularization C = 100). The dataset is split into train set (75%) and test set (25%) with the selected features. Then, a balanced bagging classifier using decision trees as estimator, with 10-fold cross validation and the selected traits on the train set (without under-sampling dataset) is implemented.

TABLE VII. NUMBER OF POSITIVE AND NEGATIVE EVENTS PER PARTICIPANTS [53]

			Events							
			Training						Test	
Num. of Participants	Gender	N. Traits	Acc		F1		AUC		Acc	F1
			Mean	Std	Mean	Std	Mean	Std		
1	Male	21	0.97	0.05	0.98	0.03	0.99	0.01	0.91	0.94
2	Male	2 (Pz)	0.99	0.03	0.99	0.02	0.99	0.0	0.99	1.00
3	Male	2 (PO3)	0.99	0.04	0.99	0.04	0.99	0.01	0.97	0.98
4	Female	2 (Pz)	0.99	0.02	1.00	0.01	0.99	0.0	1.0	1.0
5	Male	2 (Oz)	0.99	0.03	0.99	0.02	0.99	0.0	0.98	0.98
6	Male	2 (POz)	0.99	0.02	0.99	0.01	0.99	0.01	0.98	0.99
9	Male	2 (Pz)	1.0	0.01	1.0	0.01	1.0	1.0	1.0	1.0
10	Female	2 (P8)	1.0	0.01	1.0	0.01	0.99	0.0	1.0	1.0
11	Male	2 (PO4)	0.99	0.02	1.0	0.02	0.99	0.0	0.98	0.98
12	Female	2 (P3)	0.99	0.05	0.99	0.05	0.99	0.0	1.0	1.0

In the table VII, the classifier's performance with accuracy (Acc), F1, and area under the curve (AUC) scores are showed. From the scores, it was clear the good performance of the classifiers to discriminate between positive and negative time events, the features selected from the majority of the participants were DE from the theta band and form the full EEG frequency spectrum, only for participant 1 the features selected were: DE theta (F1, C1, PO3, CPz, AFz, Fz, F2, FCz, CP2), DE alpha (F1, CPz, F2), and DE from the full EEG frequency spectrum (F1, C1, Pz, CPz, AFz, Fz, F2, FCz, CP2). The results proved that with each of the EEG signals is possible to identify emotional states in the participants using a smaller time window with game events and these events have a relation with the arousal-valence scores gave by the participants.

4.4. Discussion

Prior works have reported that higher frequency bands relate with emotions while using pictures, videos or recalling past experiences as emotional stimuli. For example, in [83], authors reported the presence of decrease of peaks in the alpha band for fear and sorrow emotions, in contrast of increase in peaks frequencies in the alpha band for joy and anger. Works using the SEED dataset [36], have reported that alpha, beta, and gamma had better emotion classification performance [84], [85]. Their findings showed that for positive emotion, beta and gamma frequency bands energy increases, and for neutral and negative emotions beta and gamma frequency bands have lower energy. In addition, neutral emotions have higher energy on alpha band [86]. Other works support these findings using pictures as emotional stimuli [87], and using the DEAP dataset [88]. EEG traits correlated with arousal for the majority of the participants shows that emotions can be described from the theta band of the central electrodes (FCz, FC1, FC2, Cz, C1, C2, CPz, CP1, CP2), however, on an individual approach, when performing regression to predict emotional values with feature selection, the selected features for each participant showed that traits from the alpha, beta and gamma band were selected in higher rate than the theta band traits. In addition, some time domain features like complexity and standard deviation of the channel's signals are selected in the majority of the participants.

Comparing with other works using videos as emotional stimuli [24], [36], [38], [39], [40], arousal and valence labels classification metrics (accuracy and F1 scores) have values around 0.5 and 0.8 using machine learning methods as support vector machines and k-nearest neighbors. Arousal label classifications scores have better results than valence label classifications. In our case, we used regression analysis to predict the arousal and valence scores. The results showed error metrics higher than 0.4 but lower than 1.2 for MAE and the MSE showed values higher than 0.3 but lower than 3.5 for arousal scores, and error metrics higher than 0.7 but lower than 2.1 for MAE and the MSE showed values higher than 0.8 but lower than 4.0 for valence scores. As previous works, classification/prediction of valence labels/scores is more difficult than for arousal. This is not only reflected on the performance metrics but also in the EEG features correlation with valence scores. This can be related to the process of rating overall positive and negative emotional experience on a task (video or videogame) that had diverse contents within the same long-time window. Depending on the content showed to the participant in a long time-window, the participant can have both positive and negative emotions on different moments inside the time window, making difficult to summarize the overall experience with one score and, at the same time, generalize the signal activation through the whole-time window, narrows the possibility of identify specific moments where each participant could feel different emotions related to high or low valence responses. Although, the results are consistent with previous works, more studies that analyze emotional reactions and EEG traits under interactive virtual environments are needed to contrast our findings using the same emotional stimuli nature.

Chapter 5

Event related potentials (ERP) activation in the presence of emotional stimuli tool time events

5.1. Aim of this section

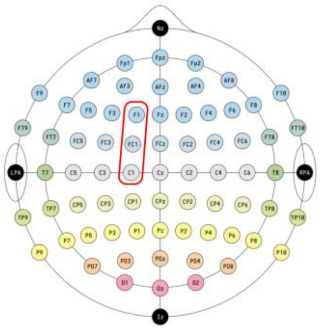
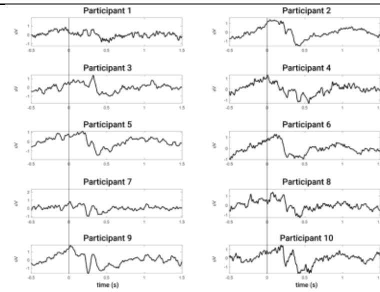
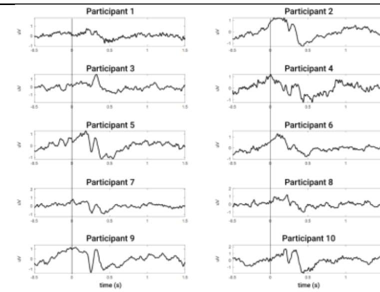
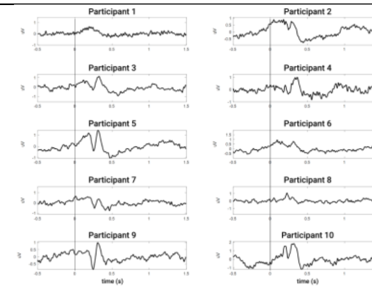
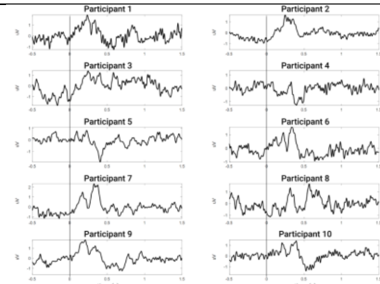
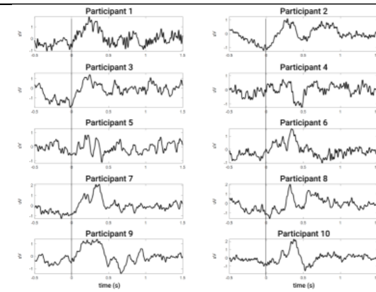
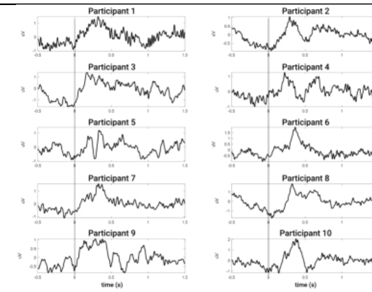
While plotting the EEG signals related with the channels that highly correlated with emotional stimulus time events, some ERP components were found. The ERPs have some common characteristics related to positive and negative events across the 10 participants of this study.

The main objective of this chapter is to describe the ERP activation related with emotional stimulus time events, that allow to describe emotional reactions using the EEG signals. The 27 channels' signals that correlated with the different events were inspected for this analysis. The channels' signals portion around the analysis time window (0.5 s from the event occurrence) was extracted, and the mean values of the channel's amplitude were calculated and plotted. Some clear ERP activations that manifested in the presence of a game time event were obtained, for some of the channels correlated with positive and negative events.

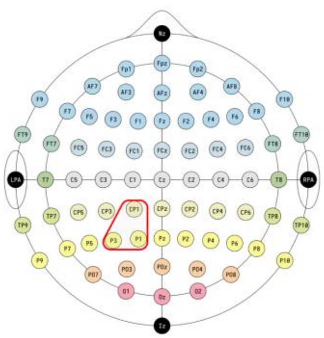
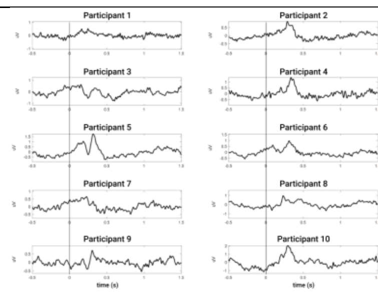
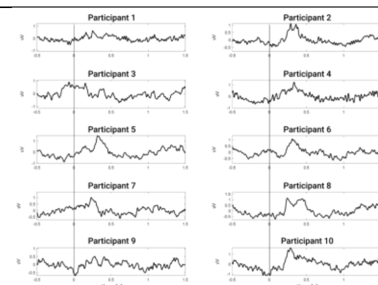
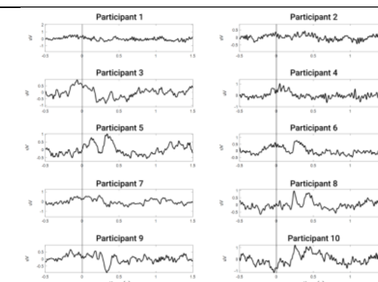
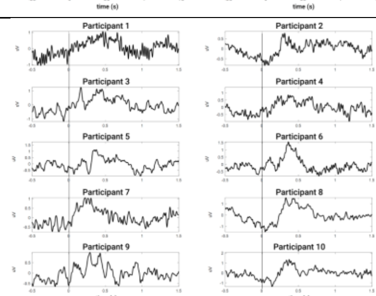
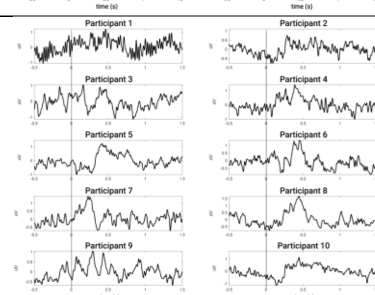
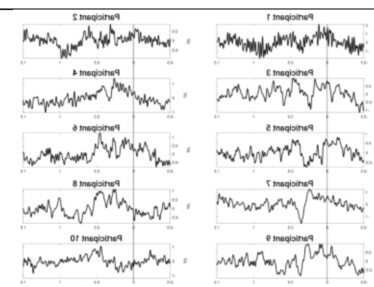
In tables VIII, IX, and X, left, central and right brain regions channels' signals are showed, respectively. To be able to distinguish the ERP behavior the channel's signals were plotted with a time window of 2.0 seconds, event onset at 0 seconds, 0.5 seconds before event onset and 1.5 seconds after event onset. Although, for all the correlated channels there are amplitude characteristic's contrast between positive and negative time events. Only the description for ERP components of channels FCz, F2, FC2, C2, P5, P7, P9, P4, P6, and P8 are described, due to the similar amplitude patterns that are found across all participants. For all channels, the signal's frequencies are higher for negative time events signals than frequencies for positive time events signals. The presence of ERP components, are evidence of participants reacting of the occurrence of game time events, the frequency characteristics showed a distinction between the nature of the events that were distinctive according to each participant.

TABLE VIII. LEFT BRAIN REGION ERPS

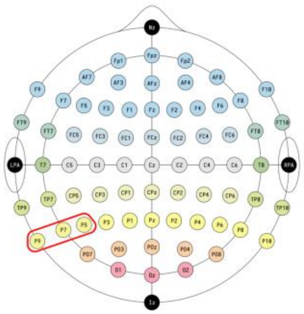
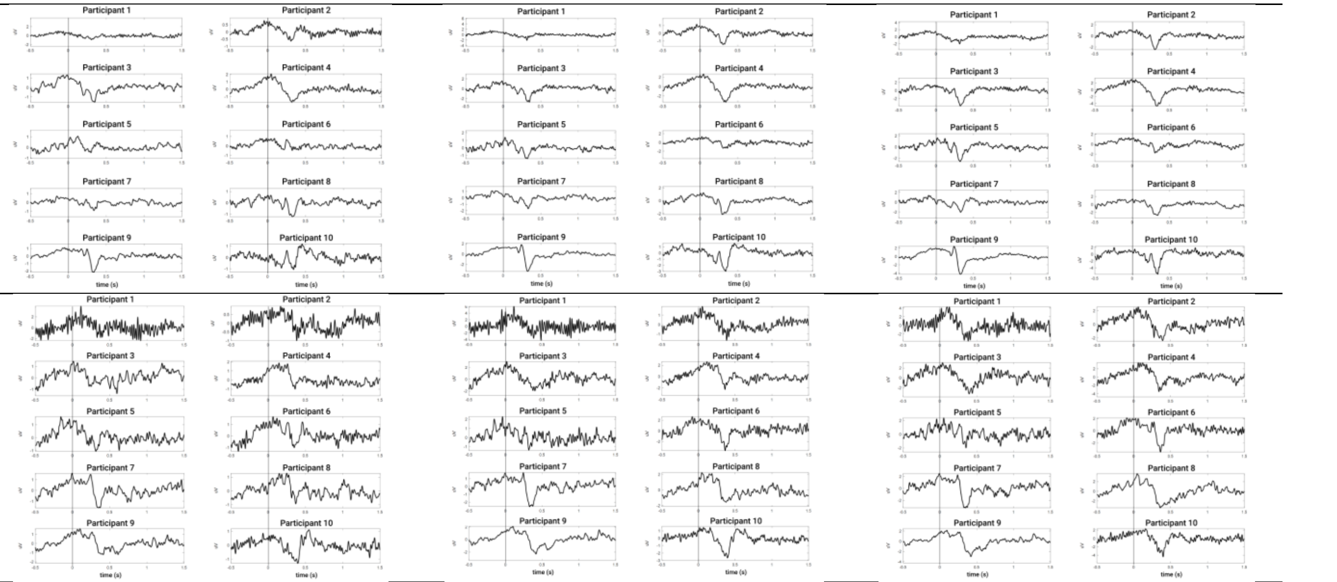
a.

		F1	FC1	C1
	**Positive time events**			
	Negative time events			

b.

		CP1	P1	P3
	**Positive time events**			
	Negative time events			

c.

	P5	P7	P9
 Positive time events			
Negative time events			

d.

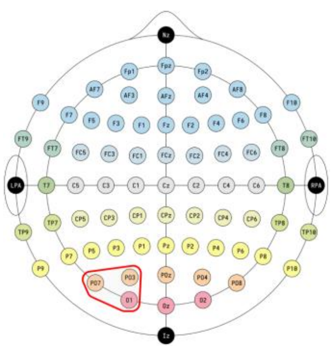
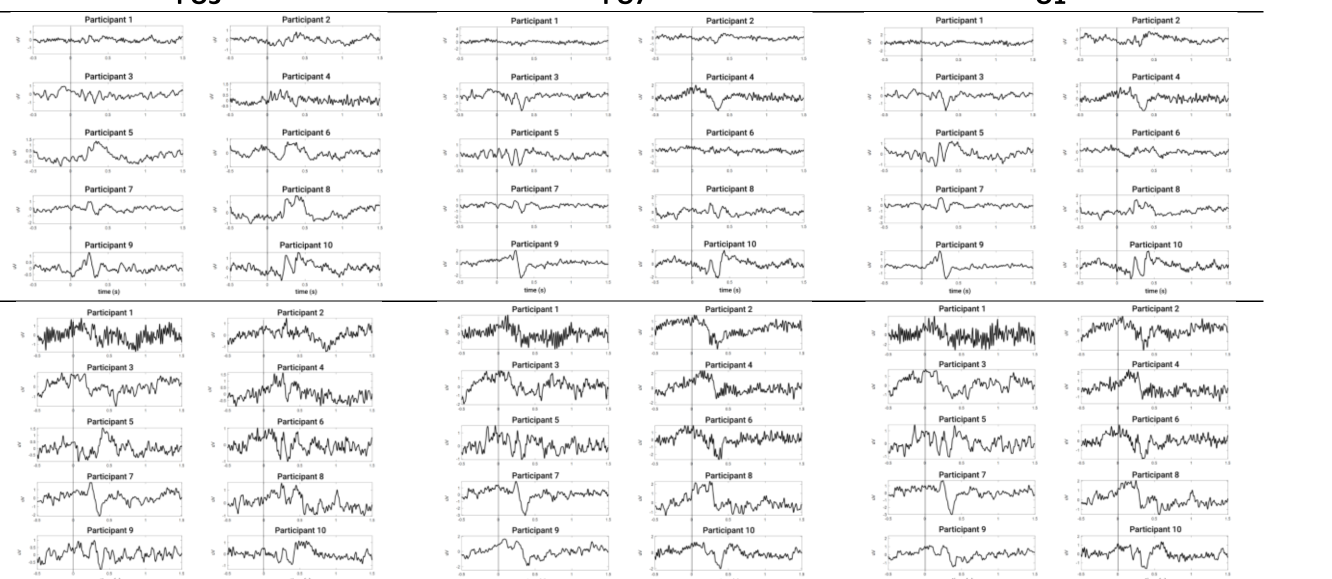
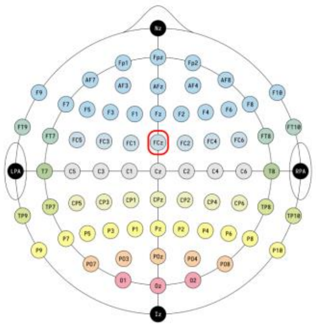
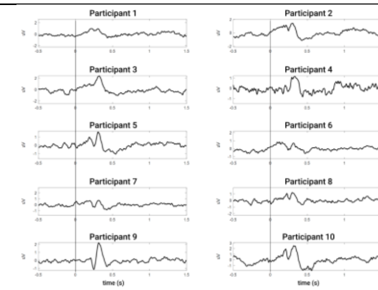
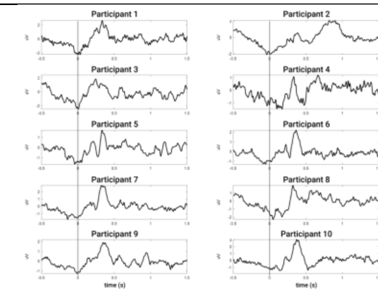
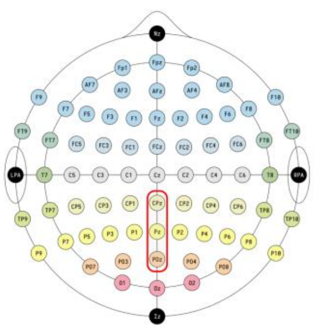
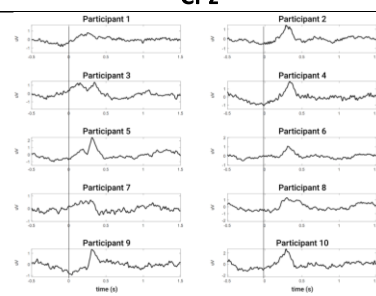
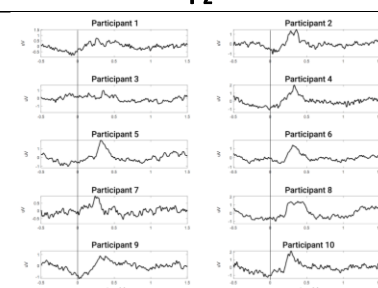
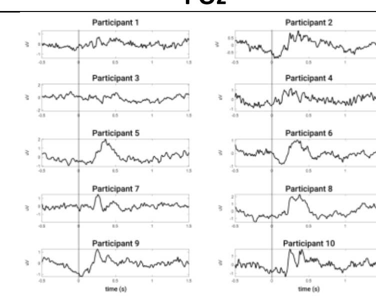
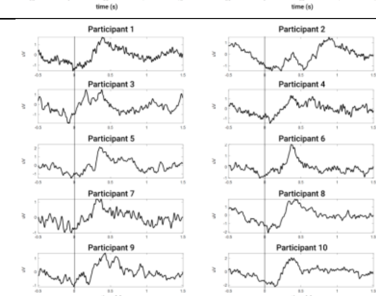
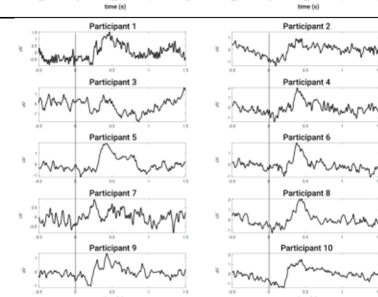
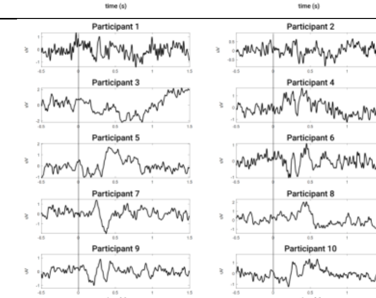
	PO3	PO7	O1
 Positive time events			
Negative time events			

TABLE IX. CENTRAL BRAIN REGION ERPS

a.

		Fz
	Positive time events	
	Negative time events	

b.

		CPz	Pz	POz
	**Positive time events**			
	Negative time events			

c.

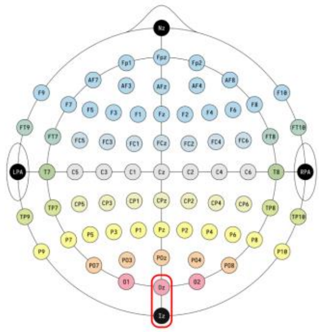
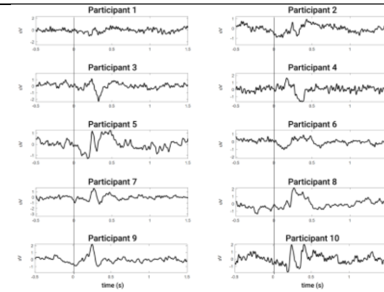
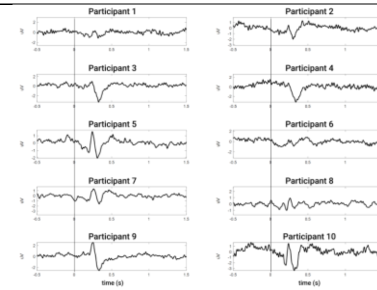
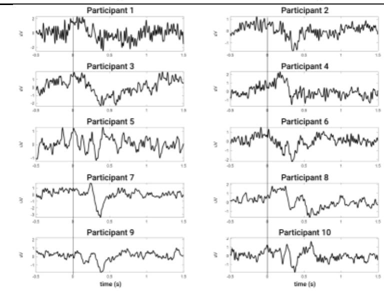
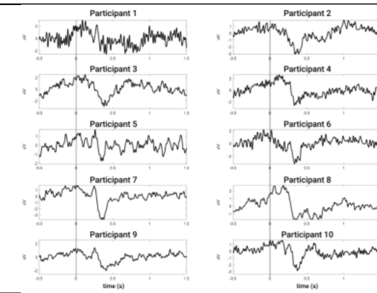
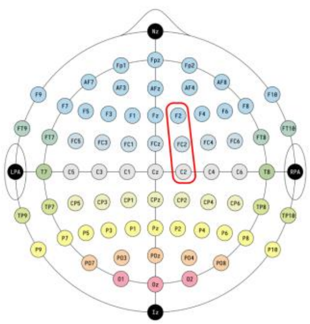
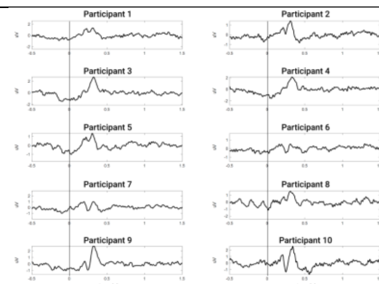
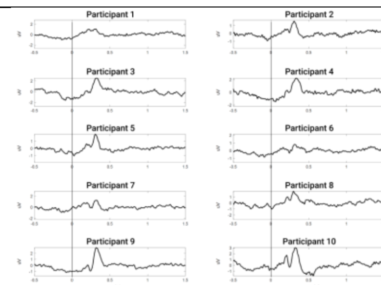
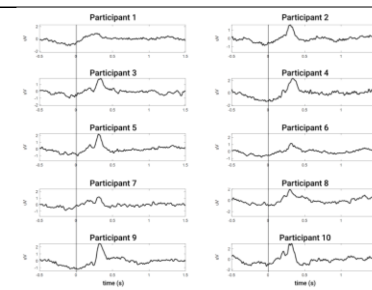
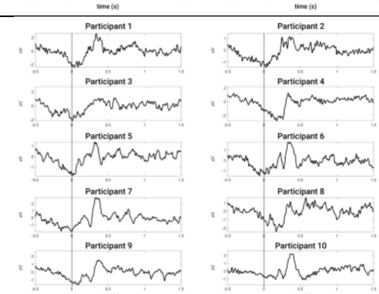
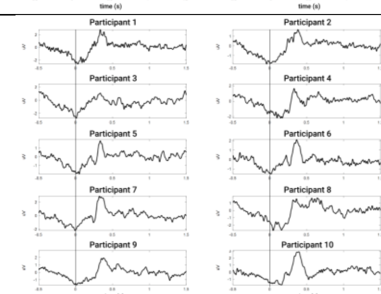
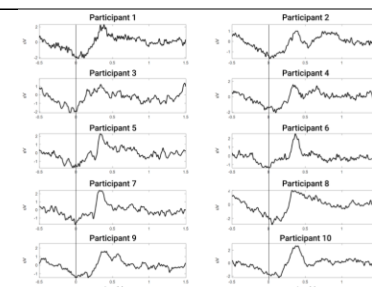
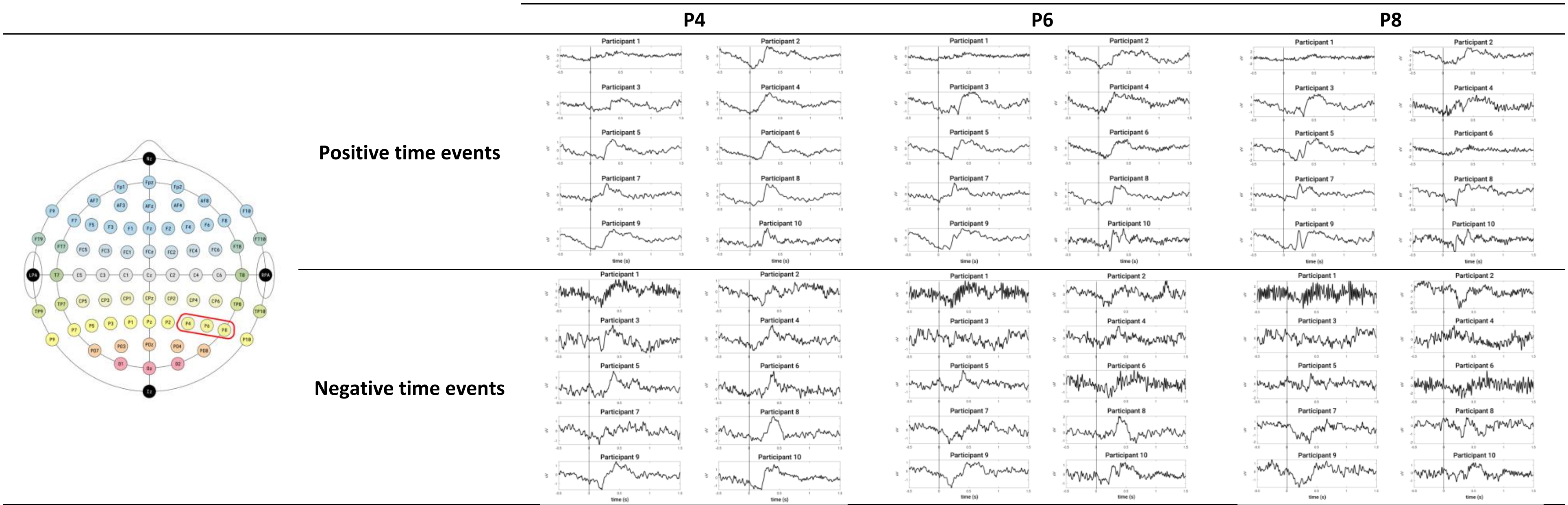
		Oz	Iz
	**Positive time events**		
	Negative time events		

TABLE X. RIGHT BRAIN REGION ERPS

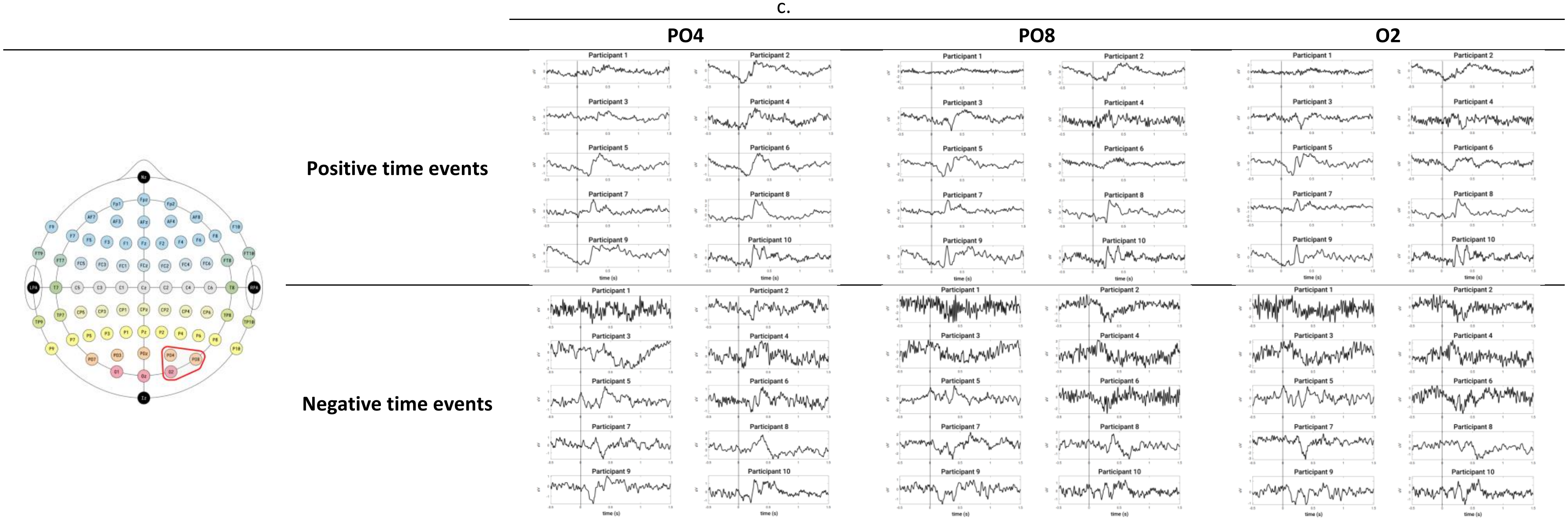
a.

		F2	FC2	C2
	**Positive time events**			
	Negative time events			

b.


P4
P6
P8
Positive time events
Negative time events

c.


PO4
PO8
O2
Positive time events
Negative time events

5.2. ERP on the left-brain region

5.2.1. P5 – P7 – P9 channels

- For positive time events: a negative ERP component is present around 300 ms (N300), with a stable patter before and after the ERP.

- For negative time events: an amplitude increase before event time onset is present, peaking after the event and 100 ms time window, with a negative component around 250 ms (N250). Amplitudes are bigger than amplitudes for positive time events.

5.3. ERP on the central brain region

5.3.1. FCz channel

- Positive time events: the signal exhibits similar characteristics of an ERP P200 (200 milliseconds and positive amplitude), with a stable patter before and after the ERP.

- For negative events: a slightly amplitude decrease before time event onset, and an ERP activation similar to P300 were found.

5.4. ERP on the right brain region

5.4.1. F2 – FC2 – C2 channels

- Positive time events: an ERP component is present at 300 ms after event on set.

- For negative time events: a amplitude decrease before time event onset is perceived, peaking around 0 seconds, with a P300 component.

5.4.2. P4 – P6 – P8 channels

- Positive time events: In contrast with the electrodes positioned in the same parietal line of left region, a slightly decrease on amplitude before event onset and a positive peak around 200 ms after a time event on set is perceived.

- Negative time events: The ERPs activation for negative events presented higher signal's frequency than the signal's frequency for positive time events.

5.5. Discussion

Some ERP components elicited by visual stimuli can be modulated when the images have emotional content: P1 modulated by the presentation of emotional faces, N1 modulated by both pleasant and unpleasant stimuli component may represent early processes in the evaluation of emotional stimuli, N170 and P2 modulated by emotional faces [2]. However, when playing video games, several brain areas are stimulated (occipital lobe for visual processing, parietal and temporal lobes from auditory stimuli, and the frontal lobe for emotional processing, concentration and decision making) [57], and is important to study how particular events can be represented by EEG activity. Some works have focus their analysis on the relation of EEG and time events [89] [90]. When EEG activation in presence of game events was considered, strong correlation of features related to theta, alpha and beta frequency bands across all participants were found, with the strongest scores belonging to the theta and the alpha bands from electrodes positioned on the front, central and occipital brain regions.

The findings are consistent with works that also analyze game events and the related EEG activation. In [51], the authors collected information about a variety of relevant game events, presumably negative events had strong responses with delta band signal component and delta/theta power increase, together with a strongly differentiated ERPs. In addition, rewarding events caused an increase in low delta signal component and high delta power with a robust ERP peaking and plateauing around the time of the P2 component. In this study, P2 characteristics on the EEG signals in presence of positive events were found for FCz, CPz, Pz, and POz channels, with stabilization before the event onset and after 0.5 seconds of the event onset. In contrast, a pronounce peak in amplitude in presence of negative events, and a decrease in amplitude (negative amplitude) before the occurrence of the event was found for FCz channel, this suggest that the participants' signal can predict when a negative event is about to occur (0.5s before the event).

Although, the number of participants in our study is relatively small, the patterns found on the EEG signals across subjects suggest a common activation and perception for both positive and negative events, what is worth to highlight is that emotion appraisal (arousal and valence responses) is an individual process that depends on past experience, memories and cognitive process related to each participant, this can explain why, not too many common EEG features corralled among participants, but still there was higher number of EEG features that correlated in an individual level, for this kind of scenarios, participant tailored approaches to identify emotional reactions are more suitable due to the consideration of individual characteristics and reactions [47], [43]. It is necessary to conduct more experiments with a higher number of participants where characteristics as age, culture, sex, etc., can be considered to analyze the influence on emotion recognition process.

Chapter 6

Conclusion

It is possible to identify emotional reactions using EEG traits using videogames as an emotional stimulus. Videogames are powerful tools to elicit emotions due to the combination of digital media as music, picture, videos, game mechanics and story-telling, with the interaction with virtual environments, becoming an efficient tool to study emotional reactions under HCI scenarios. It is important to analyze how to identify emotional reactions under interactive scenarios using EEG data that allows to detect patterns or correlated information with answers and reactions related to emotion.

The emotional reactions came represented in self-assessment responses, and also, in game time events. For self-assessment responses, theta, alpha, beta and gamma bands of electrodes from the central, occipital, and temporal regions allows to predict arousal values better, in contrast, with the performance of predicting valence values. Addressing game events, we found that EEG traits related with the theta, alpha and beta band had strong correlations. Also, distinctive event-related potentials were identified in the presence of both types of game events that correlated with the emotional responses.

6.1. Aim result

a. Which input information is suitable to increase the accuracy of emotion recognition compare with an existing public dataset, from EEG and individual characteristics.

Due to the nature of the individual characteristics (age, sex, and personality), and the small number of participants, this kind of information is not suitable to use in arousal-valence classifications. For that reason, information related to behavioral reactions related to the emotional stimulus tool were used to improve the classification performance.

There are works that have study the common characteristics and also the difference between emotion appraisal and emotion recognition, with some assuring that culture is useful to increase the classification performance, however there is still need of more studies and evidence that support this conclusion. For the results using the dataset AMIGOS, there is no information about the nationality or cultural background of the participants in this particular study (the study belongs to the Queen Mary University of London, UK, University of Trento, Italy). From here, there is no data available for me to identify if the cultural background has some influence on the results. In the conducted experiment, participants from different nationalities were selected, because the aim of the study was to identify common characteristics in the EEG signals related to emotional appraisal, and as individual characteristics, to focus on how participants reacted to the emotional stimulus and each of the mechanics proposed with the videogame. As a conclusion, the definition of emotions and how each emotion is understood depends in a strong level on the cultural background, but how participants react to emotional events can be related more to the personal experience, and events that shaped the growth and the daily life perception.

Regarding EEG features, PSD and DE features were more suitable to increase the classification performance of both arousal-valence scores and emotional stimulus time events.

b. How to build an emotional elicit tool that allows the study of emotional reaction within and after the exposure time window.

An emotional stimulus tool that allows the study of emotional reactions within and after an experimental time window was designed and implemented. The videogame is very simple because this allows to target the specific emotions for the experiment:

- The challenges are easy to understand, this is related to the simple game mechanics, that were modified to target the different emotions.
- The learning curve is shorter, this is related with the experiment time window and with the experiment length.
- The tasks are engaging, this is related with the purpose of the game, where it is easy to identify the outcome of each level and the whole videogame, and how the performance would affect the final results that is win or lose.

Taking these points into account, if the mechanics get more complex, or the structure change modifying the time exposure, the objective of the game, and the aim of each mechanics, getting specific emotions would be more difficult and more variables would be needed to consider for the analysis.

To identify if the skill level of the participants affected the results obtained in the experiment, two sources of information were correlated with the arousal – valence responses from the participants. One, is the final game scenario score, related to the points got at the end of each videogame scenario while collecting good tokens and avoiding bad tokens. And two, is the final outcome of the videogame, that was defeating a final enemy, if the participant defeated this final enemy, they win the videogame objective, if not, they lose the videogame objective.

Related to the level final scores, in the figure 19, the correlation scores from the participant and the dispersion of scenarios level score between the arousal – valence answers are shown. For arousal, only strong correlations were found for participants 1, 7, 9 and 10. However, from the distribution, it is possible to see that there is some pattern or tendency between the scores and the answers. For valence, only strong correlations were found for participants 1, 7 and 9, in this case the distribution does not have a clear tendency, what it makes more difficult to relate level scores with the valence answers.

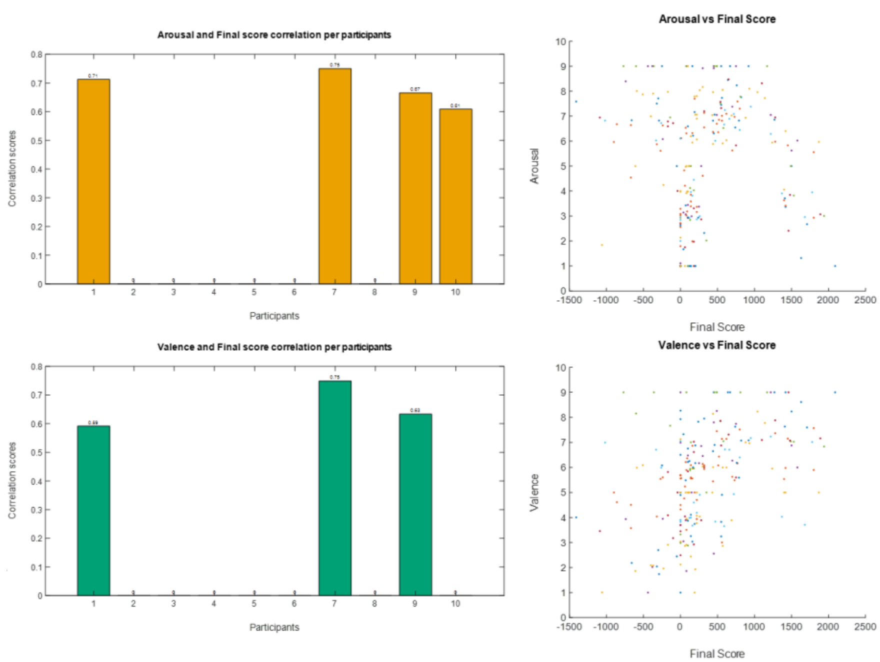


Figure 19. Correlation and dispersion between final game scenario scores and arousal valence answers.

Regarding final videogame outcome (winning or losing the videogame objective), the behavior of the finals scores in each scenario according to the participant skills was inspected, the graphics (figure 20) showed that inside each game scenario, the participants who win have more positive scores and less negative scores in contrast with the participants who lost at the end. In this case, the participants had different performance levels as showed in the charts, and it is no clear that the skill level has a correlation with the emotional self-assessment responses. From the results and the data showed, the participant skill level does not affect the results from the experiment.

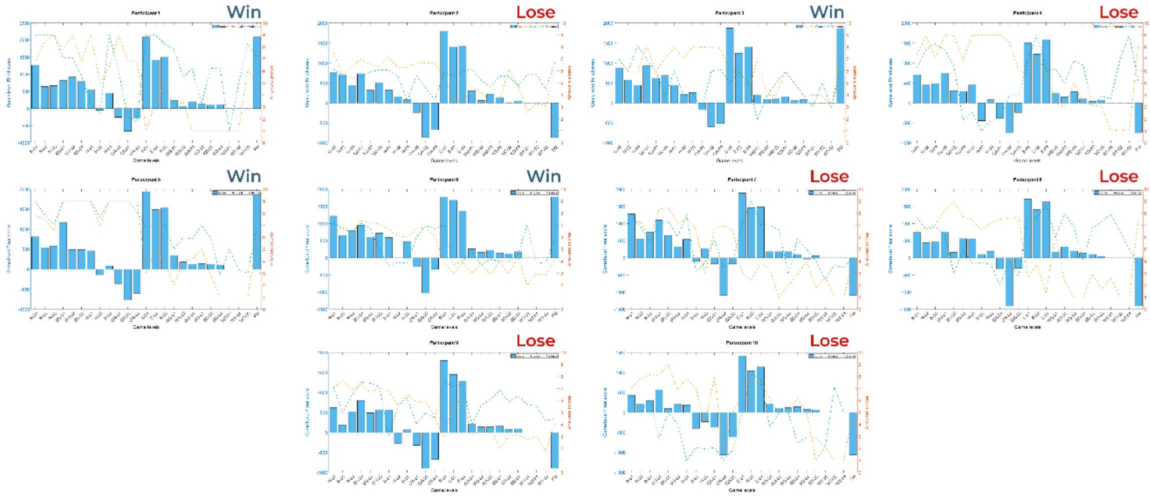


Figure 20. Game scenarios scores across all game scenarios with arousal and valence answers, and final videogame outcome (win or lose).

To identify if finger pressing keys have a correlation with emotion the time traces where the participant pressed a key related with the avatar movement, pressing up and down key arrows were recorded. In figure 21, it is possible to see that the number of pressed keys had a strong positive correlation with the arousal responses, looking at the distribution of the arousal response vs the number of keys pressed in each scenario, it is possible to see some kind of linear tendency. In contrast with the valence scores, only participant 5 had a strong positive correlation with the number of pressed keys. Inspecting the distribution with valence scores, there was no clear tendency.

Inspecting the behavior of pressed keys in each scenario according to the participant skills, figure 22 showed that inside each game level, all the participants tend to have around the same number of pressing keys, the only strong difference is presented in the final game scenario, where participants pressed the keys with higher or lower number of times having different outcomes that can be related with emotion, for example: participant 01 and 05 had around the same number of keystrokes in the main levels and in the final level, both participants won the videogame objective. In contrast participants 02, 03, and 04, had a higher number of keystrokes in the final scenario, compare with the main scenarios, where 2 of them lost the videogame objective and one of them won the objective.

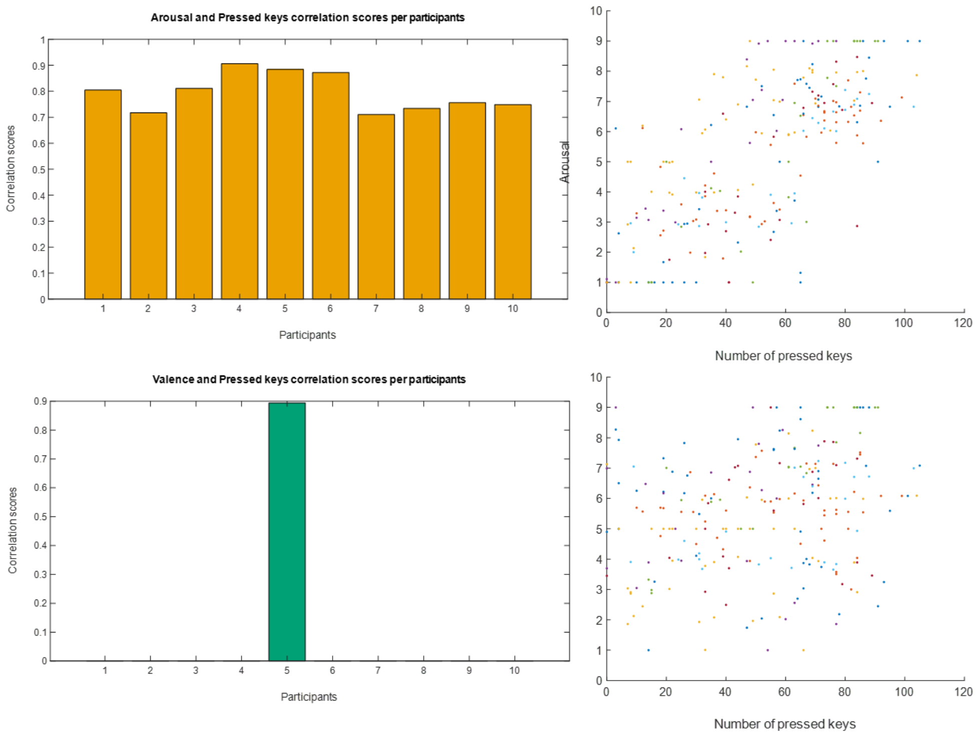


Figure 21. Correlation and dispersion between number of pressed keys and arousal valence answers.

As other example participant 06 and 10 had around the same number of keystrokes in the main levels and in the final level, in this case one participant won the videogame objective, and the other one lost. In contrast, participants 07, 08 and 09, had a higher number of keystrokes in the final level, and the three participants lost the videogame objective. What would be worth to explore is the aim or the intention of the key pressing task, for example, if the participant was trying to collect a token or if the participant was trying to avoid collision with a token, this information can give us more insights about the emotional reaction represented as correlation or ERPs in the EEG signals.

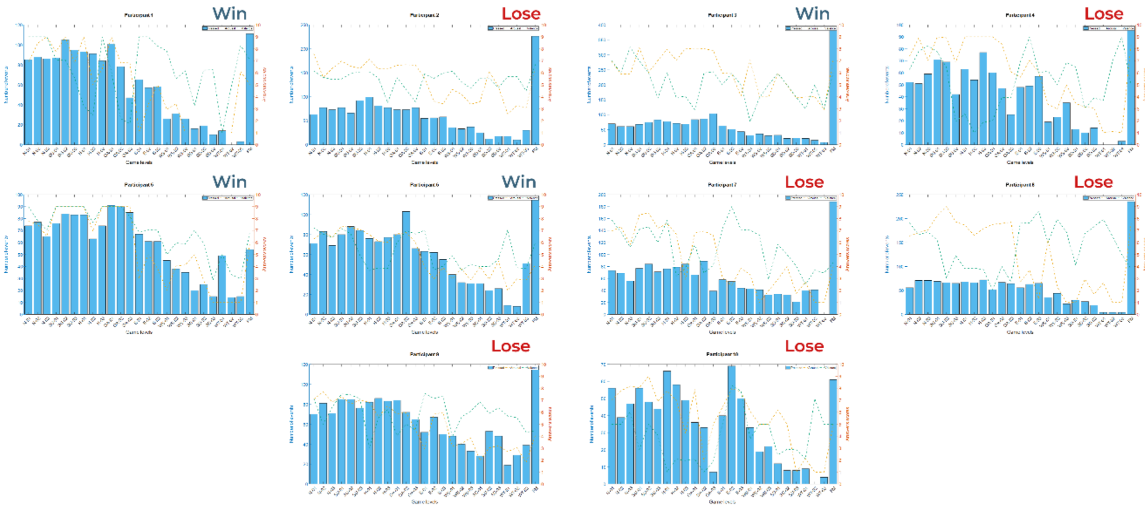


Figure 22. Number of pressed keys across all game scenarios with arousal and valence answers, and final videogame outcome (win or lose)

c. Which EEG characteristics are correlated with self-assessment emotional responses, and emotional stimuli time events.

The correlation scores of the common features among participants was explored. PSD and DE traits on the theta band for channels in the frontal (F), central (C), and parietal (P) regions: FCz, CPz, Cz, FC1, FC2, C1, CP1, CP2 had positive correlation for the majority of the participants. In figure 15. a., the features correlated with arousal for all the participants and the dispersion of each trait is showed. In contrast, in figure b., the only common correlated feature with valence for 4 participants is shown (complexity of PO3 channel had both positive and negative correlations and lower rho scores compare with the features correlated with arousal), from the dispersion is possible to identify that is more difficult to identify a clear pattern. There is a higher number of features that correlated with arousal scores than valence scores across all participants. Arousal scores are related with features in the theta frequency band and electrodes located in the frontal central, central, and central parietal regions.

There is a high number of common EEG features that correlated with game time events for all the participants, PSD and DE features of theta and alpha bands for all the EEG channels except of P2 correlated negative with positive events and positively with negative events. With lower rho scores, PSD features of beta band for all EEG channels except of: Fp1, AF7, F7, F5, FT7, FC5, Fpz, AF8, F8, and P2; and DE features of beta band for all EEG channels except of: Fp1, AF7, F7, F5, FT7, FC5, Fpz, AF8, F8, P2, Fp2, AF8, correlated negative with positive events and positively with negative events.

d. The performance/regression classification of self-assessment emotional responses and emotional stimuli time events using EEG and individual characteristics information.

For arousal regression, MAE had an average of 0.973 ± 0.316 on train set, and 1.199 ± 0.321 on test set, MSE had an average of 1.786 ± 1.122 on train set and on test set 3.186 ± 2.876, we achieved the best result for Participant 6 in the train set (MAE: 0.406 ± 0.299, MSE: 0.328 ± 0.406). For valence regression, the MAE had an average of 1.199 ± 0.321 on train set, and 1.670 ± 0.784 on test set, the MSE had an average of 2.504 ± 1.2112 on train set and on test set 4.680 ± 4.032, we achieved the best result for participant 5 in the train set (MAE: 0.723 ± 0.317, MSE: 0.806 ± 0.655). From the prediction models' implementation, better result for arousal scores were obtained than valence scores in both the train set and the test set (only participant 5 got better results for valence than for arousal scores in both train and test set).

From the classification scores, it was clear the good performance of the classifiers to discriminate between positive and negative time events, the features selected from the majority of the participants were DE from the theta band and form the full EEG frequency spectrum, only for participant 1 the features selected were: DE theta (F1, C1, PO3, CPz, AFz, Fz, F2, FCz, CP2), DE alpha (F1, CPz, F2), and DE from the full EEG frequency spectrum (F1, C1, Pz, CPz, AFz, Fz, F2, FCz, CP2). The results proved that with each of the EEG signals is possible to identify emotional states in the participants using a smaller time window with game events and these events have a relation with the arousal-valence scores gave by the participants.

6.2. Future work

In future research, it is important to address specific challenges like: the access to a wider and diverse population where participant exhibit different demographic characteristics, personality traits, and behavioral cues; the nature of the emotional stimuli, whether they are passive or active; the data gathered and its evaluation during stimuli exposure time; and the interaction type that the participant can experience during using HCI systems. For personalized HCI, it is important to analyze, not only intrinsic characteristics as demographic or personality traits, but also behavioral cues that manifest when using HCI systems and its context. For future works we would like to focus our approach on the capturing and analyzing behavioral cues, together with physiological signals, related to the use of a specific technology or the task being performed.

Bibliography

[1] P. Brey and J. H. Søraker, "Philosophy of Computing and Information Technology," in *Philosophy of Technology and Engineering Sciences*, Elsevier, 2009, pp. 1341–1407.

[2] M. Jeon, *Emotions and Affect in Human Factors and Human-Computer Interaction*. 2017.

[3] R. W. Picard, "Affective Computing." Accessed: Jan. 05, 2020. [Online]. Available: http://www.media.mit.edu/˜picard/.

[4] T. Hewett *et al.*, *ACM SIGCHI Curricula for Human-Computer Interaction*. Association for Computing Machinery, 1992.

[5] B. K. Chakraborty, D. Sarma, M. K. Bhuyan, and K. F. MacDorman, "Review of constraints on vision-based gesture recognition for human-computer interaction," *IET Computer Vision*, vol. 12, no. 1. Institution of Engineering and Technology, pp. 3–15, Feb. 01, 2018, doi: 10.1049/iet-cvi.2017.0052.

[6] R. A. Calvo and S. D'Mello, "Affect detection: An interdisciplinary review of models, methods, and their applications," *IEEE Trans. Affect. Comput.*, vol. 1, no. 1, pp. 18–37, Jan. 2010, doi: 10.1109/T-AFFC.2010.1.

[7] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clin. Neurophysiol.*, vol. 113, no. 6, pp. 767–791, Jun. 2002, doi: 10.1016/S1388-2457(02)00057-3.

[8] G. G. Molina, T. Tsoneva, and A. Nijholt, "Emotional brain-computer interfaces," 2009, doi: 10.1109/ACII.2009.5349478.

[9] C. Mühl, B. Allison, A. Nijholt, and G. Chanel, "A survey of affective brain computer interfaces: principles, state-of-the-art, and challenges," *Brain-Computer Interfaces*, vol. 1, no. 2, pp. 66–84, Apr. 2014, doi: 10.1080/2326263X.2014.912881.

[10] Z. He *et al.*, "Advances in multimodal emotion recognition based on brain–computer interfaces," *Brain Sci.*, vol. 10, no. 10, pp. 1–29, Oct. 2020, doi: 10.3390/brainsci10100687.

[11] M.-K. M. Kim, M.-K. M. Kim, E. Oh, and S.-P. Kim, "A Review on the Computational Methods for Emotional State Estimation from the Human EEG," *Comput. Math. Methods Med.*, vol. 2013, pp. 1–13, 2013, doi: 10.1155/2013/573734.

[12] K. H. Kim, S. W. Bang, and S. R. Kim, "Emotion recognition system using short-term monitoring of physiological signals," *Med. Biol. Eng. Comput.*, vol. 42, no. 3, pp. 419–427, 2004, doi: 10.1007/BF02344719.

[13] J. Tao and T. Tan, "Affective computing: A review," *Affect. Comput. Intell. Interact.*, 2005.

[14] A. Al-Nafjan, M. Hosny, Y. Al-Ohali, and A. Al-Wabil, "Review and Classification of Emotion Recognition Based on EEG Brain-Computer Interface System Research: A Systematic Review," *Appl.*

Sci., vol. 7, no. 12, p. 1239, Dec. 2017, doi: 10.3390/app7121239.

[15] C. G. Coogan and B. He, "Brain-Computer Interface Control in a Virtual Reality Environment and Applications for the Internet of Things," *IEEE Access*, vol. 6, pp. 10840–10849, Feb. 2018, doi: 10.1109/ACCESS.2018.2809453.

[16] S. M. Alarcao and M. J. Fonseca, "Emotions Recognition Using EEG Signals: A Survey," *IEEE Trans. Affect. Comput.*, vol. 3045, no. 3, pp. 1–20, 2017, doi: 10.1109/TAFFC.2017.2714671.

[17] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 10, pp. 1175–1191, Oct. 2001, doi: 10.1109/34.954607.

[18] K. Gouizi, F. B. Reguig, and C. Maaoui, "Analysis physiological signals for emotion recognition," in *7th International Workshop on Systems, Signal Processing and their Applications, WoSSPA 2011*, 2011, pp. 147–150, doi: 10.1109/WOSSPA.2011.5931436.

[19] S. Jerritta, M. Murugappan, K. Wan, and S. Yaacob, "Emotion recognition from facial EMG signals using higher order statistics and principal component analysis," *J. Chinese Inst. Eng. Trans. Chinese Inst. Eng. A/Chung-kuo K. Ch'eng Hsuch K'an*, vol. 37, no. 3, pp. 385–394, Apr. 2014, doi: 10.1080/02533839.2013.799946.

[20] X. Li, D. Song, P. Zhang, Y. Zhang, Y. Hou, and B. Hu, "Exploring EEG Features in Cross-Subject Emotion Recognition," *Front. Neurosci.*, vol. 12, p. 162, Mar. 2018, doi: 10.3389/fnins.2018.00162.

[21] A. Patil, C. Deshmukh, and A. R. Panat, "Feature extraction of EEG for emotion recognition using Hjorth features and higher order crossings," in *Conference on Advances in Signal Processing, CASP 2016*, Nov. 2016, pp. 429–434, doi: 10.1109/CASP.2016.7746209.

[22] H. Y. Ping, L. N. Abdullah, A. A. Halin, and P. S. Sulaiman, "A Study of Physiological Signals-based Emotion Recognition Systems," *Int. J. Comput. Technol.*, vol. 11, no. 1, pp. 2189–2196, Oct. 2013, doi: 10.24297/ijct.v11i1.1190.

[23] D.-W. Chen *et al.*, "A Feature Extraction Method Based on Differential Entropy and Linear Discriminant Analysis for Emotion Recognition," *Sensors*, vol. 19, no. 7, p. 1631, Apr. 2019, doi: 10.3390/s19071631.

[24] J. A. Miranda Correa, M. K. Abadi, N. Sebe, and I. Patras, *AMIGOS: A Dataset for Affect, Personality and Mood Research on Individuals and Groups*, no. 1. 2018.

[25] N. C. Ebner and H. Fischer, "Emotion and aging: Evidence from brain and behavior," *Front. Psychol.*, vol. 5, no. AUG, p. 996, Sep. 2014, doi: 10.3389/fpsyg.2014.00996.

[26] Y. Deng, L. Chang, M. Yang, M. Huo, and R. Zhou, "Gender differences in emotional response: Inconsistency between experience and expressivity," *PLoS One*, vol. 11, no. 6, p. e0158666, Jun. 2016, doi: 10.1371/journal.pone.0158666.

[27] C.-L. Tai, L.-C. Chen, C.-M. Chang, and J.-Y. Hong, "The Impact of Individual Characteristics,

Personality Traits, Perceived Risk on Young People's Intention to Participate in Overseas Working Holiday," *Procedia - Soc. Behav. Sci.*, vol. 57, pp. 388–395, Oct. 2012, doi: 10.1016/j.sbspro.2012.09.1202.

[28] R. R. McCrae and P. T. Costa Jr., "A Five-Factor theory of personality," in *Handbook of personality: Theory and research, 2nd ed.*, New York, NY, US: Guilford Press, 1999, pp. 139–153.

[29] D. Singh and S. K. Singh, "Human Ear Recognition System," in *Encyclopedia of Information Science and Technology*, Third., M. Khosrow-Pour, Ed. IGI Global, 2015.

[30] P. Ekman, "An Argument for Basic Emotions," *Cogn. Emot.*, vol. 6, no. 3–4, pp. 169–200, May 1992, doi: 10.1080/02699939208411068.

[31] J. Russell, "A circumplex model of affect," *J. Pers. Soc. Psychol.*, vol. 39, pp. 1161–1178, 1980, doi: 10.1037/h0077714.

[32] S. N. Mohammed, A. Karim, and A. Hassan, "A Survey on Emotion Recognition for Human Robot Interaction," *CIT. J. Comput. Inf. Technol.*, vol. 28, no. 2, pp. 125–146, Jun. 2020, doi: 10.20532/cit.2020.1004841.

[33] M. M. Bradley and P. J. Lang, "Measuring emotion: The self-assessment manikin and the semantic differential," *J. Behav. Ther. Exp. Psychiatry*, vol. 25, no. 1, pp. 49–59, Mar. 1994, doi: 10.1016/0005-7916(94)90063-9.

[34] P. J. Lang *et al.*, "International Affective Picture System (IAPS): Technical Manual and Affective Ratings," 1997. Accessed: Feb. 08, 2019. [Online]. Available: https://www2.unifesp.br/dpsicobio/adap/instructions.pdf.

[35] B. M. M. and L. P. J., "The International Affective Digitized Sounds (2nd Edition; IADS-2): Affective ratings of sounds and instruction manual.," Gainesville, Fl., 2007.

[36] W.-L. Zheng, W. Liu, Y. Lu, B.-L. Lu, and A. Cichocki, "EmotionMeter: A Multimodal Framework for Recognizing Human Emotions," *IEEE Trans. Cybern.*, vol. 49, no. 3, pp. 1110–1122, Mar. 2019, doi: 10.1109/TCYB.2018.2797176.

[37] L. A. Martinez-Tejada, N. Yoshimura, and Y. Koike, "Classifier comparison using EEG features for emotion recognition process," *SAMI 2020 - IEEE 18th World Symp. Appl. Mach. Intell. Informatics, Proc.*, pp. 225–229, Jan. 2020, doi: 10.1109/SAMI48414.2020.9108746.

[38] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A Multimodal Database for Affect Recognition and Implicit Tagging," doi: 10.1109/T-AFFC.2011.25.

[39] S. Koelstra *et al.*, "DEAP: A Database for Emotion Analysis using Physiological Signals." Accessed: Jul. 26, 2019. [Online]. Available: http://www.eecs.qmul.ac.uk/mmv/datasets/deap/.

[40] R. Subramanian *et al.*, "Ascertain: Emotion and personality recognition using commercial sensors," *IEEE Trans. Affect. Comput.*, vol. 9, no. 2, pp. 147–160, 2018, doi: 10.1109/TAFFC.2016.2625250.

[41] N. Wang and S. Marsella, "Introducing EVG: An Emotion Evoking Game," in *International Workshop*

on Intelligent Virtual Agents, Springer, Berlin, Heidelberg, 2006, pp. 282–291.

[42] N. Ravaja, T. Saari, M. Salminen, J. Laarni, and K. Kallinen, “Phasic Emotional Reactions to Video Game Events: A Psychophysiological Investigation,” *Media Psychol.*, vol. 8, no. 4, pp. 343–367, Nov. 2006, doi: 10.1207/s1532785xmep0804_2.

[43] F. Bevilacqua, H. Engström, and P. Backlund, “Game-Calibrated and User-Tailored Remote Detection of Stress and Boredom in Games,” *Sensors*, vol. 19, no. 13, p. 2877, Jun. 2019, doi: 10.3390/s19132877.

[44] C. S. H. Yeh, “Exploring the effects of videogame play on creativity performance and emotional responses,” *Comput. Human Behav.*, vol. 53, pp. 396–407, Jul. 2015, doi: 10.1016/j.chb.2015.07.024.

[45] H. Perez Martínez, M. Garbarino, and G. N. Yannakakis, “Generic Physiological Features as Predictors of Player Experience,” Springer, Berlin, Heidelberg, 2011, pp. 267–276.

[46] A. Alhargan, N. Cooke, and T. Binjammaz, “Multimodal affect recognition in an interactive gaming environment using eye tracking and speech signals,” in *ICMI 2017 - Proceedings of the 19th ACM International Conference on Multimodal Interaction*, Nov. 2017, vol. 2017-Janua, pp. 479–486, doi: 10.1145/3136755.3137016.

[47] A. Fortin-Côté *et al.*, “FUNii: The Physio-Behavioural Adaptive Video Game,” Springer, Cham, 2019, pp. 14–28.

[48] F. Pallavicini, A. Ferrari, and F. Mantovani, “Video games for well-being: A systematic review on the application of computer games for cognitive and emotional training in the adult population,” *Frontiers in Psychology*, vol. 9, no. NOV. Frontiers Media S.A., p. 2127, Nov. 07, 2018, doi: 10.3389/fpsyg.2018.02127.

[49] J. M. Kivikangas *et al.*, “A review of the use of psychophysiological methods in game research,” *J. Gaming Virtual Worlds*, vol. 3, no. 3, pp. 181–199, Sep. 2011, doi: 10.1386/jgvw.3.3.181_1.

[50] P. Sawangjai, S. Hompoonsup, P. Leelaarporn, S. Kongwudhikunakorn, and T. Wilaiprasitporn, “Consumer Grade EEG Measuring Sensors as Research Tools: A Review,” *IEEE Sensors Journal*, vol. 20, no. 8. Institute of Electrical and Electronics Engineers Inc., pp. 3996–4024, Apr. 15, 2020, doi: 10.1109/JSEN.2019.2962874.

[51] J. F. Cavanagh and J. Castellanos, “Identification of canonical neural events during continuous gameplay of an 8-bit style video game,” *Neuroimage*, vol. 133, pp. 1–13, Jun. 2016, doi: 10.1016/j.neuroimage.2016.02.075.

[52] B. Kerous, F. Skola, and F. Liarokapis, “EEG-based BCI and video games: a progress report,” *Virtual Real.*, vol. 22, no. 2, pp. 119–135, Jun. 2018, doi: 10.1007/s10055-017-0328-x.

[53] L. A. Martínez-Tejada, A. Puertas-González, N. Yoshimura, and Y. Koike, “Exploring EEG Characteristics to Identify Emotional Reactions under Videogame Scenarios,” *Brain Sci. 2021, Vol. 11, Page 378*, vol. 11, no. 3, p. 378, Mar. 2021, doi: 10.3390/BRAINSCI11030378.

[54] T. B. Alakus, M. Gonen, and I. Turkoglu, “Database for an emotion recognition system based on EEG signals and various computer games – GAMEEMO,” *Biomed. Signal Process. Control*, vol. 60, p. 101951, Jul. 2020, doi: 10.1016/j.bspc.2020.101951.

[55] J. Marín-Morales *et al.*, “Affective computing in virtual reality: emotion recognition from brain and heartbeat dynamics using wearable sensors,” *Sci. Rep.*, vol. 8, no. 1, pp. 1–15, Dec. 2018, doi: 10.1038/s41598-018-32063-4.

[56] L. Ghosh, S. Saha, and A. Konar, “Decoding emotional changes of android-gamers using a fused Type-2 fuzzy deep neural network,” *Comput. Human Behav.*, vol. 116, p. 106640, Mar. 2021, doi: 10.1016/j.chb.2020.106640.

[57] G. Chanel, C. Rebetez, M. Bétrancourt, and T. Pun, “Emotion assessment from physiological signals for adaptation of game difficulty,” *IEEE Trans. Syst. Man, Cybern. Part ASystems Humans*, vol. 41, no. 6, pp. 1052–1063, Nov. 2011, doi: 10.1109/TSMCA.2011.2116000.

[58] S. Rukavina, S. Gruss, H. Hoffmann, J. W. Tan, S. Walter, and H. C. Traue, “Affective computing and the impact of gender and age,” *PLoS One*, vol. 11, no. 3, 2016, doi: 10.1371/journal.pone.0150584.

[59] W. L. Wei, C. H. Wu, J. C. Lin, and H. Li, “Interaction style detection based on Fused Cross-Correlation Model in spoken conversation,” in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2013, pp. 8495–8499, doi: 10.1109/ICASSP.2013.6639323.

[60] A. Guo and J. Ma, “Archetype-based modeling of persona for comprehensive personality computing from personal big data,” *Sensors (Switzerland)*, vol. 18, no. 3, Mar. 2018, doi: 10.3390/s18030684.

[61] X. S. Wei, C. L. Zhang, H. Zhang, and J. Wu, “Deep Bimodal Regression of Apparent Personality Traits from Short Video Sequences,” *IEEE Trans. Affect. Comput.*, vol. 9, no. 3, pp. 303–315, 2018, doi: 10.1109/TAFFC.2017.2762299.

[62] F. Nasoz, C. L. Lisetti, and A. V. Vasilakos, “Affectively intelligent and adaptive car interfaces,” *Inf. Sci. (Ny).*, vol. 180, no. 20, pp. 3817–3836, 2010, doi: 10.1016/j.ins.2010.06.034.

[63] R. Jenke, A. Peer, and M. Buss, “Feature Extraction and Selection for Emotion Recognition from EEG,” *IEEE Trans. Affect. Comput.*, vol. 5, no. 3, pp. 327–339, Jul. 2014, doi: 10.1109/TAFFC.2014.2339834.

[64] H. Becker, J. Fleureau, P. Guillotel, F. Wendling, I. Merlet, and L. Albera, “Emotion recognition based on high-resolution EEG recordings and reconstructed brain sources,” *IEEE Trans. Affect. Comput.*, pp. 1–1, 2017, doi: 10.1109/TAFFC.2017.2768030.

[65] L. A. Martínez-Tejada, Y. Maruyama, N. Yoshimura, and Y. Koike, “Analysis of Personality and EEG Features in Emotion Recognition Using Machine Learning Techniques to Classify Arousal and Valence Labels,” *Mach. Learn. Knowl. Extr. 2020, Vol. 2, Pages 99-124*, vol. 2, no. 2, pp. 99–124, Apr. 2020, doi: 10.3390/MAKE2020007.

[66] F. Pedregosa *et al.*, “Scikit-learn: Machine Learning in Python,” *J. Mach. Learn. Res.*, vol. 12, no. Oct,

pp. 2825–2830, 2011, Accessed: Jul. 11, 2019. [Online]. Available: http://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html.

[67] A. Boschetti and L. Massaron, *Python Data Science Essentials - Second Edition*. Packt Publishing, 2016.

[68] R. Hunicke, M. Leblanc, and R. Zubek, "MDA: A formal approach to game design and game research," in *AAAI Workshop - Technical Report*, 2004, vol. WS-04-04, pp. 1–5.

[69] R. Dillon, *On the way to fun: An emotion-based approach to successful game design*. A K Peters/CRC Press, 2010.

[70] K. Isbister, *How games move us : emotion by design*. MIT Press, 2017.

[71] J. Kim and E. André, "Emotion recognition based on physiological changes in music listening," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 12, pp. 2067–2083, 2008, doi: 10.1109/TPAMI.2008.26.

[72] Y. H. Yang, Y. C. Lin, Y. F. Su, and H. H. Chen, "A regression approach to music emotion recognition," *IEEE Trans. Audio, Speech Lang. Process.*, vol. 16, no. 2, pp. 448–457, Feb. 2008, doi: 10.1109/TASL.2007.911513.

[73] Y. E. Kim *et al.*, "Music emotion recognition: A state of the art review," in *Proceedings of the 11th International Society for Music Information Retrieval Conference, ISMIR 2010*, 2010, pp. 255–266, Accessed: Mar. 03, 2021. [Online]. Available: http://www.allmusic.com.

[74] Y.-H. Yang and H. H. Chen, *Music emotion recognition*. CRC Press, 2011.

[75] L. A. Martinez-Tejada, A. Puertas Gonzalez, N. Yoshimura, and Y. Koike, "Videogame design as a elicit tool for emotion recognition experiments," in *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2020, vol. 2020-Octob, pp. 4320–4326.

[76] J. A. Urigüen and B. Garcia-Zapirain, "EEG artifact removal - State-of-the-art and guidelines," *J. Neural Eng.*, vol. 12, no. 3, p. 031001, Jun. 2015, doi: 10.1088/1741-2560/12/3/031001.

[77] T.-P. Jung *et al.*, "Removing electroencephalographic artifacts by blind source separation," *Psychophysiology*, vol. 37, no. 2, pp. 163–178, Mar. 2000, doi: 10.1111/1469-8986.3720163.

[78] T. P. Jung, S. Makeig, M. Westerfield, J. Townsend, E. Courchesne, and T. J. Sejnowski, "Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects," *Clin. Neurophysiol.*, vol. 111, no. 10, pp. 1745–1758, Oct. 2000, doi: 10.1016/S1388-2457(00)00386-2.

[79] S. Chandra, G. Sharma, A. A. Salam, D. Jha, and A. P. Mittal, "Playing Action Video Games a Key to Cognitive Enhancement," in *Procedia Computer Science*, Jan. 2016, vol. 84, pp. 115–122, doi: 10.1016/j.procs.2016.04.074.

[80] A. K. Singh, Y. K. Wang, J. T. King, and C. T. Lin, "Extended interaction with a bci video game changes resting-state brain activity," *IEEE Trans. Cogn. Dev. Syst.*, vol. 12, no. 4, pp. 809–823, Dec. 2020, doi: 10.1109/TCDS.2020.2985102.

[81] J. R. Vergara and P. A. Estévez, “A review of feature selection methods based on mutual information,” *Neural Computing and Applications*, vol. 24, no. 1. Springer, pp. 175–186, Jan. 13, 2014, doi: 10.1007/s00521-013-1368-0.

[82] G. Lemaitre, F. Nogueira, D. Oliveira, and C. Aridas, “Ensemble of samplers — imbalanced-learn 0.7.0 documentation,” 2016. https://imbalanced-learn.org/stable/ensemble.html (accessed Feb. 02, 2021).

[83] M. B. Kostyunina and M. A. Kulikov, “Frequency characteristics of EEG spectra in the emotions,” *Neurosci. Behav. Physiol.*, vol. 26, no. 4, pp. 340–343, 1996, doi: 10.1007/BF02359037.

[84] X. W. Wang, D. Nie, and B. L. Lu, “EEG-based emotion recognition using frequency domain features and support vector machines,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Nov. 2011, vol. 7062 LNCS, no. PART 1, pp. 734–743, doi: 10.1007/978-3-642-24955-6_87.

[85] W. Zheng, “Multichannel EEG-Based Emotion Recognition via Group Sparse Canonical Correlation Analysis,” *IEEE Trans. Cogn. Dev. Syst.*, vol. 9, no. 3, pp. 281–290, Sep. 2017, doi: 10.1109/TCDS.2016.2587290.

[86] W. L. Zheng and B. L. Lu, “Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks,” *IEEE Trans. Auton. Ment. Dev.*, vol. 7, no. 3, pp. 162–175, Sep. 2015, doi: 10.1109/TAMD.2015.2431497.

[87] N. Jatupaiboon, S. Pan-Ngum, and P. Israsena, “Emotion classification using minimal EEG channels and frequency bands,” in *Proceedings of the 2013 10th International Joint Conference on Computer Science and Software Engineering, JCSSE 2013*, 2013, pp. 21–24, doi: 10.1109/JCSSE.2013.6567313.

[88] J. Zhang, P. Chen, S. Nichele, and A. Yazidi, “Emotion Recognition Using Time-frequency Analysis of EEG Signals and Machine Learning,” in *2019 IEEE Symposium Series on Computational Intelligence, SSCI 2019*, Dec. 2019, pp. 404–409, doi: 10.1109/SSCI44817.2019.9003057.

[89] A. Stein, Y. Yotam, R. Puzis, G. Shani, and M. Taieb-Maimon, “EEG-triggered dynamic difficulty adjustment for multiplayer games,” *Entertain. Comput.*, vol. 25, pp. 14–25, Mar. 2018, doi: 10.1016/j.entcom.2017.11.003.

[90] D. Chen, J. James, F. S. Bao, C. Ling, and T. Fan, “Relationship between video game events and player emotion based on EEG,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016, vol. 9733, pp. 377–384, doi: 10.1007/978-3-319-39513-5_35.