UNIVERSITÀ DEGLI STUDI DI MILANO

PhD School in

Computer Science

Computer Science Department
"Giovanni Degli Antoni"

PhD in

Computer Science

XXXI Cycle

Emotions Recognition in Video Game Players Using Physiological Information

INF/01

PhD candidate:

Marco Granato

Advisor:

Dr. Davide Gadia

School Director:

Prof. Paolo Boldi

Academic Year 2017/18

Dedicated to Caffeine,
faithful colleague, in work and in life

“ *Turn your computer off and go to sleep!* ”

The Secret of Monkey Island

Acknowledgements

When I started my Ph.D., I had no idea on what really means the term "research" and how to conduct it (and I'm not sure I understood it fully now). In the different years of my Ph.D., I followed a path that developed my research knowledge, such as my person. This path was not drawn by me alone, but I have to thank all the *islands* that I have met during in these years.

The first big thank is dedicated to Davide, also known as mine Advisor. I know that you addressed several difficulties to follow a student like me. I am aware of being very far from the perfect student. However, you had put almost all your effort to literally try to transform me, and I hope to have given to you at least some gratifications. As a consequence, thank for all the time that you have dedicated to me. A second big thank is dedicated to Dario and Laura. Dario, you supported me not only as professor but also as teacher, I think that you can not even imagine how much this helped me grow as a person. Laura, you patiently addressed my knowledge in the right way, suggesting resources, and giving me a ticking off when there was a need. Thank you, Davide, Dario, and Laura to have faced every difficulty with wise advice.

I would like to sincerely thank the reviewers for their careful comments and suggestions that supported me in the dissertation improving. I have never met most of you, but I am sure you are fantastic people.

Another thank to my colleagues and the students of the PONG laboratory, with you I experienced so many wonderful moments. In particular, thank Marco and thank Alessandro!

Thank Alieno, without you this thesis would was impossible. You supported my research days and nights, you are tireless.

Thank to all my friends. Wherever you are you still feel very close and give me strength. And thank the Geas Sesto San Giovanni Rugby team, you always gave me happiness even on the most difficult days. Thank my family that supported me in this path. I know that I am not always easy to manage, and I was very absent. You accepted everything and you still support me.

Andrea, I have no word to explain how you were important during my Ph.D. path and in my path. I made you endure everything, but you still stand by my side. I believe that almost everything I have done is thanks to you.

And of course thanks to all the people that supported my thesis, I have no space for thank all of you, but I appreciated all your effort that you gave me. I am convinced that many of them do not know how essential they are. However, it is very well known that "what is essential is invisible to the eye".

Ah, I was forgetting, a big thank also to myself!

Contents

List of Figures

List of Tables

List of Most Used Abbreviations

ADC Analog to Digital Converter

ANS Autonomic Nervous System

AS Affective Slider

BCI Brain-Computer Interface

BPM Beats Per Minute

BVP Blood Volume Pressure

CNS Central Nervous System

CV Cross Validation

DAPIS Data Acquisition of Physiological Information Software

DDA Dynamic Difficulty Adjustment

ECG Electrocardiogram

EDA Electrodermal Activity

EEG Electroencephalography

EMG Electromyography

EOG Electrooculogram

ESAT Emotion Self-Assessment Tool

FACS Facial Action Coding System

GPR Gaussian Process Regression

GSR Galvanic Skin Response

GUR Game User Research

HCI Human-Computer Interaction

HR Heart Rate

MAE Mean Absolute Error

ML Machine Learning

MSE Mean Square Error

NPC Non-Player Character

NRMSE Normalized Root Mean Square Error

PAD Pleasure(Valence)-Arousal-Dominance

PCars Project Cars

RAGA RAcing GAme Dataset

RF Random Forest

RMSE Root Mean Square Error

RO RedOut

SAM Self-Assesment Manikins

SC Skin Conductance

SCL Skin Conductance Level

SCR Skin Conductance Response

SFFS Sequential Floating Forward Selection

SR Skin Resistance

SRL Skin Resistance Level

SRR Skin Resistance Response

VA Valence-Arousal

VR Virtual Reality

UNIVERSITÀ DEGLI STUDI DI MILANO

Abstract

Computer Science Department
"Giovanni Degli Antoni"

Ph.D. Thesis

Emotions Recognition in Video Game Players Using Physiological Information

by Marco GRANATO

Video games are interactive software able to arouse different kinds of emotions in players. Usually, the game designer tries to define a set of game features able to enjoy, engage, and/or educate the consumers. Through the gameplay, the narrative, and the game environment, a video game is able to interact with players' intellect and emotions. Thanks to the technological developments of the last years, the gaming industry has grown to become one of the most important entertainment markets. The scientific community and private companies have put a lot of efforts on the technical aspects as well as on the interaction aspects between the players and the video game. Considering the game design, many theories have been proposed to define some guidelines to design games able to arouse specific emotions in consumers. They mainly use interviews or observations in order to deduce the goodness of their approach through qualitative data.

There are some works based on empirical studies aimed at studying the emotional states directly on players, using quantitative data. However, these researches usually consider the data analysis as a classification problem involving, mainly, the game events.

Our goal is to understand how the feelings, experienced by the players, can be automatically deducted, and how these emotional states can be used to improve the game quality. In order to pursue this purpose, we measured the mental states using physiological signals in order to return a set of quantitative values used to identify the players emotions. The most common ways to identify emotions are: to use a discrete set of labels (e.g., joy, anger), or to assess them inside an n-dimensional vector space. Albeit the most natural way to describe the emotions is to represent them through their name, the latter approach provides a quantitative result that can be used

to define the new game status. In this thesis, we propose a framework aimed at an automatic assessment, using physiological data, of emotions in a 2-dimensional space, structured by valence and arousal vectors. The former may vary between pleasure and displeasure, while the latter defines the level of physiological activation. As a consequence, we considered as most effective to infer the players' mental states, the following physiological data: electrocardiography (ECG), electromyography on 5 facial muscles (Facial EMG), galvanic skins response (GSR), and respiration intensity/rate. We recorded a video, during a set of game sessions, of the player's face and of her gameplay. To acquire the affective information, we showed the recorded video and audio to the player, and we asked to self-assess her/his emotional state over the entire game on the valence and arousal vectors presented above.

Starting from this framework, we conducted two sets of experiments. In the first experiment, our aim was to validate the procedure. We collected the data of 10 participants while playing at 4 platform games. We also analyzed the data to identify the emotion pattern of the player during the gaming sessions. The analysis was conducted in two directions: individual analysis (to find the physiological pattern of an individual player), and collective analysis (to find the generic patterns of the sample population).

The goal of the second experiment was to create a dataset of physiological information of 33 players, and to extend the data analysis and the results provided by the pilot study. We asked the participants to play at 2 racing games in two different environments: on a standard monitor and using a head mounted display for Virtual Reality. After we collected the information useful to the dataset creation, we analyzed the data focusing on individual analysis. In both analyses, the self-assessment and the physiological data were used in order to infer the emotional state of the players in each moment of the game sessions, and to build a prediction model of players' emotions using Machine Learning techniques.

Therefore, the three main contributions of this thesis are: to design a novel framework for study the emotions of video game players, to develop an open-source architecture and a set of software able to acquire the physiological signals and the affective states, to create an affective dataset using racing video games as stimuli, to understand which physiological conditions could be the most relevant in order to determine the players' emotions, and to propose a method for the real-time prediction of a player's mental state during a video game session. The results suggest that it is possible to design a model that fits with player's characteristics, predicting her emotions. It could be an effective tool available to game designers who can introduce innovative features to their games.

Chapter 1

Introduction

HUIZINGA [1] defines the action to play a game as "a voluntary activity or occupation executed within certain fixed limits of time and place, according to rules freely accepted but absolutely binding, having its aim in itself and accompanied by a feeling of tension, joy and the consciousness that it is *different* from *ordinary life*". One of the main reasons people play games is that they stimulate and generate all types of emotional responses [2]. Thus, generally, the development of a game is focused on the players, with the main goal of inducing entertainment and engagement. Hence, a game designer has to consider which is the target audience and, consequentially, how the players' intellect and emotions will interact with the game.

A video game uses audio/visual information presented through electronic devices in order to communicate the game structure. Furthermore, Sid Meier, a famous video game designer, described the games as "a set of interesting decisions" [3]. Thus, the players' decisions may provide a specific gaming experience that, usually, may have a different emotional impact. In order to investigate the connection between decision making and emotions, Damasio developed the Somatic Marker Hypothesis (SMH) [4]: the author described how the decisions are defined by previous outcomes. In particular, humans, consciously or unconsciously, associate somatic markers to their past outcomes. Consequentially, if a person perceives a *positive* somatic marker, she may be encouraged to continue in her behavior, feeling a sensation of "happiness" (vice versa for a *negative somatic marker*). In addition, the author described [5] how emotions alter physiological condition in relation to a specific stimulus, and how they can modify the future decisions. Starting from this concept, many researches investigated the relationship between video games and emotions [6, 7, 8]. A good video game can create engagement with the player, maximizing the emotions induced by the game choices. Thus, a game designer should consider and balance the different video game features in order to maintain players' attention and to generate the desired emotional response [9]. Furthermore, the players should adhere to a subset of game rules in order to receive a game output (e.g., use a sword to kill an enemy). This mechanism lays the foundation to an *Affective Loop* [10, 11], where the human-computer interaction becomes an emotional communication process. *Spatial Presence* [12] and *Flow* [13] are two theories of positive psychology that are commonly used to identify how a video game, or a generic entertainment product (e.g., a movie), interacts with the human emotions. *Spatial Presence* is a psychological condition describing how much a player has the illusion to be transported in a virtual environment. This condition can be better elicited by immersive technologies (e.g., Virtual Reality). Researchers have suggested that a high sense of *Spatial Presence* can improve players' entertainment and it may also facilitate the players' performance [14]. The *Theory of Flow* tries, instead, to define a mental state where a user is completely absorbed in a task. This theory describes a balanced channel between challenge and ability: when in the *flow* state, a person can benefit of an experience of achievement and happiness. In a video game session, the sensation of *Flow* seems to be connected with an increase of dopamine level, a neurotransmitter that increases human attention [15].

These theories are only a subset of approaches aimed at describing the link between video games and emotions: usually, studies on emotions and games involve different disciplines, like e.g., psychology, physiology, computer science, etc. *Affective Computing* tries to merge these disciplines: its main purpose is to automatize the recognition and/or simulation of human emotions through a computer [16]. Our contribution aims to extend the researches in the affective computing field, exploring a novel area regarding emotions recognition during video games fruition. Improving the studies of emotions during video games fruition may improve the overall quality of the product, and, consequentially, the engagement with the consumers. Moreover, this work may provide a contribution to Game User Research (GUR) [17]. It is a novel approach which aims to consider practical and research methods in order to ensure an optimal game quality. This approach involves different disciplines, and it considers all the aspects in game development, such as, infrastructures, controls, menus, customer support, etc. Our research follows a scientific robust methodology, taking care to follow the GUR guidelines, and providing a framework able to optimize the user experience. This dissertation can also provide a contribution to the development of more effective *Serious Games*. A *Serious Game "is a game in which education (in its various forms) is the primary goal, rather than entertainment"* [18]. Thus, serious games differ from the classic educational tools as they use a different framework, namely that of the game, to achieve an educational purpose using an entertaining product. As an entertaining and interactive software, a serious game should arouse specific emotions in order to engage the players and to transmit adequately the educational message as intended by the game designers. As a consequence, improving the video game emotional effect does not affect only the playful aspect, but it may also improve the overall life quality, providing a powerful tool able to enhance the general welfare, like e.g., entertainment, education, sanity, etc.

As we mentioned above, some theories were developed to improve the possibility to design a video game able to have an impact on players' emotions. However, these researches are based mainly on the authors' expertise, and on the observations of the players' behavior. Consequentially, the aim of our research is to design a framework that uses affective computing approaches to understand, with empirical evidence, how to reveal the players' emotions during video games fruition. Thus, we performed a set of experiments on players during gaming sessions. We acquired physiological and emotional information in order to predict the mental state during video game fruition. Starting from the results of the experiments, the main contribution of this thesis is to develop a valid framework aimed at providing to game designers useful hints regarding how to adapt their games to the players' conditions, like, e.g., to apply a novel and real-time method of Dynamic Difficulty Adjustment (DDA) [19]. Furthermore, this thesis will provide other academic contributions:

- The creation of an affective dataset named RAcing GAme (RAGA). It will be freely available to the scientific community, and it will contain physiological and emotional (self-assessment) data of players, acquired during video game sessions
- Starting from the collected data, to provide an analysis of the more effective physiological information used to predict the players' emotions
- To propose a supervised learning method for the real-time prediction of a player's mental state during a video game session.

The dissertation is structured as follows: in Ch 2 we provide the necessary background information useful to understand the concepts presented in this thesis. We discuss about how Affective Computing and the physiological information connected with players' emotions can be integrated in video game research field. In Ch 3, we contextualize the research in affective computing and video games researches, investigating the different connection points between the disciplines. The thesis proceeds in Ch 4 with an overview of the methodologies applied to design the research framework. Thus, we provide details about the different types of emotions and the physiological data involved in the research. Furthermore, we provide the architectural and software design in Ch 5. in Ch 6, we describe the first experimental environment (a pilot study) used to validate the framework. We have acquired data from 10 participants and we have applied Machine Learning (ML) techniques to predict the emotional participants' states considering an individual and a generic analysis. Then, in Ch 7, we describe an extension of the previous experiment: we have involved 33 participants, and we have improved the experimental setup, enhancing the quality of the physiological signals acquisition, developing a novel method of feature selection, and using a more accurate ML algorithm, in order to propose a more effective framework and to extend the results. Furthermore, we present the RAGA Dataset, and we describe the analysis applied to the collected data to predict the player emotional state during video games fruition. in Ch 8, we discuss the achieved results and we provide final considerations on the overall research. Lastly, we suggest some possible future works that will involve the design of a general model able to predict the emotional state without a specific training on the player. Finally, the appendices provide further details on the outcomes of the analysis of RAGA dataset. In particular, in Appendix A, we present detailed information about the features considered in each experiment, in Appendix B we show the boxplots which represent the overall accuracy of the different applied models, and in Appendix C, we compare the outcomes of the final model, considering different indexes.

Chapter 2

Background

IN the following sections, we present the background information useful to understand the overall dissertation. In particular, in Sec. 2.1, we discuss the different proposed definition of a game, and an overview of game design elements able to arouse emotions in players. The arguments proposed in this section are valid for video games as well, since they are a subgroup of games.

In Sec. 2.2, we describe how an emotion arouses in humans, the physiological effects of emotions in the human body, and the possible approaches to detect these phenomena.

2.1 Emotions in Games

Playing is a natural activity practiced by humans, such as the mammals and some birds [20]. It fills the natural need to develop an emotional flexibility through the experience of different emotional aspects. Even if the activity of play can be observed across the species, humans are the only mammals that play at games. Games are activities with the main goal to entertain the players and they are limited in time and space. However, to find an exhaustive definition of *game* can be a challenge. Usually, the game definitions are focused on particular game characteristics, and, therefore, they can not provide a unique and exhaustive description of a *game*. Moreover, the definitions of *game* are not necessarily exclusive, the different descriptions (or parts of them) can coexist. In our opinion, the most extensive and complete definition of "*game*" is provided by Huizinga [1] and it is reported at the beginning of Ch. 1. However, different authors have provided a definition of *game*. Frasca, in his Ph.D. thesis [21], underlines a crucial point in the definition of game, since it should be related to both system and activity. Both elements are important in games, and a definition which excludes one of these two aspects can be considered uncompleted. Another *game* definition is provided by Sid Meier, which considers a video game as a set of interesting decisions, as already discussed in Ch. 1. Another definition provided by Salen and Zimmerman describes a game as "a system in which players engage in an artificial conflict, defined by rules, that results in a quantifiable outcome" [22]. Fullerton provides a description of what is a game from the game designer point of view. She describes the *art of game design* as the ability "to create that elusive combination of challenge, competition, and interaction that players just call *fun*" [23].

Arjoranta observed that the game definitions are quite similar, since, usually, the authors look the previous definitions, find common elements and problems, discern problems and provide a synthesis able to fix them [24]. The author also quoted the dissertation of Wittgenstein [25] (first philosopher that defines the word *game*), and he summarizes the philosopher argumentation declaring that it is not possible to provide a unique definition of games, since they have not a common core of attributes, but they "share attributes as family resemblances, which vary from one instance to another". Starting from this concept, we reduced the definitions to the minimum, and we consider, for our study, "*a game as an activity which is structured by rules able to arouse emotions in players*", which is quite similar to dictionary definition provided by Merriam-Webster: "activity engaged in for diversion or amusement"[1]. As a consequence, we studied different approaches, provided by expert authors in the game design field, to study how to use the games mechanics and environment to arouse emotions in players.

Hunicke et al. [26] tried to design, to the best of our knowledge, the first formal framework to study the concept of *game*. This formal approach tries to develop a tool which considers the games as artifacts, transforming the game components in the design counterparts. As a consequence:

- the *Rules* become ***Mechanics***, with this term the authors "describes the particular components of the game, at the level of data representation and algorithms",
- the *System* becomes ***Dynamics***, with this term the authors "describes the run-time behavior of the mechanics acting on player inputs and each others' outputs over time",
- the *"Fun"* becomes ***Aesthetics***, with this term the authors "describes the desirable emotional responses evoked in the player, when she interacts with the game system".

Thus, MDA (Mechanics, Dynamics, and Aesthetics) has the main goal to study the games providing a common approach to different game actors: designers, developers, critics, and researchers. Lazzaro [27], through the observation of subjects playing at their favorite video games, identified 4 elements able to arouse emotions without an explicit narrative:

- *Player*, that is focused on player's feeling during and after the game session. This point is also divided into four different sub-elements that describe the players' emotional reactions to four

[1]`https://www.merriam-webster.com/dictionary/game`

games aspects: *Visceral*, which describes the reaction to the game environment, *Behavioral*, which describes the reaction to the product interaction, *Cognitive*, which describes the reactions to the ideas, memories, and association, and *Social*, which describes the reaction with the interaction with other players

- *Hard Fun*, that involves the game challenge. The challenge should be balanced with the player skill and the game difficulty. The author has listed a set of hard fun features able to engage the players: adjusting the level of difficulty, amount of commitment required, progress and feedback, development of skills and feelings from growth, modifying games, custom rules and messing around, and having choice between multiple strategies, skills, and goals

- *Easy Fun*, that involves the player in game activity. It provides the sensation of immersion in a virtual world. In order to improve this feature, the game designers have to consider the different types of interactions between the player and the environment.

- *Other Players*, that involves the cooperation or competition between players. This factor is able to intensify the players' emotions. During her study, the author has observed that people playing in groups demonstrate more signs of emotions than the "single-players".

Freeman [28] listed 33 game design techniques useful to elicit players' emotions during the games' fruition, summarizing these techniques with the term "*Emotioneering*". In contrast to the Lazzaro's approach, the author focused his attention on the narrative aspects of the game, proposing a set of methods in order to develop a game story (e.g., plot, dialogues) conceptually interesting and exciting. Koster published a popular book [15] which places the players' emotions at the center of the game experience. An interesting statement extracted by his book is: "the destiny of games is to become boring, not to be fun". It is because the enjoyment in video games persists until the player has the feeling to learn something new that helps her to master the game mechanics. Salen and Zimmerman [22] developed the concept of *Magic Circle* starting from the ideas developed by Huizinga [1]. The *Magic Circle* defines the boundaries of a game, i.e. a frame, limited in space and time, in which the player plays at a game. In the circle, the game rules develop a special meaning of the environments and the behaviors, since a new reality is created. The authors develop also the concept of "*Lusory Attitude*" in which the player accepts to play in an environment with a limited set of rules in order to experience the pleasure provided by the game. Yannakakis and Paiva [29] described the games as "*Emotion Elicitors*", since the emotions can be elicited through the interaction with the game elements. They define two different clusters able to arouse emotions in player: *game content*, and *game Non-Player Characters (NPCs)*. In the *game content*, the authors considered the game environment, the game mechanics, story plot points, and reward system. These elements can be further divided into more specifics elements, like, e.g., the audio/visual settings. The NPCs, if they are implemented in the game, can be used as triggers to arouse the desired emotions in the player. If an NPC has a credible behavior, it may arouse emotional reactions if something good or bad happens. In addition, the NPC may have an impact on both game environment and the virtual relationship. For example, if in a video game, for narrative choices, a team member dies, the player loses the virtual relationship with the NPC and the support of a team member (e.g., the forced decision of which crew member to sacrifice in Mass Effect).

Different studies also proposed to understand the types of sensations elicited by these entertainments products. These researches are based mainly on two approaches: *model-based*, and *model-free* [30]. The first uses a top-down approach, building the model on a theoretical framework which maps the players' affect. For example, a model can be based on game emotion theory ([27, 15, 22]), or on a general emotion theory ([31]), such as the definition of a set of parameters which describes the players' behavior. Modeling the players' characteristics allows to understand their behavior and, consequentially, to suppose the emotions elicited by a specific video game feature through simulations. In [32, 33] the researchers studied the players' satisfaction regarding the game rewards (Looting System) in Massive Multiplayer Online (MMO) and Multiplayer Online Battle Arena (MOBA) games, through a set of simulations where different agents have peculiar features (like, e.g., Bartle type [34], time usually spent playing video games, etc.). The second approach (bottom-up) infers the players' emotions mapping them without the use of any particular theoretical framework. This type of players' analysis can be performed in two ways: studying the player's behavior in a video game, or using her physiological information to infer her emotions. The former is usually considered in several moments during the game development, and sometimes also after the game release to the market. This procedure is called *playtest*, and it has the main purpose to have an overview of "the entire design process to gain an insight into whether the game is achieving your player experience goals" [23]. Moreover, the *playtest* is usually used to evaluate the player behavior in specific environments, and to find issues or lacks in the game levels. Usually, it is carried out involving a set of participants (*Game Testers*), or using a software-based approach such as AI-driven bots which mimic the humans behavior (e.g., [35]). Playtesting is widely used in the industry, but, usually, it is not useful to extract quantitative data about the players. Instead, using directly the annotated emotional physiological information of the player may provide to the researchers quantitative results based on players' physiological conditions. An example of study which uses the physiological information in order to improve the quality of the gameplay information can be found in [36]. Unfortunately, it does not provide any direct information about the participants' emotions elicited by the video game fruition. Further details on the methods used to annotate the affective states will be explained in the next section.

2.2 What Our Body Says

Emotions arise spontaneously in humans, and they may modify our decisions and actions. The first studies on emotions are mainly focused across the nineteenth and twentieth century. In 1872, Darwin published a book [37] which reports the results of his research on emotions. He stated that the emotions of animals and humans are "*homologues*". To support this hypothesis, he compared photographs of animals and humans during the experience of different emotional states. The author, also, hypothesized the existence of a set of universal observable emotions across culture and species. After some years, James [38] proposed a new theory about emotions. He stated that they are an outcome of physiological changes. Thus, the author assumed that a solicited sensory system sends information to the brain, which defines the appropriate emotion and, in turn, it sends signals to the whole body in order to induce the correct reaction. In parallel, Lange proposed a similar idea [39], and, as a consequence, the theory is widely known with the name of "*James-Lange theory*". After some years, Cannon published in a paper [40] some criticisms to the work of James and Lange. These criticisms were supported and partially corrected by the Cannon's doctoral student, Bard [41]. In order to prove his theory, Cannon inhibited the sensory system (destroying the sympathetic nervous system) of a set of animals, and he observed their emotional reactions. He discovered that the removal of this body function has only a little effect on the animals' emotional response. Thus, the authors formulated a new theory, known today as "*Cannon-Bard theory*". This theory states that when a stimulus was perceived, the information is directly communicated to the brain (involving, in particular, the hypothalamus), which provides the physiological reaction, and, as a consequence, the emotional experience. In more recent years, Ekman [42] conducted an empirical research, and he defined six basic and universal human emotions that can be revealed on human face: *anger*, *fear*, *sadness*, *happiness*, *disgust*, and *surprise*, validating part of the Darwin theory. These emotions have the same intrinsic meaning and physiological outcome across the different cultures. Moreover, the author defined a new set of 11 possible universal emotions, but he has also stated that "the evidence [of these emotions] is certainly not available now" [43].

Dalgleish [44] presented an overview of these and more theories on the relation between emotions and physiological conditions.

Summarizing, different researches proved that there is a connection between emotions and physiological changes in the human body. Thus, the emotional reactions can be described and acquired by at least three human output systems [45]: *self-report* measures (e.g., through verbal expressions), *behaviors* (e.g., facial expressions), and physiological reactions of Autonomic Nervous System (ANS), like, e.g., Heart Rate (HR), Brain Activity, etc. Hence, if emotions have a physiological response, then they may be identified through a set of tools able to detect the variations in human physiology.

The *self-report* is a direct measure of the emotional states provided directly by the subject. It can be reported mainly in three ways: asking the users to report the nature of their experience with a "*free-response*", considering a set of *discrete* emotions using, for example, the Ekman's universal emotions [42] (i.e., happiness, fear, etc.), and through a *dimensional* model using two or three vectors [46]. Using the *free-response* format, the researchers can provide a comfortable environment to the participants [47]. With this method, the users can report their experienced emotion using each type of label or any type of expression able to better describe what they have felt. However, many participants may have a problem to use the appropriate label to identify the emotions, since they normally do not communicate their emotional states. In addition, it is not possible to study this type of emotion self-assessment using a quantitative analysis, since the frequency of some labels is quite low. In order to reduce this problem, some researchers sorted the free-response labels in a limited number of categories. For example, Geneva Affect Label Coder (GALC)[2] identified 36 affective states by parsing the text and looking for specific words (or their synonyms). The *discrete* self-assessment of the emotions uses human language in order to define a set of markers able to describe, and clearly separate, the emotions. Thus, this type of categorization is similar to the semantic organization of the emotions available in natural languages (with unique patterns able to identify a specific feeling). The main approaches used to identify the discrete emotions are [47]: **Nominal Scale**, it presents a set of terms that describe the experienced emotion, **Ordinal Scale**, which indicates in a Likert scale the intensity of a specific emotion, and/or **Interval Scale**, which uses an analog scale to indicate the level of experience regarding an emotion. In order to standardize the list of discrete emotion labels commonly used in the affective studies, Izard presented in the book "The Psychology of Emotions" [48] the *Differential Emotion Scale*. It listed 10 standardized emotions (joy, surprise, anger, disgust, contempt, shame,

[2] https://www.affective-sciences.org/home/research/materials-and-online-research/research-material/

guilt, fear, interest, and sadness) described through a checklist structured by thirty-item adjectives, each one evaluated in a 5-point scale. The last self-report method is the *dimensional* model. It is structured by a 2D or 3D model, where each point in the dimensional space represents a particular mental state. Usually, the considered vectors are Pleasure(Valence)-Arousal-Dominance (PAD), and VA, where:

- *Valence* measures the emotion "quality" (from *averseness* to *attractiveness*). It defines how pleasant (or unpleasant) is the participant feeling regarding an event
- *Arousal* measures the emotion "energy" (from *very calm* to *very excited*). It defines the level of the participant physiological activation in front of a particular event
- *Dominance* measures the user "potency" (from *submissive* to *dominant*), which represents the amount of influence a user feels in a specific environment.

However, also different vectors were applied, such as the *motivation*, which describes the affect associated to approach or avoid a stimulus [49]. An example of other affective states used in the dimensional model are *tension*, and *control*, which usually replace the *Dominance* in the PAD [47]. As a consequence, the different intersections of these dimensions are able to provide a very fine classification of emotions to the researchers. Moreover, different works tried to map the discrete emotions in the dimensional model (e.g., [50]). An estimation of discrete emotions, mapped in the VA 2-dimensional space, can be see in Fig. 2.1.

In order to provide a greater accuracy in the dimensional emotion identification, some researches were addressed to develop markers able to support the users in emotions labeling. The most common method used in different research is the Self-Assesment Manikins (SAM) [52]. This tool is structured by a series of anthropomorphic figures representing different human emotions. The authors designed these manikins in order to support the mapping of the PAD value. Moreover, the authors provided, for each vector, 3 set of figures (with 5, 7, 9 manikins), in order to cover almost all types of approaches used for the emotion self-assessment. A different tool was developed by Betella and Verschure [53]. They developed the Affective Slider (AS) as a valid alternative to the SAM. AS is structured by a set of emoticons that represent the emotions limits of PAD vectors. It also uses a bow-tie graph where the narrow area indicates a neutral emotion. Through their research, the authors showed that AS can be a valid alternative to SAM.

Each self-report measure is reliable only if the time spent between the measure and the emotional experience is short (i.e., few hours) [54].

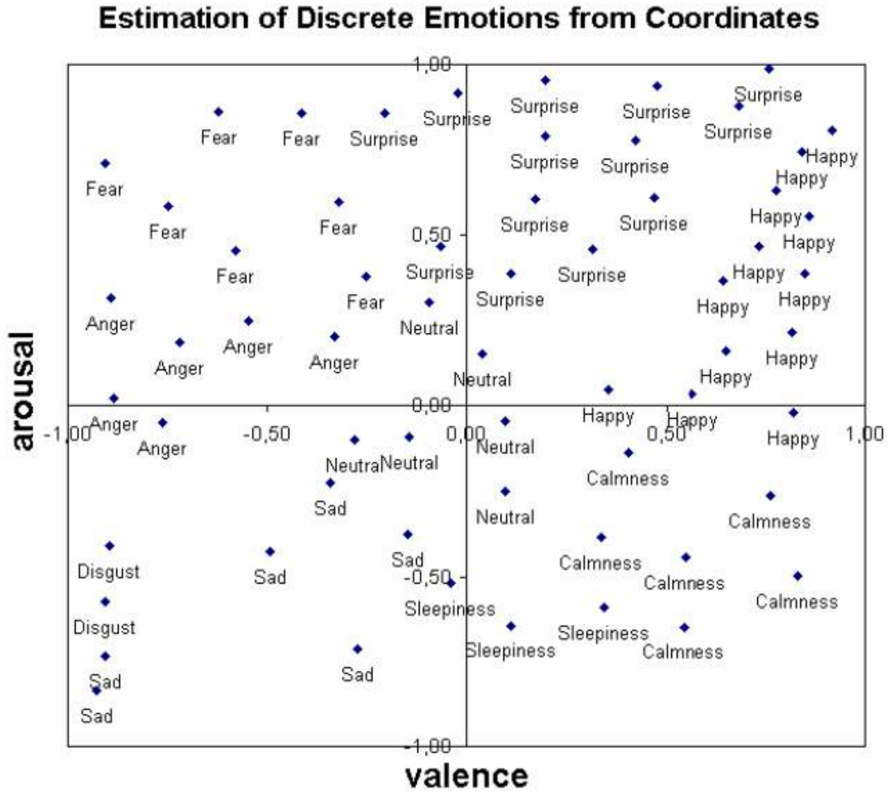


FIGURE 2.1: Estimation of Discrete Emotion in VA dimensional space. The image is taken from [51]

.

The measure of emotions defined by the *behavior* is mainly described by three components: vocal features, body movements, and facial expressions. This measure describes how humans communicate with each other their emotional states. The speech contains, besides a semantic content, some information related to the tone. It has different features (e.g., amplitude and pitch) that can be used to infer the speaker's emotion. An overview of speech datasets, and on the techniques commonly used to analyze emotional speech signals can be found in [55]. The study of the body posture received less attention than the analysis of the others human behaviors. However, some researches are related the study on human posture and the link to specific emotional states. For example, a work of Tracy et al. [56] demonstrated that pride has a universal body outcome, i.e., a visibly expanded posture, and arms raised above the head or hands on hips. The last body outcome is the facial expressions. Albeit Ekman's studies have found only a limited number of universal facial expressions able to represent a corresponding number of discrete emotions, the information acquired on human face can support the identification of

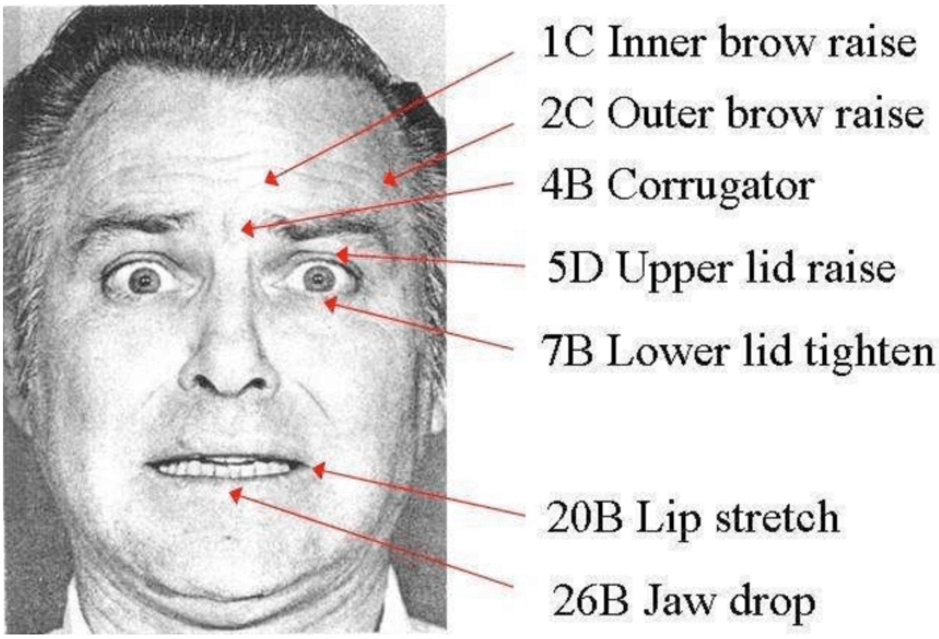


FIGURE 2.2: Example of facial decomposition based on FACS. The image is taken from [60]

the emotions in the dimensional model. The facial behavior seems to have a strong connection with the emotion valence [57]. For example, the contraction of the muscles that surround the eye is usually linked with positive emotions [58], while when the eyebrows are lowered and closed together, it reports, commonly, negative emotions [45]. In order to define a universal methodology to study the facial expressions, Ekman et al. [59] developed a coding system, named "Facial Action Coding System (FACS)", which assesses 44 movements of facial muscles that can be observed on a human face (see Fig. 2.2).

Thus, more emotional information can be revealed on facial behaviors through an Electromyography (EMG) analysis. Usually, the muscles most involved during the EMG studies are the *Corrugator Supercilii* and the *Zygomatic* (see, e.g., [61]) which provide an approximation of the overall facial expression.

Furthermore, the periventricular area of the hypothalamus controls the *ANS*. It is a system which provides coordinated and autonomous functions at the various organs. The ANS is commonly divided into two systems, the sympathetic and parasympathetic divisions. The former manages the stressful situations (e.g., increasing the HR and the Blood Volume Pressure (BVP), etc.), while the latter manages the calm situations. For example, two physiological outputs produced by ANS that can be acquired with non-invasive techniques through different types of sensors are: the Electrodermal Activity (EDA), and the *cardiovascular response*, usually acquired through an ECG.

EDA is mainly measured as the Skin Conductance (SC) or its inverse quantity, the skin resistance Skin Resistance (SR). Until the 1980s, the common term of skin electrical measurements was Galvanic Skin Response (GSR) [62], thus, in this paper, we use these names alternately. GSR can be considered as a reflection of the sympathetic system, it measures the electrical characteristics of the epidermis, mainly altered by the sweating of the skin, and it is physically interpreted as conductance [63]. Sweat glands, distributed on the skin, receive input only by the sympathetic nervous system, and, as a consequence, sweat is a good indicator of arousal [64]. Its measure can be quantified in terms of: *raw* data (i.e., EDA/SC/GSR/SR), Skin Conductance Level (SCL) or Skin Resistance Level (SRL), and Skin Conductance Response (SCR) or Skin Resistance Response (SRR). The measure of the *Level* provides the tonic information of EDA, which defines information not related to a specific and immediate stimulus, but it provides average information on slow ANS changes (over a second). On the other side, the *Response* is a phasic measure which provides information on the specific and immediate stimulus [62]. Summarizing, the sympathetic activity is linked to the emotions and, therefore, GSR is often suggested as emotions index [65]. The ANS has also the functions to innervate the heart and blood vessels. These functions provide alterations on HR, as well in other related physiological information: BVP, Total Peripheral Resistance (TPR), Cardiac Output (CO), Pre-Ejection Period (PEP), and Heart Rate Variability (HRV) [66]. Other patterns managed by ANS, which can support emotions recognition, are: Respiration [67], Skin Temperature [68], and Pupils Dilation [69]. Also, these physiological outcomes can be observed through a set of non-invasive sensors. For a full list of the ANS physiological information solicited by emotions and that can be acquired through non-invasive techniques see [70].

The last type of physiological information, usually acquired to infer emotional states, is the electrical activity of the brain, which is usually measured through the Electroencephalography (EEG). This kind of measure is categorized in the Central Nervous System (CNS), and in particular, the non-invasive techniques have been addressed to acquire the electrical activity of the neocortex. Empirical experiments revealed a particular behavior of brain electrical activity, named Event-Related Potential, which generates an action potential (i.e., a positive signal deflection) after around 300 milliseconds from the presentation of a stimulus [71]. Commonly, this characteristic is used as a diagnostic tool (in both psychiatry and neurology), as well as for the Brain-Computer Interface (BCI) research field [72], where the passive BCI is a common measure to infer the user's cognitive state. The BCI analysis addressed to study the humans' affective state is named aBCI, its approach is based on stimulus-independent passive BCI, which includes general affect sensing for

Human-Computer Interaction (HCI) scenarios, inferring the emotional states and adapting the environment in order to engage better the users [73]. Other common techniques used to measure the brain activity are: Positron emission tomography (PET), Functional magnetic resonance imaging (fMRI), and Magnetoencephalography (MEG). Albeit these methods are usually more informative than EEG, they are also more invasive, more expensive, and with a greater demand in terms of space.

Part of the content of this section were presented in UBIO workshop and it will be published in " Granato, M., Gadia, D., Maggiorini, D., and Ripamonti, L.A., "Feature Extraction and Selection for Real-Time Emotion Recognition in Video Games Players", *Proceedings of International Workshop on Ubiquitous implicit BIOmetrics and health signals monitoring for person-centric applications, IEEE*" [74].

Chapter 3

State of Art

THE aim of this chapter is to provide a review of the literature. In particular, in Sec. 3.1 we compare different affective datasets, describing the types of stimuli, the experimental procedure, and the eventual data analysis outcomes. The Sec. 3.2 aims to describe researches that have used the players' physiological data in video game research field. Thus, we describe studies addressed to understand the players' emotions during the game fruition, and studies aimed at the developing of video games which use the player mental state as an additional input.

In Tab. 3.2, we provide a comparison between part of these datasets. Thus, we compare the number of participants, the type of stimuli used for the experiment, the collected physiological information, and the type of emotion assessment.

3.1 Affective Studies and Datasets

Many Affective Computing researches were addressed to recognize the human emotions, and to develop a computer behavior which simulates, as closely as possible, their expression. In this section, we focus on the emotion recognition starting from data acquired on humans and on the description of different data sets developed for academic purposes.

Usually, affective datasets contain different types of human output information (e.g., video of humans' face, voice, etc.). These databases can be unimodal, designed to collect a unique type of data (e.g., images with different facial expressions), or multimodal, which provides different types of information (e.g., video and audio). Below, we describe a set of affective datasets which provide, at least, a physiological information generated by ANS. These datasets were designed to be studied in the emotion recognition research field. Thus, they also contain an emotional target variable used by the researchers to estimate the accuracy of their model (i.e., annotation).

3.1.1 Multivariate response patterning of fear and anger

Sinha and Parsons in 1996 [75] were, to the best of our knowledge, the first researches which collected an affective dataset using physiological information. In particular, they acquired: ECG, BVP, GSR, facial EMG, skin temperature (acquired on a finger), and eye movements. In order to collect the eye movement, the authors placed two electrodes (GRASS Gold Cup) near to the right eyeball: one below and the other one on the lateral canthus. They also acquired the EMG signals of the electrical activity of 4 facial muscles: zygomatic, corrugator, masseter, and depressor. The experiments were conducted on 27 participants, each volunteer was male with an age between 21 and 35. The participants were asked to listen a set of auditory scripts which suggest imagining different kinds of scenarios. The imagery trials were divided in mental trains able to arouse six different emotions, and one physical activity (a basket match). After the experimental session, the authors asked the participants to assess their emotions, during each auditory script, using a short version of the *Different Emotion Scale* [76]. Thus, the authors classified the physiological signals according to the scenarios of fear, anger, and neutral, reaching an accuracy of 98.8%. Unfortunately, the dataset was not released to the academic community.

3.1.2 Eight-Emotion Sentics Data

In 2001, Picard et. al [77, 78] collected a free and public dataset[1] named *Eight-Emotion Sentics Data*. *Sentic* is a term used to refer to the emotion categorization, which "basically means understanding signals that convey emotions" [79]. The researchers considered 4 physiological information (i.e., BVP, GSR, respiration, and a facial EMG along the masseter), involving a single participant in an experiment in a 20-day long experiment. Each day, the subject had to participate to an experimental session, listening for 25 minutes to a set of auditory stimuli. During the experiment, the participant listened a set of auditory stimuli. In particular, through a headphone a random sequence of emotions was announced, and each element of the sequence was followed by a series of soft metronome clicks. At each click, the participant had to press a finger on a sensor placed in front of him in order to evoke the specific emotions through physical expression (see [80]). Also, in this case, the emotions were elicited through an imagery technique, where the participant imagines different scenarios. The considered emotions were: neutral, anger, hate, grief, love, romantic love, joy, and reverence. The authors reached the 81% of accuracy in the classification of the emotion recognition over the eight discrete emotions.

[1] `https://affect.media.mit.edu/share-data.php`

3.1.3 Wearable Sensors to Recognize Humans Emotions

Lisetti et. al [81] collected, in 2004, a physiological dataset using a wearable sensor (BodyMedia SenseWear Armband[2]). With this device, the authors were able to acquire the following data: GSR, HR, and skin temperature. Thus, they used a set of movie clips and mathematics problems in order to arouse, in the participants, the emotions of: sadness, anger, surprise, fear, frustration, and amusement. A preliminary experiment was conducted on 14 students, which have participated simultaneously in the study. They watched 14 movie clips and they reported, for each clip: the experienced emotions (in the discrete subset), the emotional intensity, the concurrence of different emotions, and a brief description of the video clip. This preliminary study was conducted in order to select the movie clips able to elicit a particular emotion. Thus, the clips that elicited the correct emotion in at least the 90% of the participants, with an average intensity greater than 3.5, were considered. The main experiment was conducted on 29 participants, which participated simultaneously in groups of 1 to 3. A set of slide-show with images of naturalistic landscapes were shown to the participants in order to acquire a baseline of physiological data. After the slides, the researches showed to the subjects the following sequence, repeated for each clip: the video of a selected movie clip which is able to arouse a specific emotion, a slide which asks the participants to answer some questions about the clips (the questions were similar to the pilot study), and a slide which asks the participants to relax with some music. Thus, the authors classified the emotions using three different ML algorithms, achieving an accuracy between 72.3% and 84.1%. Unfortunately, the authors did not make the data available to the scientific community.

3.1.4 Emotion Recognition During Music Listening

In 2008, Kim and André [82], acquired an affective dataset on 3 participants with an age between 25 and 38 years old. The authors collected the EMG on the upper trapezius muscle, ECG, GSR, and respiration signals using the Procomp Infinity device, produced by MindMedia[3]. They used as stimuli 4 songs, each one selected directly by the participants in order to evoke a set of emotional memories: positive/high arousal, negative/high arousal, negative/low arousal, and positive/low arousal. The selection of the songs was carried bout each participant, because the emotional responses to music may vary between people, due to past experiences and culture. Thus, the authors used ML techniques in order to classify the physiological data in the four types of emotional memory (using the leave-one-out CV). The average accuracy of data prediction on

[2]The company have been acquired in April 2013, and, unfortunately, the product is no longer purchasable

[3]https://www.mindmedia.com/

the 3 participants, considering the individual analysis, was 87%, while the accuracy of the model trained on all the subjects' data was 65%. Unfortunately, the authors did not make the data available to the scientific community.

3.1.5 HCI-Tagging Database

The *MAHNOB HCI-TAGGING* is a multimodal database [83] that was acquired in 2012. It provides a set of physiological signals and the emotion self-assessment on 27 participants. The biofeedback measures include eye gaze, ECG, EEG (on 32 channels), GSR, respiration, and skin temperature. Moreover, the authors recorded the participants' voice with a microphone, and a video of the participants through 6 video cameras. As described in the paper, the authors took special care of the data synchronization. The experiment was structured in two parts: *Emotion Recognition Experiment*, and *Implicit Tagging Experiment*. In the former, the participants watched a neutral video clip, followed by a clip selected by the authors (in random order). 20 movie clips (with an average length of 81.4 seconds) were selected for this experiment. The authors acquired these videos with two methodologies: 6 videos have been defined manually and 14 through a collective and web-based preliminary study [84]. Each video had an emotion annotation in arousal and valence vectors, provided by more than 50 participants. After each stimulus, the participant had to complete a set of 5 questions: the first to identify the experienced emotion in a discrete interval (neutral, anxiety, amusement, sadness, joy, disgust, anger, surprise, and fear), the others questions were designed to identify, in a 9 point scale, the information of PAD, and predictability [85]. The authors applied a data classification on VA values, collecting the results presented in Tab. 3.1.

Table 3.1: Manhob-HCI Classification Results

	Recognition Rate	
Trainig Variables	**Arousal**	**Valence**
Physiological Data	46.2%	45.5%
EEG	52.4%	57.0%
Eye Gaze	63.5%	68.8%
Eye Gaze + EEG	67.7%	76.1%

In *Implicit Tagging Experiment*, images or video fragments were shown to 27 participants. These stimuli present a tag at the bottom of the video, which describes the presented situation. The tag can be correct or incorrect, and the participant had the task to select a correct button according to his opinion, i.e., if she agreed with the image/video description. Thus, the authors classified the participants' facial expression and the eye gaze according to the correctness of the displayed tags in image stimuli. The classification accuracy obtained a result of 75% with a fusion modality (facial expression + eye gaze). Additional details on methods, analysis, and results can be found in [86]. The dataset is freely available to the academic community[4].

3.1.6 DEAP

DEAP [87] is a multimodal affective dataset which provides a 32 channels EEG information, GSR, BVP, ECG, EMG on Zygomaticus and Trapezius muscles, respiration, skin temperature, and Electrooculogram (EOG). The database is based on a pilot study [88] which involved 6 participants and 20 music videos as stimuli, while *DEAP* has considered 40 music videos and 33 participants. The faces of a subset (22) of participants were also recorded during the overall experiment. In order to select the 40 music videos, the authors had initially selected 120 different stimuli, with half of them selected automatically. Thus, through an affective highlight algorithm, the authors had extracted a one-minute video for each stimulus. As the last step, the authors had selected the final 40 stimuli through a web-based subjective assessment experiment. Each experiment had followed the same procedure, structured, for each video, in 4 different steps: 2 seconds screen which informs the participant of her progress, 5 seconds baseline recording, 1 minute of the video clip, and the self-assessment stage. The latter was acquired on 4 different values: PAD, and liking. In order to acquire the PAD, the researchers used the SAM [52] with a Likert scale on 9 levels. For the liking scale, the authors used three thumb figures: up, neutral, and down. In order to describe the meaning of the valence information to the participants, the authors defined this measure as a report of "the participants' tastes, not their feelings. For example, it is possible to like videos that make one feel sad or angry". Thus, the authors computed a classification between a set of features extracted by the EEG and the scales of arousal, valence, and liking, obtaining an accuracy of, respectively, 62%, 57.6%, and 55.4%. The dataset is freely available to the academic community[5].

3.1.7 RECOLA

In 2013, another multimodal database was acquired. Its name is *RECOLA* [89] and it provides audio, video, and physiological (ECG and GSR) information. It involved 46 different participants, however, only on 18 of them the authors had acquired the information for the dataset. The database was addressed to investigate the humans emotions in collaborative work. Thus, the participants were divided into two separate rooms and they had to complete a survey in order to

[4] https://mahnob-db.eu/hci-tagging/
[5] http://www.eecs.qmul.ac.uk/mmv/datasets/deap/

evaluate their emotional state. After an individual task, the participants were engaged in a remote discussion. The emotion self-assessment were acquired on around 3.8 hours of audiovisual data, and 2.9 hours with physiological data. To the best of our knowledge, this database is the first to assess the emotions in a continuous time. The emotion was self-assessed, and identified by 6 external annotators on two affective vectors (VA), the annotation values ranging from -1 to +1, with a step of 0.01. The dataset is freely available to the academic community[6].

3.1.8 DECAF

DECAF [90] is a multimodal dataset acquired in 2015. It provides the physiological information of 30 participants during the fruition of 36 movie clips. The clips were selected through a preliminary stage, where 42 volunteers watched 58 movie segments, and they self-assessed their emotional state on valence, arousal, and a set of discrete emotions. Thus, according to the acquired data, they removed outliers movie clips. Moreover, the authors collected the emotional response on 40 music videos already used in *DEAP* [87]. The authors' main goal was to compare different types of stimuli in order to investigate their effectiveness to elicit similar emotions. The acquired physiological data are: Magnetoencephalogram (MEG), EOG, ECG, a EMG on the trapezius muscle, and an infra-red facial video. Each participant completed two separate experimental sessions: in the first, the movie clips selected by the authors were presented, while, in the second, the subject watched the music videos. The videos were shown in random order, taking care that two clips with similar VA values did not follow one another. The emotion self-assessment were acquired through a microphone, the participants had to rate (in a range from 0 to 4) the values of PAD. The authors declared that "*MEG signals are seen to effectively encode arousal and dominance, while peripheral physiology signals efficiently encode valence. Facial expressions are also seen to best encode valence, while audio-visual features achieve best arousal recognition for music clips with PB[7] labels*". The dataset is freely available to the academic community[8]

3.1.9 OPEN_EMOREC_II

In 2015, Rukavina et al. acquired a dataset named *OPEN_EmoRec_II* [61]. It is an evolution of a previous pilot work (see [91]). In order to collect the dataset, the authors collected the video, audio, and a set of physiological data on 30 participants. The biofeedback information acquired for *OPEN_EmoRec_II* are: ECG, BVP, GSR, facial EMG (on zygomaticus, and corrugator supercilii

[6]https://diuf.unifr.ch/diva/recola/download.html
[7]The authors have used the acronym PB in order to indicate the stimuli label provided by Population-Based.
[8]http://mhug.disi.unitn.it/wp-content/DECAF/DECAF.html

muscles), and respiration. The experiment was structured in two separate sessions. In the first session, a set of pictures, acquired from IAPS [92] database, were showed to the participants. A stimulus was composed by a set of 10 pictures (each one has been displayed for 2 seconds) with similar emotional ratings. For each participant, the authors showed 10 sets of pictures, 2 for each of the 5 considered affective states. The second part of the experiment aimed to investigate the affective state in HCI. The subjects were informed that they had to interact with a virtual environment controlled by a computer, able to receive users' natural inputs (i.e., natural speech). Actually, the computer was controlled by an experimenter located in another room (Wizard of Oz experiment). Thus, the participant was proposed to solve 6 task of mental training. Each task, described in [91], was designed to induce different emotions through the computer feedback (i.e., the operator in the other room). At the end of the experiment, the participants rated each task on PAD. Moreover, 4 external annotators provided, for each participant, the information of valence and "*intensity*, which is useful as it is identical to the dimensional approach like the whole emotion induction" [61]. The annotators also provided a label for the facial reaction after positive/negative events. Unfortunately, the authors had not provided a data analysis of the acquired data. In addition, albeit the dataset is available for research groups, unfortunately, the website which hosts the data is no longer available. However, the authors provided, in the paper, an email in order to provide technical support.

3.1.10 AMHUSE

AMHUSE [93] is a multimodal and affective dataset collected in 2017. It involved 36 participants with an age between 18 to 54 years old. Four videos were showed to each subject, 1 neutral, and 3 videos with comic contents. The comic videos were selected to induce amusement in participants. During the stimuli fruition, 2 videos, with RGB and Depth cameras, and a set of physiological data were collected. All biofeedback signals were acquired on the fingers of the left hand: on the index and the middle were placed the electrodes for the GSR sensor, on the ring finger was acquired the BVP, and on the pinke a was placed thermometer to acquire the skin temperature. At the end of each stimulus, the participants annotated the levels of PAD using the AffectButton tool [94]. This tool allowed the users to intuitively self-asses their emotions in a range between [-1,1]. Moreover, the emotions were identified by 4 external annotators on VA vectors. Unfortunately, the authors had not provided a data classification/regression of the acquired data. The dataset is freely available to the academic community[9].

[9] `http://amhuse.phuselab.di.unimi.it`

3.1.11 Video Games Datasets that Use Physiological Data

To the best of our knowledge, there are not freely available datasets which consider the relationship between a dimensional model and the video games as stimuli. However, two studies investigated and collected datasets on the relationship between specific game mechanics or game events, and the physiological information and the players' mental state. Unfortunately, these datasets use different approaches to assess the emotion rather than the dimensional model.

In 2010, Yannakakis et al. [95] recorded a dataset of physiological information in order to investigate the players' experience under different camera perspectives. Thus, the authors developed a 3D game (named Maze-Ball) with mechanics similar to *Pac-Man*, and they collected the physiological information of HR ,GSR and BVP on 36 participants. During the experiment, different in-game camera conditions were tested and the participants had to choose which one provides a better user experience considering the affective state of fun, challenge, boreness, frustration, excitation, anxiety, and relax. The authors reached a classification accuracy above the 80% on the majority of the above mentioned affective state. The dataset is freely available[10].

In 2015, Karpouzis et al. collected the *Platformer Experience Dataset* (PED). They acquired the participants outcomes using only a HD camera. Thus, they collected the video of 58 players playing at to play at 2 automatic generated level of *Infinite Mario Bros* (IMB), a clone of the Nintendo *Super Mario Bros*. The authors marked the game events in order to permit the researchers to investigate the visual behavior in relation to different game states. Moreover, the players rated in a range between 0 and 4, in both game levels, the values of: engagement, frustration, and challenge. Different studies were applied to the dataset, for example, regarding the player experience [96], or the study of the spontaneous movements according to "how the level ends" (i.e., if the player finishes the level successfully or not) [97]. The dataset is freely available[11].

3.2 Affective Computing in Video Games

One of the main argument at the basis of the Affective Computing research field [16] suggests that any computer has the ability to express and recognize the people affects. In video games, Affective Computing is relevant in three main aspects: *players emotions recognition* - which leads to the game response to the emotions -, generation of "*affective behaviors*" in the game characters to enchant the realism with a credible output to various game events, and "*modelization*" of the

[10] `http://www.hectorpmartinez.com/`
[11] `http://ped.institutedigitalgames.com/`

TABLE 3.2:
Comparison among available affective datasets. The physiological signals considered are: EEG = electroencephalography (with the number of channels), ECG = electrocardiogram, BVP = blood volume pulse, GSR = galvanic skin response, Facial EMG = electromyography placed on participant face (with the number of muscles considered), Resp = respiration, Temp = temperature, Gaze = eye gaze tracking. The last 3 columns define the type of emotion identification, where: ES = Emotion Space, in type column D = Discrete, C = Continuous, and in Annotator column S = Self Report, E = External Report with, in brackets, the number of annotators.

Dataset	Stimulus	Subj.	EEG	ECG	BVP	GSR	Facial EMG	Resp	Temp	Gaze	ES	Type	Annotator
Eight-Emotion Sentics Data [77]	Sentic [80]	1	-	-	✓	✓	1	✓	-	-	8-D emot.	D	S
MAHNOB-HCI [83]	Video, Images	30	32 Ch.	✓	-	✓	-	✓	✓	✓	PAD	D	S
DEAP [87]°	Music Video	33	32 Ch.	✓	✓	✓	1	✓	✓	-	PAD + Liking	D	S
RECOLA [89]	Collaborative Work	16	-	✓	-	✓	-	-	-	-	VA	C	6(E) + S
DECAF [90]*	Movie Clips/Music Video	30	-	✓	-	-	-	-	-	-	PAD	D	S
OPEN_EmoRec_II [61]	Mental Puzzles	30	-	-	✓	✓	2	✓	-	-	VA	D	4(E) + S
AMHUSE [93]	Movie Clips	36	-	-	✓	✓	-	-	✓	-	VA	C + D	4(E) + S

° *The dataset provides also the signals of: EMG on Trapezius muscle, and EOG to investigate the eye movements*

* *The dataset provides also the signals of: EMG on Trapezius muscle, MEG in order to measure the brain activity, and EOG to investigate the eye movements*

emotions' generations on the game characters in order to represent a believable physical reaction (e.g. facial expressions) [98]. The possibilities of recognizing emotions, through physiological information, during the video games fruition, allows the researchers and developers to investigate what kind of game events can generate a specific emotion and to design games aimed at inducing specific emotions in the players.

A common way to infer the players' emotions is to consider their in-game behavior and the physiological information. The latter can be used to identify the user's emotions ([99]), or to provide an input to a software or a device (e.g., [100]). In the following sections, we analyze different case studies considering two aspects: studies that have tried to understand the players' emotions, and work that have used physiological information as input for a video game. Other surveys on the relation between Affective Computing and video games can be found in [101, 102].

3.2.1 Inferring Emotions in Video Games

In the video game research field, the main methods used to infer the players' emotions are mainly two: the *gameplay*, and the *physiological information*.

The gameplay considers the player behavior inside the virtual world. The player provides a set of input and decisions basing on her expertise in the game. This cognitive process may alter the emotions, and, as a consequence, may influence the type of interaction within the game. This approach can use different types of measures, according to the kind of game. For example, the researchers can evaluate the time spent on a task, or the selected weapon. These decisions may support the researchers to infer the players' emotion in different game stages [29]. However, this type of metric is quite sensitive to the type of players, since the players may have different approaches to the game. Thus, it is mainly reliable with game prototypes designed to study specifics effects of a subset of mechanics or level areas.

In contrast, using physiological information may provide a more reliable feedback also in commercial games. This approach can be used to explore the relationship between the physiological signals and the gameplay experience [103, 104, 105], as well to study the elicited emotional response.

Hazlett [106] applied, during a racing game session, two EMG sensors on the face of 13 teenage players. Through this study, the author has shown that the Zygomaticus muscle is more involved during positive events, while the Corrugator muscle is involved in the negative ones.

Tognetti et al. [107] used an open-source racing game (TORCS) in order to understand the players' preferences under different game environment. They performed different experiments with

FIGURE 3.1: An example of screenshot acquired during a *biometric-based* study conducted during the research presented in [108]

a set of participants, recording the following physiological signals: ECG, GSR, respiration, and temperature. Thus, the players played at the same game 6 times, 2 for each customized opponent, where: (W) the opponent keeps a distance of +100 meters, (C) it drives with a skill similar to the player, and (L) it keeps a distance of -100 meters. The participants played the levels in a specific order (CLWLCW), and, after each level, they provided the information of preference between the current and previous level. Thus, they classified the data through a Linear Discriminant Analysis, and they reached an accuracy of 74%.

Mirza-Babei et al. [108] compared two methods useful to provide feedback to the developers, and often used in GUR: *observation-based*, and *biometric-based*. Thus, the authors selected two *First Person Shooters* with a different user experience quality, considering the Metacritic score as index: "Call of Duty: Modern Warfare 2" (with a score of 94), and "Haze" (with a score of 55). Thus, the authors involved 6 participants without any previous experience with these games, and they acquired the GSR signal via BIOPAC[12] suite. Then, the participants played at both games, and, during the game sessions, the two above mentioned methods were applied. The *biometric-based* approach uses the SCR to log the micro-events on a per-individual basis (see Fig. 3.1). The *observation-based* approach involves two experts to evaluate the gameplay. The authors observed that even if the *observation-based* approach can expose the majority of game design issues related to the usability, the *biometric-based* approach is able to acquire latent issues about players' feeling.

Another work [109] analyzed the effect of a horror game on the players' affect. The authors involved 11 participants to play at "*Slender: The Eight Pages*". In this game, the player, from the first person point of view, has to collect eight pages of a diary, avoiding the *Slender-Man*.

[12] https://www.biopac.com/

FIGURE 3.2: Screenshot of the Affect Annotation Tool GUI developed by Vachiratamporn et al. [109]. It uses a discrete set of markers that were assessed in particular phases of the gameplay.

The player is equipped with only a flashlight (with limited battery duration) useful to see through the fog and the dark. The antagonist character has the ability to teleport in different map areas, and if the *Slender Man* reaches the player, before she collects all the diary pages, the game is over. The authors collected the physiological information via EEG, and ECG, and the log of the keyboard and mouse inputs. They also developed a particular self-assessment tool, which presents a video of the gameplay and the player face during the experiment, and in the salient points, the participants had to mark the emotions (using the numeric key on the keyboard) in a discrete interval: neutral, anxiety, suspense, low-fear, mid-fear, and high-fear. The authors classify the data using an algorithm based on decision trees, obtaining an average precision of 0.88 (a similar value have been reached also for recall).

3.2.2 Games that Interact with Players

The evolution of video games provided intensive studies on the input devices. The modern controllers are developed to be comfortable and they have to provide an intuitive approach to the game. Beside controllers designed for specific use and/or people [110], the tendency for the innovative controller design is to interact with the games using a natural and realistic interaction. Some examples of commercial game peripherals, that use this approach, are: Nintendo Joy-Con, Microsoft Kinect, and Oculus Touch. The implementation of biofeedback sensors in these devices (or, more in general, in the use of biofeedback data game applications) may provide a further contribution to expanding the HCI field.

Liao et al. [111] developed and described how to design a dry BCI device in order to interact with games. The device uses 3 electrodes located on the prefrontal cortex, and a reference electrode positioned on the earlobe. In their paper, the authors illustrated all the steps of the development of the device (e.g., architecture, communication module), and they presented a computer game which acquires as input only the brain activity. The game consists in an archery level, where the player has to hit, with his bow, the center of the target. The accuracy of the aim is based on the "*focus*" level revealed by the EEG player's signal: if the concentration is high, the shot will be more accurate. The "*focus*" algorithm was developed by the authors, and it is validated using 10 participants.

Similar, but commercial, approach was conducted by Neurosky[13]. The company produces different low-cost EEG devices that acquire the information, with a single dry electrode, on the prefrontal cortex (the reference is positioned on earlobe). These devices are not invasive, and they communicate with the computer via a wireless protocol (i.e., Bluetooth). Moreover, the company developed a proprietary algorithm able to provide the information of "*attention*", and "*meditation*". Another BCI device named EPOC+[14], (produced by Emotiv) follows a similar approach. It involves 14 channels, using wet electrodes (with a saline solution). Another version of this device is the INSIGHT[15]. It is lighter than EPOC+, and it acquires the information on 5 channels using a "semi-dry" electrodes. The Emotiv products include a gyroscope, in order to detect the head movements. Different games (both digital and analogical) were designed to interact with this type of devices. Moreover, the device is successfully used in educational, and research fields. In an interesting paper [112], the authors described how to combine these 3 topics using the EPOC+ in order to interact with 2 games. The first game is named *BrainPong*, and it involves a competition between two players which moving the paddle using only the brain activity. During the data classification of the EEG signals, a tutor explains the role of motor cortex in motor plan generation and the related outcomes that can be detected via EEG. The second game is *EmoBlaster*, a single player twin-stick shooter controlled via a gamepad. Two variables can be controlled through the EEG: the difficulty (represented by enemy speed and spawn rate), and the enemy health. A high value of anxiety or frustration increases the enemy health, but reduces the difficulty (vice versa for calmness state). These two different states are presented to the players through visual feedback. The player can also defeat all the enemies on the screen through a motor imagery action. During the experimental phase, a staff member guided the player using different relaxation techniques, while, occasionally, she broke the focus disturbing the participant. Then, she started a conversation with the player and the spectators about the involved brain structures. These experiments were performed during two years in a summer camp, involving students between 8^{th} and 10^{th} grades. The researchers

[13] `http://neurosky.com/`

[14] `https://www.emotiv.com/epoc/`

[15] `https://www.emotiv.com/insight/`

obtained an overall positive feedback, and the students expressed an increased interest in science, and in the scientific career (especially for 10^{th} grade students).

As *EmoBlaster*, different studies and games also tried to merge the standard gaming controllers with further inputs provided by physiological information.

An example of a commercial video game which uses biofeedback as input is Nevermind[16], developed by Flying Mollusk. It is an adventure-horror game, which modifies the environment and the mechanics according to the player emotions. The game supports two kinds of devices: physiological feedback sensors able to reveal the HR (e.g., Wild Divine IomPE[17]), and emotional biofeedback (i.e., Affective Tool[18]). These devices can be used separately or combined, in order to provide a better user experience. In the game, the player has to explore and solve puzzles in the surreal subconscious of a psychiatric patient. The physiological sensors are used to identify the status of anxiety and fear, and, according to these felling, the game adapts dynamically its environment. For example, when a player feels a sensation of anxiety, the map starts to fill with water. Thus, she must return to a state of calm in order to continue the level.

Nacke et al. [113] used different physiological data in order to support the control of a 2d side-scrolling shooter game. The game involves different obstacles, enemies, and a final boss. The authors implemented 5 different power-ups, each one controlled by the physiological data:

- **Enemy Target Size**, which increases the enemy dimension, represented as a shadow, in order to facilitate hitting the target. This power-up is controlled by GSR, and respiration physiological information
- **Flame Length**, which increments shooting range of a specific weapon. This power-up is controlled by GSR, and respiration
- **Speed and Jump**, which increases the avatar speed and jump height. The power-up is controlled by ECG, and EMG
- **Weather Condition**, which modifies the quantity of falling snow presented on the screen, and, as a consequence, improving the visibility of the platforms and the enemies. This feature is implemented only in the boss area, which behavior is controlled according to the quantity of snow (more simple with a low quantity). This power-up is controlled by ECG, and skin temperature

[16] `http://nevermindgame.com/`
[17] `https://wilddivine.com/products/iompe-biofeedback-sensor`
[18] `https://www.affectiva.com/`

- **Medusa's Gaze**, which temporarily freezes the enemies and the moving platforms, when the player looks at these elements. In order to enable this power-up, the player has to collect a special item.

The authors performed a set of experiments on 7 participants and they collected their opinion through a survey. The results show that the participants prefer the features controlled by "*direct*" physiological outputs (e.g., EMG) rather than "indirect" (e.g., ECG).

BioPong [114] is a game which uses the biofeedback information (i.e., HR and GSR) to implement a DDA at the classic Pong video game. The authors developed two versions of the game, the first becomes harder when the physiological signals increase, and easier when they decrease, and the second vice versa. Thus, the researchers invited 12 participants to play at the two versions of the game, and at the classic version of Pong. After the experiments, the authors reported that albeit the biofeedback implementation has improved the user experience, it had not affected the players' performance between the different versions of the game.

A similar study was conducted by Chanel et al. [115]. They acquired different physiological signals (i.e., GSR, BVP, respiration, skin temperature, and EEG) in order to adapt the game difficulty of Tetris game. The participants played at 6 consecutive game sessions, using 3 different difficulties. After the game sessions, they had to answer a questionnaire on the emotions aroused during the game fruition and the level of involvement in the different experimental stages. Moreover, they had to rate their emotions in the VA vector using SAM. The main experiment result shows that the different 3 difficulties had aroused different types of emotions in players, which the authors identified as:

- easy difficulty: boredom
- medium difficulty: engagement
- hard difficulty: anxiety.

Tan et al. [116] conducted a study in which the participants played at *Portal* game. During the game session, the physiological signals of GSR, ECG, and 2 facial EMG (on zygomaticus and supercilii muscles) were acquired on the participants. After the game session, the participant had to verbalize her prior experience (i.e., *think-aloud*) while watching a replay of their gameplay. Then, the authors designed a set of think-codes in order to convert the think-aloud in discrete emotions domain. Both types (physiological and think-codes) of data were used to investigate different game events, providing interesting results.

A study on Free-to-Play Mobile game was conducted by Petersen et al. [117]. They investigated the first few minutes of gameplay (which can be considered the most important in this game category) of 3 mobile games, considering two physiological signals (GSR, PPG) and a set of self-report measures acquired through surveys and graphs. The authors identified a relationship between physiological measures and average player experience graphs, and self-reported engagement measures indicate a connection between physiological arousal response and qualitative proxy measures of engagement.

The use of biofeedback in gaming application can also improve the ability to control the desired physiologic reaction. For example, Zafar et al. [118] developed 3 mobile casual games which acquire in input the respiration of the player. The games main goal is to provide a support to the players in order to investigate the potentiality of relaxing through the breathing. The three games have simple mechanics and they are modified versions of open source *Android* games. Each one presents modifications to specific aspects of the game, in particular: *PacMan Zen*, a game inspired at the classic game of *Pacman*, modifies its game environment (where the player has to perform respiration exercises to collect points), *Dodging Stress*, inspired at the tilt-based games where the player has to move a ball from one side of the screen to the other, modifies its game controlled elements (i.e., the number of obstacles), and *Chill Out*, a version of bubble-popping game, modifies the player controlled elements (over 7 breaths per minute the cannon starts to "auto-shot" the balls, increasing the shot ratio according to the player status). The authors performed the experiment on 103 participants, in which they had to play at the games and to perform a cognitive task. The authors observed that the biofeedback games provided an improvement in players' ability to control the breath, both during the game sessions and during the cognitive task.

An exhaustive review on the different applications of physiological data in video game research is available in [119]. It investigates the research methods, and the different game aspects in which a study of physiological information can support the video game research (e.g., social game experience, game events, game features, etc.).

Chapter 4

Overview of the Applied Methods

THE main goal of our research is to infer the emotions through the physiological changes acquired on the players during video games fruition. In order to pursue this goal, we have designed a set of experimental sessions addressed to reveal these types of data, and to acquire further support information used to infer the players' emotions.

In the following sections, we present a high-level overview of the methodologies applied in the experiments. In particular, we discuss the physiological signals that we have considered as the most informative to infer emotions from video games players (in Sec. 4.2), and the different typologies of emotions identification (in Sec. 4.3). Moreover, in the following chapters (Ch. 6, and 7), we will provide a complete explanation and overview of every design choices and applied techniques of the experimental setups.

4.1 Framework Design

The automatic emotion recognition is a topic widely explored in Affective Computing research field. Usually, during the experimental sessions, the researches consider short stimuli related to a strong ability to arouse a specific emotion. However, video games are *interactive* software, and, as a consequence, they can stimulate a wide range of emotions in players.

There are different video game characteristics able to arouse emotions in players (see Sec. 2.1). For example, the game mechanics (e.g., jump) in different conditions may arouse a set of distinct emotions (e.g., jump on a platform, or jump on the head of an enemy). In addition, aesthetics features in games (e.g., music, graphics) are also fundamental elements in order to arouse emotions in players. The video games selected for our experiments should maximize the mechanics and aesthetics aspects, as they can be found, with a good approximation, in almost all video games. Instead, the selected games have to minimize other aspects able to arouse players' emotions, as social factors (e.g., multi-player), the role/behavior of NPCs, an explicit narrative,

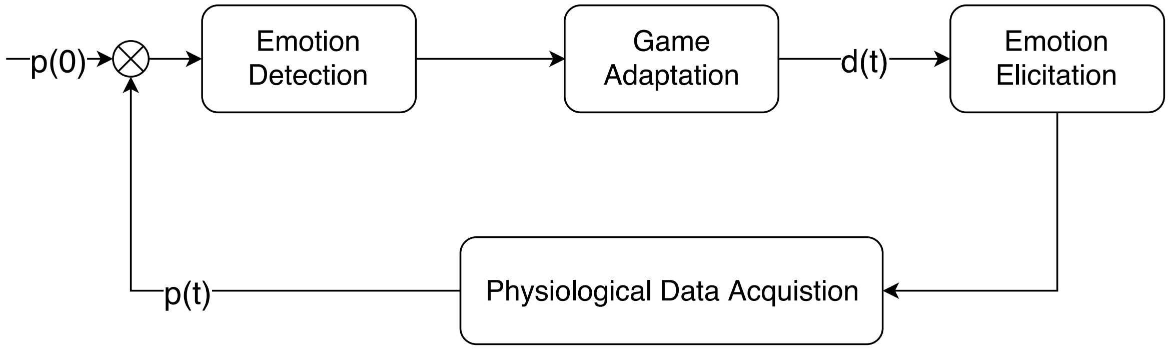


FIGURE 4.1: This Affective Loop scheme is inspired to the model purposed by Parnandi and Gutierrez-Osuna in [120]. In this scheme, p(t) describes the physiological information at time *t*, while d(t) represents the game difficulties.

etc. They can definitely have an impact on players' emotions, however, they are strongly dependent on the cultural heritage of the player, and they can not be controlled in time domain. Moreover, these characteristics are not available in all the games, and, as a consequence, they can not be considered representative of the whole media. In order to acquire a heterogeneous and large number of participants, we have to select games that do not require specific skills to be played. As a consequence, the games have to be designed for people with an age between 18 and 60 years old, and they have to involve simple and intuitive mechanics. This condition should ensure the game quality, and, as a consequence, the game ability to arouse players' emotions. These limitations do not permit to collect results representative to the whole video game panorama, however, it should provide a compromise between the generalization of the video game media and the experiment feasibility.

Our framework is based on the Affective Loop [11], in which the emotions can be acquired and they can be used to manipulate the digital environment according to the player response. In particular, the detection of the players' affect can provide the ability to the game designers to implements adjustable elements (see Sec 8.1). In our research, the Affective Loop can be viewed as a cycle between three elements (Fig. 4.1): elicitation of emotion in player (see Sec. 2.1), inferring the player's emotion (see Secc. 2.2 and 3.2.1), and adaptation of the video game to the player emotion (see Sec. 3.2.2). In this dissertation, we focused mainly on the second point, providing the basis for the design of a DDA novel method.

Our experimental approach is to acquire, through empirical experiments, information about physiological changes in participants body during video games fruition. Moreover, we asked the participants to report the emotions experienced during the experiment. Thus, we hypothesized that there is a relation between the physiological data and the self-assessment provided by the player. Starting from this hypothesis, we designed a framework to study the players' emotions during video game fruition using a quantitative approach. Thus, the framework had to provide a "*numerical*" outcomes which can be used to state the emotion in different moments in which the player is involved to play.

In order to collect the data, we had first of all analyzed the different techniques and tools used in similar researches. Thus, starting from the devices and the software used in literature, we developed new tools to use in the experimental setup, specifically designed to fit with the experiments requirements. A detailed explanation of tools developed for the experimental sessions is presented in Ch. 5.

Since the framework had to work autonomously, it was designed to use, during the data analysis, different techniques based on ML. Consequentially, all the aspects of the analysis (e.g., the procedure for the selection of the most relevant signals features) were designed to work without any external operation.

These methods were designed with two main goals: to propose a valid research environment available to the scientific community, and to develop a tool for game designers that may adapt their game at the players' emotions.

Thus, the research plan was divided into three main steps: to design the framework characteristics and to implement the necessary tools, to conduct a preliminary set of experiments in order to validate the framework and to highlight the critical aspects, and to acquire data (through a second set of experiments) with a significant number of participants. Moreover, both experimental sessions were divided into three separate steps: the architecture configuration, the experiments aimed to acquire physiological and emotional data, and the data analysis.

The participants information were anonymized, and at each participant was assigned an incremental ID. The information on the relationship between the ID and the acquired video is not provided to the scientific community, such as the videos, in order to ensure the anonymity of the data. However, the participants signed an informed consent which permits to use the audio/video information for the scientific divulgation purposes.

4.2 Considered Physiological Data

We had chosen to acquire 4 groups of physiological signals: ECG, EMG on facial muscles, GSR, and Respiration. All these signals can be revealed in two areas: on the surface of the hands and on the face. This restriction is applied in order to design (see Sec. 8.1), in a future work, an embedded device able to reveal these data without the aid of external devices. Nowadays, many video games have started to support games in Virtual Reality (VR) through the use of a headset. Consequentially, it can be a valid support for the face sensors placement. Moreover, a common way to interact with a video game is through a gamepad, which is usually held by the players with their hands. Thus, the sensors may be placed inside these game peripherals in order to acquire the data in a not-invasive way.

As stated in Sec. 2.2, the HR is an important information to infer the human emotions. In our research, to collect the HR data, we acquired the ECG following the guidelines designed by *Einthoven's Triangle* [121]. The physiologist discovered that it is possible to reveal the ECG signal placing the electrodes in three different limbs (left and right arms, and left leg), shaping an equidistant triangle centered on the heart. Thus, during the experiments, we connected the electrodes at the left and right writs, and at the left ankle. However, a novel researches [122, 123] demonstrated that is possible to acquire the HR also on human face placing a microphone on the temple. A different solution in order to infer the HR is to use a standard webcam, with the lens directed towards the user's face. Wang et al. [124] illustrated and evaluated different solutions proposed in literature using the video collected in the MAHNOB-HCI database [83]. In particular, one of the most popular and effective method to detect the HR using a video of human face is the Photoplethysmogram (PPG). The hemoglobin in human blood has a strong absorption spectrum in the light: therefore, the measure of the light variation reflected by a region of the skin provides the information of the blood vessel volume variation and, as a consequence, the HR. As a consequence, during the embedding of the architecture in video game devices, we may consider to implement a sensor able to acquire the HR directly on the player face.

The EMG data can be used to acquire information about the electrical activity generated by muscles. Different muscles reactions are involved when an emotion arouse (see 2.2). We limited the analysis on the facial muscles, permitting to embed the sensors into a VR headset.

The EDA signal is usually collected on two distal phalanges. As a consequence, it particularly fits with our purpose, as that area of skin is commonly in contact with the gamepad. A similar approach had already conducted by Bacchini et al. [125]; they designed a controller similar to the Playstation Dualshock which has installed different sensors.

A common way to acquire the information about human respiration is to place two Respiratory Inductance Plethysmography (RIP) bands on chest and abdomen. This methodology uses sensors that are placed in an area usually not involves in-game peripherals. However, the respiration signal can be also collected on human face, placing a thermometer under the player nose, as already proposed in [126].

4.3 Emotions Assessment

During the experiments, we need also to acquire information about players emotions. Thus, after the game session, we directly asked the participants to self-assess their emotional state during the video game fruition. Thus, an important decision to take is the type of emotions markers that must be applied (see Sec. 2.2).

We considered acquiring the emotion self-assessment values in a continuous time over all the game level. This created a novel challenge in the video game experimental design, since the majority of the researches in affective video game field identified the emotions in the discrete time, labeling the game sessions (or highlights) through a final survey (see Ch. 3). We agree that using a discrete time can provide a greater amount of data able to predict the emotional state, allowing an overall better accuracy on the ML model. However, to self-assess the overall game session with a single evaluation can be reductive, since a game stage can provide different and conflicting emotions. In contrast, if the research provides different evaluations using the game highlights, they may lose the focus on the overall emotions elicited by all the game session. In addition, defining the highlights as the parts of the game able to arouse emotions may be reductive for mainly two reasons: the game highlights may not affect the mental state of the player, and not all the games have emotional highlights. For example, some repeated mechanics in a game may affect the game ability to elicit emotions in players, and as a consequence, may not provide the desired emotion. Moreover, to evaluate only the salient parts of the game (excluding the context) can induce the player to assess the emotion following what he thinks it is "expected", to the *emotional bias* [127] and *response bias* [128], or to confuse the events if they are similar. Considering, for example, a basketball video game, where the player makes and receives several points. If only the most salient moments of the match are shown to the player, she loses the context, and, as a consequence, she may confuse the affective state elicited by a particular action, since each event is similar. Thus, to consider only the highlights may make the whole vision of the game lost. On the other side, some emotional games may not provide game highlights, but the implicit narrative and the environment are able to elicit the players' emotions. For example, *Journey* (developed by *Thatgamecompany*) does not provide any particular game highlights, however it is able to provide strong emotions in players [129]. A novel approach for the emotion classification may be to evaluate the physiological changes rather than game highlights. Of course, it does not resolve the issues based on biases and on the loss of the overall vision of the game stage, however it may assess the relevant information with the player point of view.

The most common way to describe an emotion is to identify it among a discrete set of labels (e.g., joy, boredom, etc.). This type of emotions identification may support the comparison between the participants, since it provides the most common method for humans to identify the emotions. Scientific researches identified several markers, however, they are often not uniformly defined across the different cultural backgrounds. In addition, different researches may adopt different sets of emotions labels. These characteristics of the discrete emotion assessment provide serious problems in order to compare the results across different studies. Moreover, using a fixed number of markers to identify an emotion may "prime"[1] the participant. This labeling method may also provide the opposite problem, where the participants want to self-assess their emotions with a category not available in the list, forcing the user to answer with the closest alternative. Even if the participant finds a category which corresponds to the experienced emotion, she may not be familiar with the label chosen by the researcher, being used to refer the affective state with a near synonym, for example, a more popular or slang expression (e.g. jittery in place of anxious) [47]. For our specific case, also the Ekman's six universal emotions [42] (see Sec. 2.2) can not be considered a valid approach. They are limited considering the range of emotions that can be experienced during video game fruition. Furthermore, Russell [130] underlined the gaps of Ekman's research by showing that the names of different emotions have an overlapped meaning in some languages. Summarizing, the discrete set of emotions may provide in specific cases a better accuracy in the comparison of the emotions between the participants of the same culture, providing a labeling near to the natural language. However, our goal is to provide a support to study the player emotions during video game fruition. Thus, due the limitations of the discrete emotion metric, we preferred to lean on a dimensional model to assess the emotions, since it naturally supports the data annotation in continuous time (see, e.g., [93, 89]).

As a consequence, we considered to map the emotions in an n-dimensional space. As stated in Sec. 2.2, the most common approach is to map the emotions into a 3D vector space, considering PAD as axes [131]. Following the approach of other datasets [89, 61, 93], in the present work we considered, for the emotions identification, only the VA vectors. This helps to reduce the time of the experimental sessions, and to attenuate the participants' bias. Moreover, the dimensional model can support the players to assess their emotions continuously over all the game session watching her face and her gameplay, without bother the game flow.

As mentioned in Sec. 2.2, a common practice to allow the players to self-assess their emotions is to use the SAM [52]. Moreover, Betella and Verschure [53] developed a valid alternative named

[1] To provide a list of markers may suggest a response that the participant would not have chosen in a natural way

AS. In our framework, we aim to maximize the support of the emotion labeling, implementing a valid combination of both tools.

Chapter 5

Tools

IN Affective Computing, several datasets were developed, each one containing information about users' emotions and a physical correspondence. In our research, the main approach used to develop tools able to study players emotions is inspired to the methodologies already presented in others papers which present affective databases [93, 83, 61, 89]. However, some characteristics of video game research (e.g., the duration of a game session, the different events that happen in a video game, and their consequent variability in the players' emotional responses) led to the need to design some novel tools.

Our solution is designed to be applied in affective experiments on video games research, however, the framework can be also used for different affective case studies; in particular studies with medium/long experimental sessions (around 10/20 minutes) which present different types of events (e.g., a short movie).

Part of the content of this section were presented in GOODTECHS conference and it will be publisheds in the "M. Granato, D. Gadia, D. Maggiorini, L. A. Ripamonti, "Software and Hardware Setup for Emotion Recognition During Video Game Fruition",*Proceeding of GOODTECHS 2018 - 4th EAI International Conference on Smart Objects and Technologies for Social Good, ACM*" [132].

This chapter is organized as follows:

- in Sec. 5.1, we provide a brief overview of the devices and the software already used to collect affective datasets

- in Sec. 5.2, we present the hardware architecture used to acquire physiological dataset, providing also our solution to define the experimental events and the data synchronization
- in Sec. 5.3.1 and 5.3.2, we present, respectively, the software designed to store and visualize the physiological information - DAPIS - and the emotion self-assessment tool - ESAT -. At the end of the section 5.3.2, we also discuss the ESAT validity, starting from a set of experiments.
- in Sec. 5.4, we provide conclusions and final considerations for future works.

5.1 Available Tools for Emotion Recognition

In our framework, we need three components: a hardware architecture able to reveal human physiological information, a software able to store the data, and another software to acquire the emotions self-assessment. In order to acquire a dataset structured by multiple physiological data, the hardware has to respect two main requirements: to record in real-time the physiological information, and to synchronize them with a common sample rate. In literature different open source and commercial tools were used, which are able to satisfy these constraints; for sake of brevity, we present only the most common devices. Biosemi[1] and Mindmedia[2] are two companies that sell to the researchers their biofeedback suites. These devices are already successfully used to create physiological databases (e.g., [83, 61], and they allow to acquire and to synchronize a predefined set of physiological data. For example, NeXus-32, developed by Mindmedia, is designed to reveal EEG, EMG, ECG, and EOG. Unfortunately, these devices present limited programmability, and they are often constrained by proprietary communication structure and sensors. Boccignone et al. [93] provides an alternative, using a more suitable methodology. They used an Arduino Uno, a programmable board, connected to an e-Health Sensor Platform shield[3]. Consequentially, the authors were able to define a personal communication protocol. However, the Arduino UNO Analog to Digital Converter (ADC) can quantize data at 10-bit precision, a resolution that can be, in particular cases, not sufficient to understand some peculiar physiological features. Furthermore, the e-Health Sensor uses the analogical Arduino pins with presets sensors. As a consequence, the inclusion of additional devices able to acquire physiological data on the board is limited.

Usually, companies that develop devices able to record physiological data also provide a software to acquire and store the digitized data on a computer. For example, Biosemi developed *ActiView*, a

[1] `https://www.biosemi.com/`
[2] `https://www.mindmedia.com/`
[3] https://goo.gl/B5qFsy

software to acquire data collected by *ActiveTwo*[4]. Albeit the software is an open source project, it is designed to work with the company proprietary device, making the porting on different hardware setup a challenge. Others software (like, e.g., OpenVibe [133]) were designed to support a wide range of devices, however, often they are focused on a specific physiological signal. LabView [134] is a general purpose software which can be used to acquire and to elaborate almost all data acquired by sensors. It also supports a wide range of devices, including Arduino boards, and can be programmed by a peculiar visual programming language. However, it is a proprietary software, with high-level minimum hardware requirements.

The current approach in emotion tagging applications is to report the users' emotions on *n* vectors (with n equal to 2 or 3). A common implementation is to support the annotators, visualizing in the software GUI the SAM [52]. Feeltrace [135], DANTE [93], and ANNEMO [89] are all valid tools used to identify the emotions in VA vectors. Unfortunately, none of these were designed for events management, like, e.g., the beginning and the end of different levels. Moreover, they usually require to tag the emotions on vectors one at a time, a solution that improves the accuracy but that can be suitable only for short video sequences.

Lastly, in our architecture, we focused on understanding how to implement an accurate synchronization between the data. Noldus has developed the Oberver-TX[5], it is a software developed for multimodal researches purpose able to synchronize different types of data, such as audio/video, physiological, and eye gaze. Unfortunately, the software is able to integrate and to synchronize data acquired through specific commercial suites, like, e.g., BIOPAC[6] (for physiological data), Tobii Pro Studio[7] (for eye tracking), and Viso[8] (for video recording).

5.2 Design of the Core Hardware Architecture for Signal Acquisition

Our research objective is to infer users' emotion using physiological data. Consequentially, the hardware architecture has to acquire physiological information, and a set of digital inputs used as "flag values". These values describe different events which happen during the experimental session. Moreover, the device has to provide an ADC, and to prepare a communication protocol to send data to the computer. Therefore, we designed a hardware architecture based on Arduino Due.

[4] https://www.biosemi.com/products.htm
[5] `https://www.noldus.com/the-observer-xt`
[6] `https://www.biopac.com/`
[7] `https://www.tobiipro.com/product-listing/tobii-pro-studio/`
[8] `http://www.noldus.com/viso`

It is a programmable board that uses an Atmel SAM3X8E ARM Cortex-M3 CPU with a 32-bit core[9]: its computational effort should guarantee a correct functioning of the data elaboration and communication. The micro-controller has also 12 pins able to read analog information and 54 pins for digital I/O. These pins tolerate a maximum tension of 3.6V, however, in our setup, we used a voltage equal to 3.3V, as suggested in Arduino Due documentation[10]. Moreover, the ADC embedded in the CPU permits a conversion at 12-bit, obtaining, as a consequence, an ADC step equal to 805.66 $\mu V/bit$.

Using a set of sensors connected to the Arduino, we acquired 4 groups of physiological signals: ECG, EMG up to 5 facial muscles, GSR, Respiration. We also considered the recent evolution of VR headsets in gaming applications, which lead to a higher level of immersivity in the players' experience. Thus, our architecture was designed to support and to be embedded in these devices. Moreover, as we will describe in the following pages, it is possible to add any analog sensor (e.g., a thermometer to measure the skin temperature) to the architecture. Focusing on physiological signals, we used 6 Olimex EKG-EMG shields[11] to acquire information about ECG and EMG. It is a device already used in biomedical engineering [136], and it is able to read a 3-lead electrode connector via 3.5 jack. Thus, we connected the leads of an Olimex shield at the wrists and the left ankle, following the guidelines provided by *Einthoven's Triangle* [121]. The other 5 shields were used to acquire the EMG information up to 5 facial muscles. Lastly, we connected a common reference at the border of the hair line. A couple of electrodes were used to collect the GSR signal connecting them to the phalanges of two adjacent fingers which are not involved during the experimental session. We used the Grove GSR sensor, which is equipped also with an amplifier (LM324) in order to improve the information quality provided by the skin potential difference. The last physiological signal is the respiration rate/intensity. It is collected by placing a thermistor under the participant's nose. We also isolated the base of the sensor to avoid the direct contact with the user's skin, minimizing the noise generated by the epidermis temperature. The thermistor provides an accuracy of ± 0.5 C° (between 25 C° and 85 C°), and it reduces the tension when the temperature increases, in our specific case when the user exhales (vice versa when she inhales).

We used a 4x4 keypad in order to defines game events. At each key was associated an integer value in a range [0,15], while the value 16 was used to identify a no event state (i.e., when no buttons have been pressed). In our experiment, the considered events are only the beginning and the end of the game levels, however, the architecture was designed to capture information up to

[9] datasheet: `https://goo.gl/ZM8zpQ`
[10] https://store.arduino.cc/usa/arduino-due
[11] datasheet: `https://goo.gl/TDD8UZj`

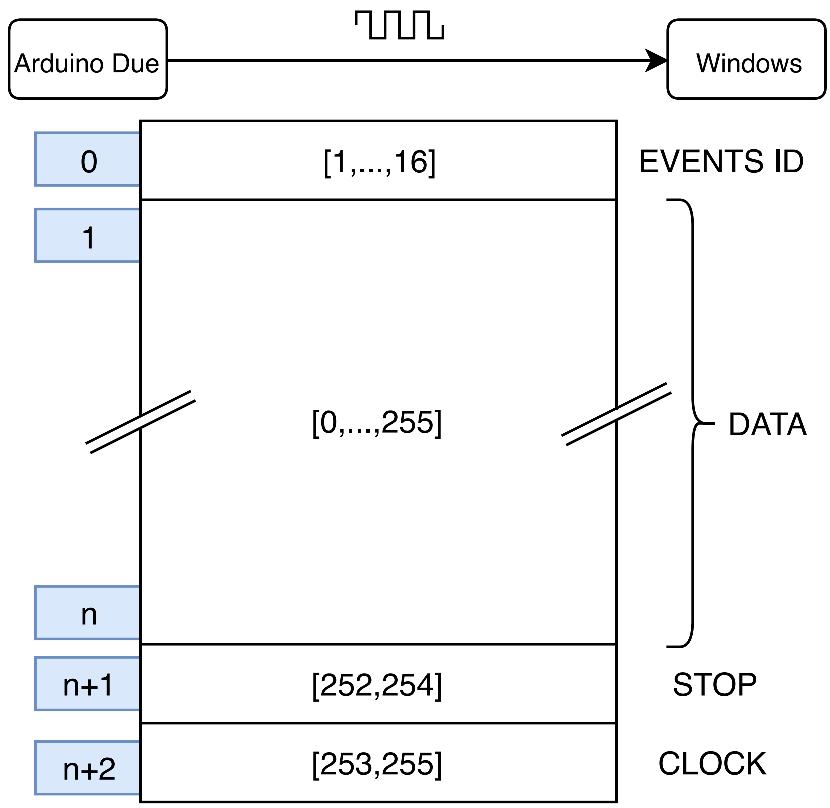


FIGURE 5.1: Packet structure sent by Arduino to the Computer.

15 events. However, these events can be used to analyze specifics and different game scenarios. In a future study, we will consider different parts of the level, as well as different game events, synchronizing the physiological and emotional data through the keypad. We also used two bytes to, respectively, delimit the buffer and clock alert advertise (STOP and CLOCK). The former was used to provide an information of buffer control and to define the end of the data buffer, while the latter was used to define the start of Arduino CPU clock. These bytes were sent after the data quantization performed by the Arduino Due ADC. Furthermore, they modify their values when Arduino starts a new clock, in order to communicate this information to the computer. The delimiter byte was also used to support the clock byte (in case of communication loss), modifying its value according to the CLOCK byte.

The analog data were converted to digital through the built-in ADC provided by Arduino Due CPU. As we mentioned, it permits to quantize the data at 12-bit precision. However, the serial communication can send only one byte at a time (2^8). Consequentially, we split each converted analog data into two bytes, that we named *h-byte*, and *l-byte*. The buffer structure is presented in Fig. 5.1. The overall architecture is presented in Fig. 5.2. If a new sensor is added to the

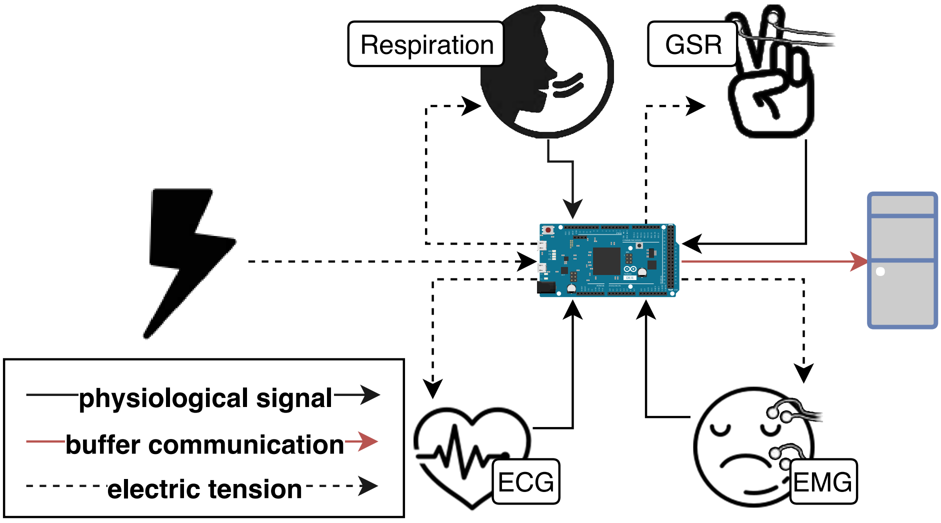


FIGURE 5.2: General architecture model. It considers 4 different types of physiological information acquired and synchronized by an Arduino and communicated to a computer

hardware setup, the acquired information may be appended to the end of DATA area using the guidelines presented in this section, increasing the packet size.

In this section, we described an overview of the hardware used to acquire physiological data, providing the information about only the architecture core. A detailed description of the implementation used for the pilot study is provided in Sec. 6.1.3. Then, starting from the outcomes of the pilot study, we had further enhanced the implementation in order to create the RAGA dataset. The final implementation of the hardware acquisition setup is described in Sec. 7.1.

5.3 Data Acquisition and Self-Assessment Software Design

In order to store the physiological data revealed by the architecture presented in the previous section, we designed a software able to visualize, and store the information communicated by the Arduino. Moreover, w designed a software to collect data about emotion self-assessment.

5.3.1 Software for Physiological Data Acquisition: DAPIS

For our experimental setup, we developed two different software: DAPIS, a software developed on the basis of an open-source project[12], used during the experimental phase to acquire, to synchronize, and to visualize the physiological data, and ESAT, which is used immediately after each

[12] https://github.com/vsquared/ECG_UNO_Processing3_2_3

video game session to self-assess the participant's emotional states. Both applications were developed in Processing[13], a programming language based on Java, which aims mainly to produce visual contents. For our experimental setup, a video of the screen during the game session was recorded, placing DAPIS GUI in the top-left area of the acquired video. In particular, the video contains the participant's face, acquired through a camera, and the video of the gameplay (Fig. 5.5). The DAPIS GUI is structured by 3 main components: the top bar, which presents a set of buttons used to interact with the software functions, two colored bars used to synchronize the physiological data with the self-assessment information acquired by ESAT, and the central area which visualizes, in real-time, the signal plot. The visualization of the plot is used only to identify, during the experiment, the quality of the acquired signals through the top and bottom colored bars. DAPIS supports the serial communication, which is used to exchange packages with external devices. Thus, selecting the correct COM port, the application is able to receive data from the architecture based on Arduino Due presented in Sec. 5.2. Moreover, the software automatically writes, in real-time, in a specific path selected by the user, all the acquired physiological data. DAPIS also analyzes the value of the *event flag* (i.e., the first byte in the buffer). According to its value, the bars switch their color, permitting the data synchronization with ESAT (Fig. 5.4). In our specific case, a button is used to identify the beginning of the game level (green bars), while another button defines the level end (red bars). DAPIS contains also different functions able to improve the overall experimental quality. The mains functions are: to change the physiological data visualized in the central area using the keyboard space-bar, to move the center of the plot through the keyboard arrows, to clear the central area, to take a screenshot of the plot, and to record a data baseline of custom duration in a separate file.

5.3.2 ESAT: Tool for Emotion Self-Assessment

When the experimental session ends, the video acquired was reproduced into ESAT. In our specific case, the video was structured by three components (Fig. 5.5): the player face (c), the gameplay (d), and DAPIS GUI (e). The staff member had to ask the participants to focus only on the area of the video that concern the face and to support the assessment with the video of the gameplay. The DAPIS GUI was used only to synchronize the acquired data as mentioned in Sec. 5.3.1. Then, ESAT synchronizes its data with the physiological data acquiring the color of the bars presented in the left area of the video. At each color it associates a specific value; in our case study, we have associated 0 for the red bars, and 1 for green. During all the video, the participant interacts with

[13] `https://processing.org/`

the left and right analog joystick of the same gamepad used to play at the video games to control the self-assessment bars, respectively the left is used to identify valence values (a), while the right is used for arousal (b). To support the self-assessment, we implemented two tools, SAM [52] and AS [53] (see Sec. 2.2). Moving the analog joysticks on the horizontal axis, the participants move the green and the red rods on the bow-tie graphs and a semitransparent square underlines the manikins faces.

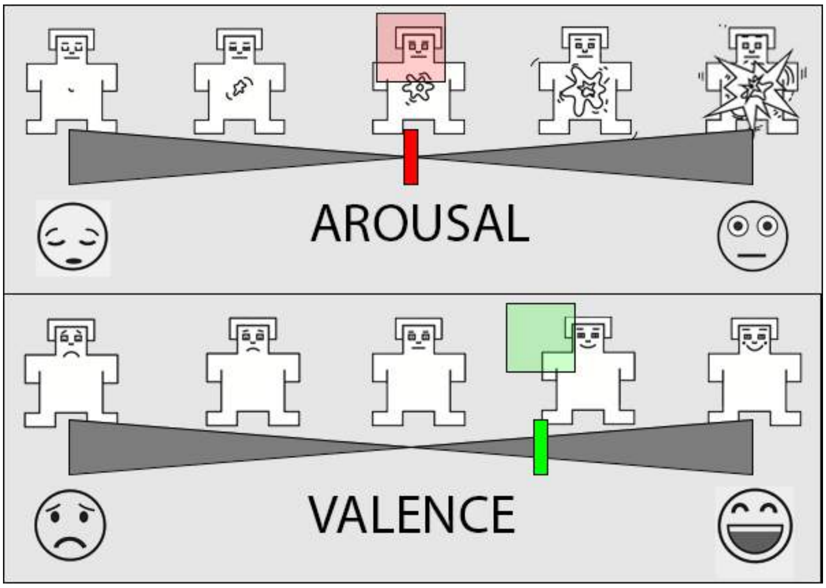


FIGURE 5.3: These bars have been used to self-assess the participants' mental state during the experiments. It involves SAM and AS tools in order to support the emotions identification

A red line (f) underlines the time spent from the beginning of the self-assessment. Moreover, the participants can stop and rollback the video; however, during the experiment, we suggested for each user to limit the uses of these functionalities.

Both applications provide in output a *CSV* file containing tables of length $(fps * seconds) \times 4$ for ESAT, and $(samplerate * seconds) \times (2 + \text{N° of Data})$ for DAPIS (Tab. 5.1). Moreover, the tools are persistent, thus, they try to minimize the data loss: if a computer crash happens, saving acquired data at regular intervals.

TABLE 5.1: The markers presented in these tables are the headers of *CSV* files.

Emotion Tag

Frame ID	Arousal	Value	Game Status

Physiological Data

Game Status	Data 1	Data 2	...	Data n	SampleRate

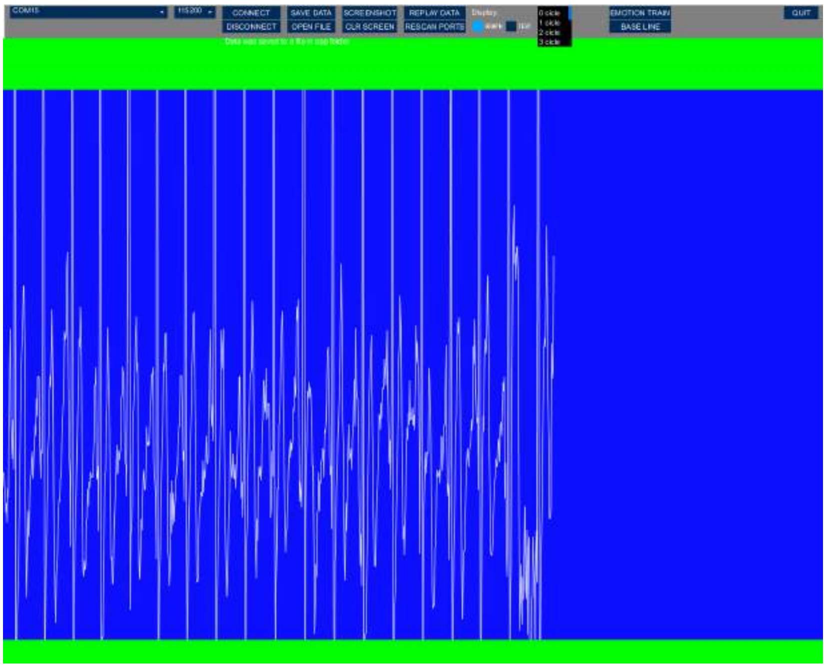


FIGURE 5.4: DAPIS GUI, it shows the colored bars used to synchronize the data, in particular, the green bar represents the start of the game level. After pressing a button connected to the Arduino, the bars change their color to red, which identifies the level end.

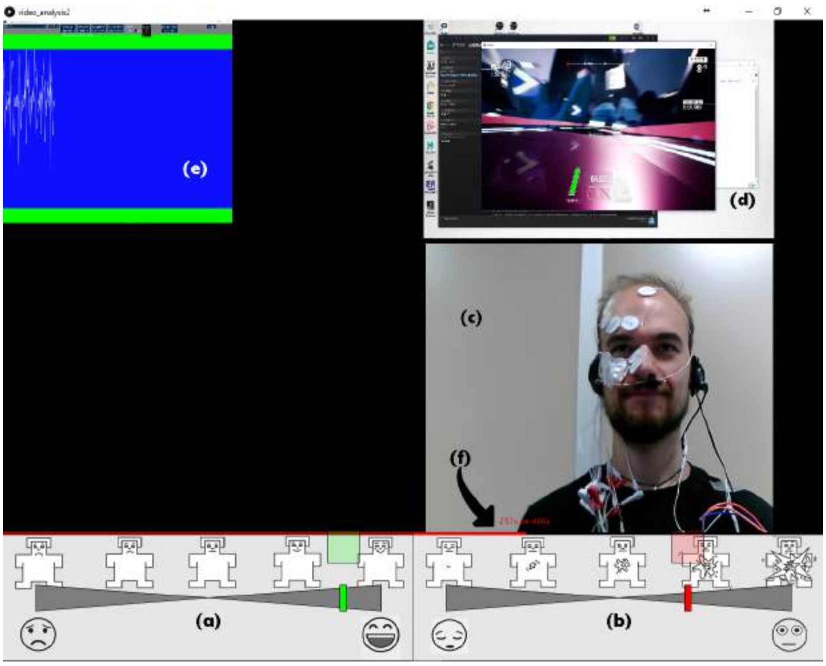


FIGURE 5.5: ESAT GUI, used to identify the emotion tagging. It shows the video of a game session and asks the user to identify hers emotions over all the playback.

The experiments and the data analysis were performed on a computer with Windows 10 OS, an i7 6700k CPU, 32gb RAM (DDR4), and a NVIDIA GeForce GTX 1080. During the experiments, a camera (a 5 MegaPixel camera with an image resolution of 1280x1024, able to record video up to 60 frames per second) was placed under the monitor used during the gaming sessions (i.e. a 32" LCD display). During the experiments, the participants used a standard gamepad (GameSir G3w) in order to interact with the games and ESAT software.

Although functional tests were performed on a computer with Windows 10 OS, both applications were developed in a multi-platform language, taking care to produce a software able to work on the most common operating systems.

5.4 Final Considerations

In this chapter, we presented a novel and flexible architecture to acquire human physiological data during video game sessions. Moreover, we provided a description of two software developed in order to store the digitized physiological information (DAPIS), and to acquire emotion self-assessment data (ESAT). We also performed a validation of ESAT concurrently with the collection of RAGA dataset, that will be presented in Ch. 7. Thus, the participants used for the ESAT validation are the same users involved to collect RAGA dataset. As a consequence, we will present the validation analysis in Sec. 7.5.1.

The descriptions of the software and of the hardware setup are freely available on GitHub[14].

Albeit the tools presented in this paper are sufficient for our experimental setup, DAPIS can be further improved, for example implementing a dynamic adaptation on the number of physiological data acquired. A similar solution can be also applied on the overall package structure, permitting to place the STOP and CLOCK byte in any position of the stack.

Other improvements can be considered for the ESAT tool, like the support for other input devices (e.g., the signal of two potentiometers connected to an ADC). Lastly, the two software collect data at two different sample rates: DAPIS samples at the same rate of the Arduino, while ESAT at the video frame rate. Thus, these data, usually, are aligned in post-processing. A future implementation of ESAT can consider a built-in algorithm for signal alignment, in order to interpolate the VA matrix length equal to the physiological data matrix, simplifying the data analysis.

[14] https://github.com/grano00/VGRDevicesAndTools

Chapter 6

Pilot Study

IN this chapter, we provide a description of the pilot study used to validate the framework presented in Sec. 4.1, and, as a consequence, to verify its reliability. Then, we conducted a set of empirical experiments, using the methodology and tools presented, respectively, in Ch. 4 and 5. In the experiments, we involved a limited number of participants to play at a set of video games and to self-assess their emotional state. The pilot study main goals are:

- to understand the reliability of our framework to infer the players' emotions
- to underline the technical and practical issues of the experimental setup
- to test the prediction on general and individual models, using only the data acquired by the participants

Part of the contents of this chapter were published in "M. Granato, D. Gadia, D. Maggiorini, L. A. Ripamonti, "Emotions Detection Through the Analysis of Physiological Information During Video Games Fruition", *Springer Lecture Notes in Computer Science (Proceedings of 6th International Conference of Games and Learning Alliance - GALA 2017)* 10653, pp. 197-207, Lisbon, Portugal, December 2017." [137].

6.1 Experiments Overview

To investigate the players' emotions during video games fruition, we performed a set of experiments with a group of participants. We asked the participants to play different video games. During the gaming sessions, we recorded a set of physiological information, and we also collected the emotional feedback using the dimensional model (see Secc. 2.2 and 4.3). The participants had to play different video games and, they were asked to self-assess their emotional state. After the data collection, we analyzed the acquired signals and we used ML techniques in order to predict the

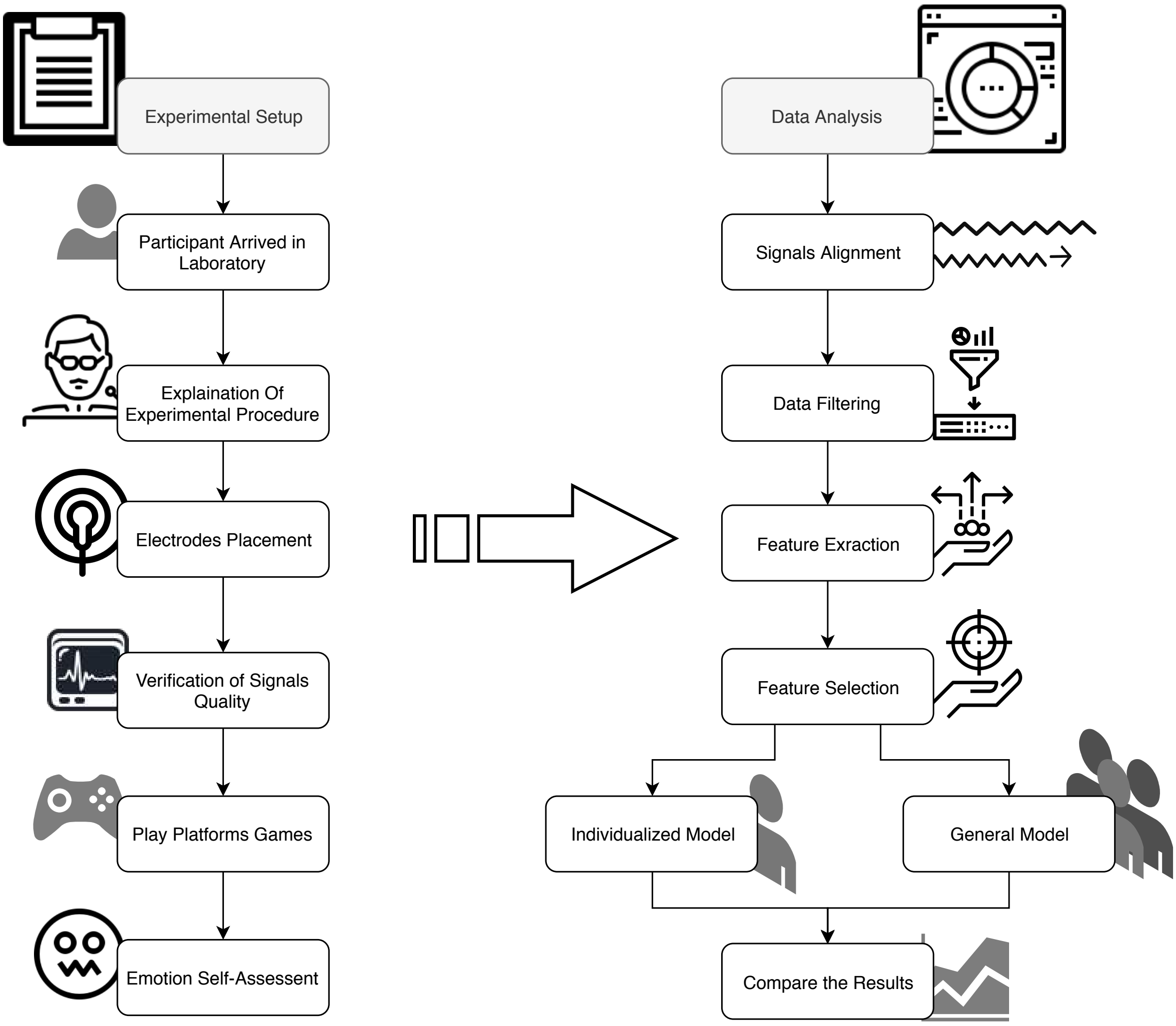


FIGURE 6.1: Pilot Study Flowchart. On the left, the figure shows the different steps involved during the experimental setup, while, in the right, it illustrates the main steps involved in the data analysis.

self-assessment values provided by the participants. The different pilot study steps are summarized in Fig. 6.1.

6.1.1 Participants

The group of participants was composed of 10 males between 18 and 38 years old (in particular, 50% of them has an age between 21 and 24 years). 90% of participants usually plays video games more than 3 days by week, and 30% plays every day. Furthermore, the 60% of participants declared that they play at home at least two hours for each game session. For each participant, we assigned an incremental ID, and the acquired data have been anonymized. Moreover, we asked the participants to take note of their ID, and to report it in case of application in future experiments.

When asked about the preferred game category, half of the participants indicated platform games, also claiming to be sufficiently skilled in this genre. Unfortunately, due to technical issues, one participant was not eligible for the study, and then his data was discarded.

All the participants were volunteers from Italy and they had not received any monetary or academic contribution for the experiment.

6.1.2 Considered Games

All video games selected for the experiment are platform games. This video game category is typically structured by a linear environment where the player starts from a position A and must reach a position B, avoiding obstacles and defeating enemies. Usually, platformers have simple mechanics (i.e., the character can jump, run, and attack the enemies), and a linear progression structured in levels.

This game category has received several changes in the game design and level design over the years (both in the game difficulty and in the interaction with the game environment). Albeit this genre received a great diffusion on *home consoles*, many important platform games have been also released to *arcades*, *mobile consoles*, and *computer*. As a consequence, its diffusion has permitted to consider this category of games as one of the most fruited, and, consequentially, it is particularly indicated for this type of study. Thus, we selected a heterogeneous topology of video games belonging to this genre. The different indexes used to define the topologies are: game environment (2D/3D), release date (recent/"classic" games), and license (commercial/non-commercial games). Moreover, all games should have received a good rating by the specialized critic ($\geq$ 80 in Metacritic[1], or $\geq$ "*C*" in Video Game Critic[2]) or, in any case, they should be considered *enough fun to play* [138, 139].

The selected games are:

- *Rayman Origins*[3] (2011) is the fourth chapter of its series and it is developed by *Ubisoft*. It consists in a 2D side-scrolling platformer with collectible items. The participants have to play the first level (as a tutorial) and the third level in the third stage.
- Six platform levels generated with *FunPledge 2.0* (developed in 2016) [139]. Evolution of a previous work [138], it is a tool able to generate 2D side-scrolling platform levels automatically, on the basis of a musical rhythm used as a basis to define challenging and

[1] https://www.metacritic.com/
[2] https://videogamecritic.com
[3] https://www.ubisoft.com/it-it/game/rayman-origins

entertaining levels. The players should finish all the levels. The game session has been also stopped if the player reaches the game over.

- *Earthworm Jim 2* (1995) is the second chapter of a series developed by *Shiny Entertainment*. It is a 2D side-scrolling platform where the protagonist can also shoot to the enemies. The game presents also some puzzle elements. The player should finish the first level. The game session has been also stopped if the player reaches the game over.
- *Crash Bandicoot* (1996) is the first game of a series initially developed by *Naughty Dog*. It is structured by levels similar to the classic 2D side-scrolling platforming and 3D levels. The players have to play the thirteenth level (a 3D level). The game session has been stopped also if the player reaches two game overs. This softening of end-of-session rules is given to allow the user to switch between an analog cursor to the directional crosses.

6.1.3 Acquisition of Physiological signals

In Sec 5.2 we described the core architecture we proposed for the acquisition of physiological data from players during video game sessions. For the pilot study, we extended the core architecture by adding the support up to 6 facial EMG, a digital thermometer in order to acquire the respiration signal, a serial communication via a virtualized USB. More details about the extensions used for the pilot study are descried in the rest of the section.

From each player, we acquired different physiological data: ECG, EMG on 4 facial muscles, GSR, and respiration rate. As stated in Sec. 5.2, the data was collected with a set of sensors connected to an Arduino Due, that performs a 12-bit quantization with its ADC, and it sends the data to a computer through Serial Communication (virtualized on USB). The Arduino was feed directly by the computer, and the sensors were powered by the Arduino pins. The video was recorded through a camera with a frame rate of 30 fps.

The EMG electrodes were placed on 4 different areas of the right side of the players' face as illustrated in Figure 6.2: on *Zygomaticus Major* (*EMG1*), *Orbicularis* (*EMG2*), *Nasalis* (*EMG3*), and *Supercilli* (*EMG4*) muscles. Furthermore, on three users we acquired an additional EMG signal on *Temporalis (EMG5)* muscle, while on other three users the electrodes were placed on *Depressor Labii Inferioris muscle (EMG6)*. Although EMG5 and EMG6 data are not used in the current analysis, they were recorded in order to study the efficacy of the related muscles to infer the players' emotions, and the ability of the sensors to acquire a valid signal on the participants.

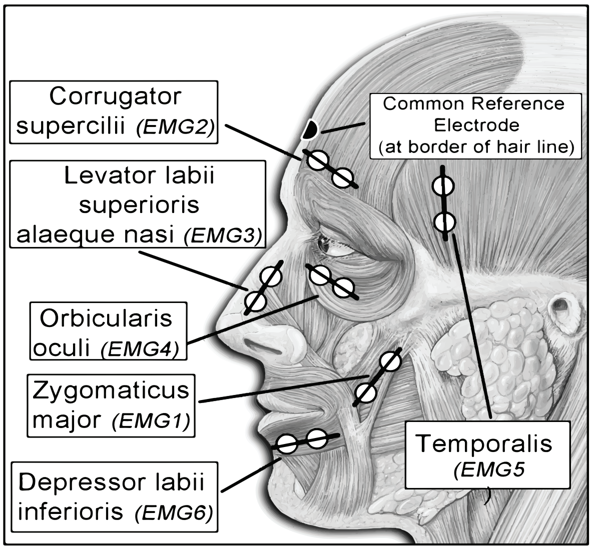


FIGURE 6.2: Representation of facial electrodes positions connected to EMG sensors. Original medical illustration from Patrick J. Lynch (https://goo.gl/ttgxo6)

These pairs of electrodes used to acquire the EMG signals have a common reference electrode, placed on the forehead, near the hair border, as suggested in [140].

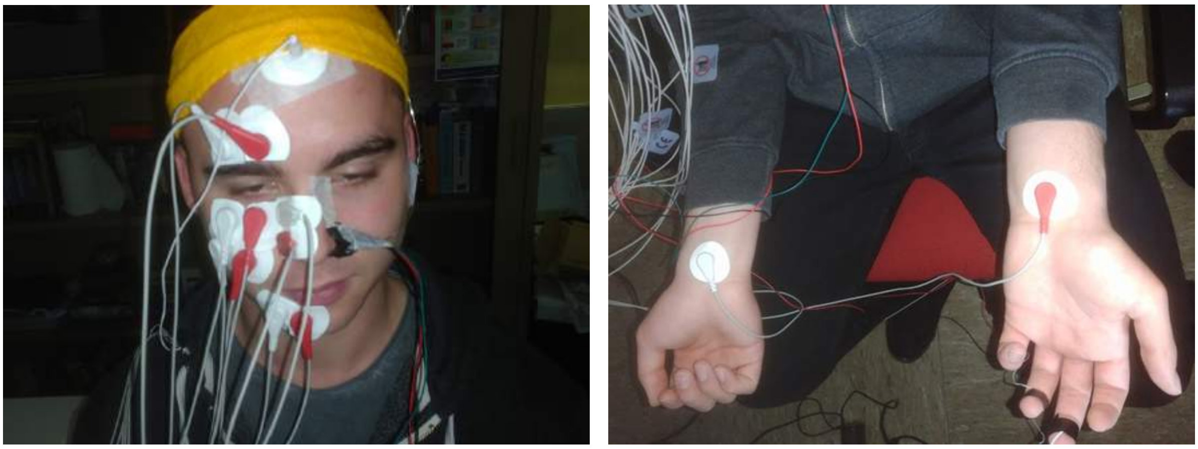
To interact with the selected games, the players used only the frontal gamepad buttons (i.e., the direction arrow, the digital joystick, and the frontal buttons). These buttons were designed to be used through the thumbs. Consequentially, we collected the GSR data placing two electrodes on the index and the middle phalanges of the left hand, because, as above said, these fingers were not used to control the selected games.

The respiration intensity/rate was measured placing a digital thermometer (DS18B20) under the player's nose. The thermometer was placed avoiding the contact with the player's skin, limiting, as a consequence, the noise generated by the epidermal temperature. The respiration data were stored with an 11-bit resolution (the sensor has a fluctuation of 0.125 C°) and it was sampled each 375 ms as indicated by the DS18B20 Datasheet[4]. Due to the sample rate of the thermometer, the overall Arduino frequency was not stable, having a fluctuation between 952Hz and 989Hz. However, this issue was corrected in post-processing analysis.

The communication buffer involved 19 bytes, each one transmitted at a frequency band of 250000, without parity control and with 1 stop bit (250000/8N1).

For each experiment, we used the Fiab F9079/100 disposable electrodes (36x40 mm) to acquire the ECG, EMG, and GSR data. Their dimension is quite sizable compared to the electrodes usually used for the facial EMG and the GSR data, because they are designed to acquire data for ECG or

[4]Datasheet: `https://datasheets.maximintegrated.com/en/ds/DS18B20.pdf`



(A) Electrodes used to acquire facial EMG

(B) Electrodes used to acquire ECG

FIGURE 6.3: Two photos captured during the experiments. They illustrate the electrodes placement on a participant

EMG on a great skin surface. However, we artificially reduced their dimension in order to acquire data on the participant's face. Moreover, for the GSR sensor, we separated the metal area from the conductive gel and the surrounded band. Thus, we cleaned the finger skin with a disinfectant alcohol, we wet the metal area surface with a conductive gel, and we placed the electrodes on the participants' fingers, firmly keeping them through an insulating tape.

6.1.4 Procedure

The participants were invited to sit on a comfortable chair. They were informed about the experimental procedure and they were invited to read and sign an informed consent, and a permission to use the video and images recorded during the experiment for research and academic purposes. The acquired data were collected anonymously, and at each participant, an incremental identifier was assigned to be used for the data analysis and for future experiments. Lastly, each user performed the experiment in a unique daily session. In order to attenuate the *observer-expectancy effect* [141], in each phase of the experiment, the staff member was positioned in a predefined area out of the participant's field of view. She also avoided any kind of interaction or observation during the different experimental phases. Moreover, we asked the participant to reduce the contacts with the staff member, and to report only technical problems.

The experiment consisted of three different stages: electrodes placement, game session, and emotion labeling.

In the first stage, the sensors were placed taking care to not bother the players during the session. Thus, we asked the participant to perform specific movements with her face in order

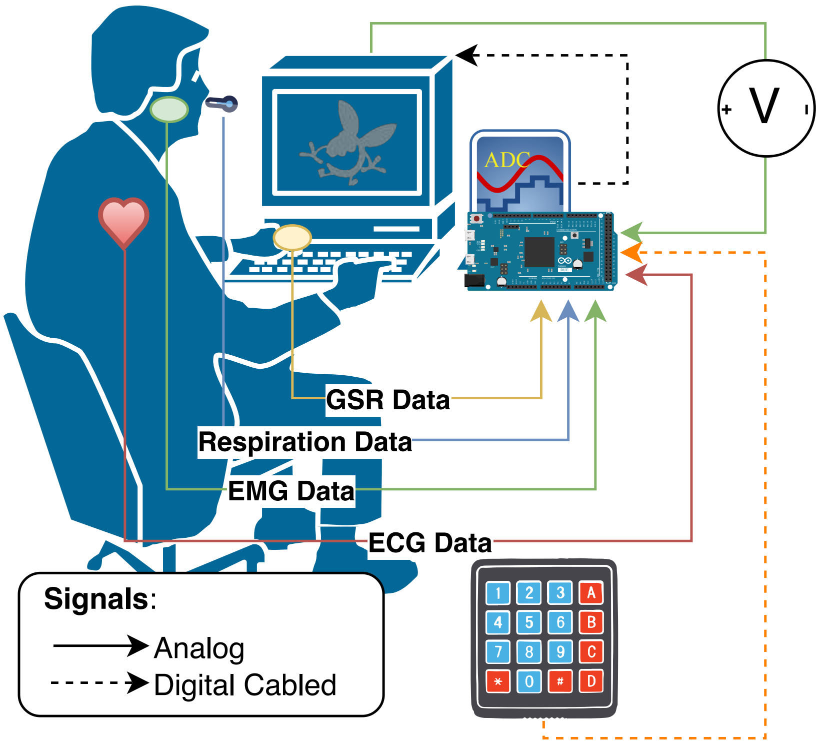


FIGURE 6.4: The figure shows the overall architecture used to acquire the physiological information during the pilot study

to check the EMG signals through the DAPIS GUI. In case of a noise-rich signal, we discarded the electrodes of the specific EMG and we replaced them with a new set. A similar procedure was applied also for the ECG. To check the GSR data, we asked the participants to perform a deep breath. Usually, a deep breath elicites the GSR signal, which provides a general sympathetic discharge [142], and, as a consequence, it increases the sweating of the fingers. Thus, if a small wave on GSR signal was not shown on DAPIS GUI, we removed the GSR electrodes, we cleaned the metal area and the participant's skin, and we replaced the conductivity gel. Lastly, we controlled for anomalies in the respiration signal. In case of a low general difference between low-peaks and high-peaks data, we repositioned the sensor in order to avoid different sources of heat (e.g., the skin warmth).

In the second stage, the participants had to play all the game levels, in the same order described in Sec. 6.1.2. Before each game session, the games mechanics and the level goal (as described in Section 6.1.2) were explained to the participant. In each game, we considered the level ended when the participant have completed it, or at game over. Moreover, before the start of each level, the electrodes were controlled and, in the case of unsticking, they were replaced. Before and

FIGURE 6.5: A screen shot captured during the self-assessment stage. Here, the participant was identifying his emotions during the playback of a video which presenting the information about the player' face and the gameplay

after each level, the player had to press a physical button (connected to the Arduino) in order to synchronize the physiological data with the game sessions as illustrated in Sec. 5.2. During all the game sessions, we asked the participants to limit the interaction with the laboratory staff.

In the third stage, we asked the participant to map their emotional states on the VA dimensional model using ESAT (see Sec. 5.3.2). In particular, we asked the participants to remove all the electrodes and to have a little break. After the small pause, a video of the recorded game session and of the participant's face, during the playing time, was shown to the player. All the videos regarding games levels were reproduced in the same order of the second stage. In particular, the videos were composed by three zones (see Fig. 6.5): in the left area, the synchronization bars of DAPIS are shown, in the central area the gameplay, while in right area the player's face.

After an initial training with ESAT, a staff member asked the participant to use the two gamepad analog joystick to set two pointers, one for arousal and one for valence, following the SAM and the AS markers as described in Section 5.3.2. She could also rewind the video in order to correct her choice evaluation, however, we discouraged the use of this feature. We have hypothesized that a fluent view of the video can evoke better the emotions experienced during the experimental session, thus we recommended to rewind only after making a serious mistake. To support the participant's focus during the emotions self-assessment, we added a little pause between each video in order to split the tagging session into small parts.

Lastly, each user was asked to answer a survey with questions regarding her gamer skills and habits, and regarding the overall experiment considerations.

6.2 Data Analysis

After collecting the physiological data and the self-assessment information on each participant, we designed a data analysis. In particular, the analysis consisted of 4 specific points:

- **Data Filter**, in order to remove the signals noise produced by the electrical current, and the participants' skin,
- **Feature Extraction**, in order to define a set of features, starting from the signals of RAW data, which we can use to predict the players' emotions
- **Feature Selection**, in order to remove the unnecessary variables, and, as a consequence, have a shorter training time, have a lighter final model, and have a panoramic on the most relevant variables
- **Self-Assessment Data Prediction**, in order to define a model able to infer the players' emotions during video games fruition

6.2.1 Data Filtering

To correct the variable sample rate, each signal second was sub-sampled to the lowest value (i.e., $lw = 952Hz$). Thus, we separated all the physiological data in different matrices of one second, where each column defines the physiological information. Thus, for each matrix in the set $S = M_1, M_2, ..., M_{second}$, we applied a linear interpolation on their columns obtaining a new set of matrixes.

The signal of each physiological information was smoothed using Savitzky/Golay filter [143], applied with a polynomials fit of order 30 and a floating window with a length equal to the odd value nearer to half sample rate (i.e., 475). This filter is typically used to smooth noisy signals, such as electrical ones. Unlike the classic low-pass filters, this filter does not cut all the high frequencies, leaving the information intact. The filtered signals were also de-noised using an orthogonal wavelet with 5 levels of decomposition. A *Penalized Contrast Function* [144] was used to identify the location and the variance change points.

Moreover, all the physiological data were normalized between [-1,1]. Let y a generic physiological signal:

$$\forall y_i \in y : \; y_i = -1 + 2 * \frac{y - min(y)}{max(y) - min(y)} \tag{6.1}$$

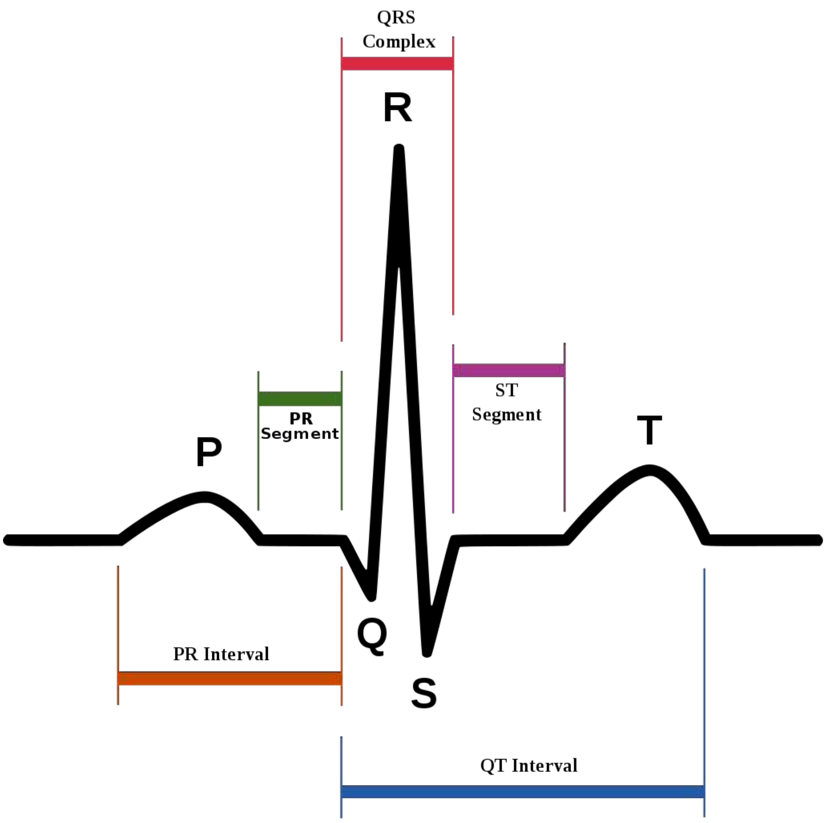


FIGURE 6.6: A schematic representation of the fiduciary points in an ECG. The central area is the QRS complex, which is usually used to acquire the HR information. The figure has been produced by Anthony Atkielski on Wikimedia commons.

The self-assessment data was acquired in a range between 0 and 100. During data filtering, we also centered and scaled this signal in a range between [-1,1] ($cse = {}^{(e-50)}/_{50}$, where e is a generic emotion tagging information).

6.2.2 Considered Features

For the validation, we considered to study the participants' emotions with a precision of half second. Thus, the features extracted by RAW signals and the self-assessment data were compacted.

In order to acquire the players' Heart Rate (HR) during the game session, we analyzed the ECG signal over all the game session. The ECG data presents three signal deflections that repeat over time and occur in rapid succession. They correspond to the depolarization of the heart ventricles. The name of these signal deflections is *QRS complex* [145], where Q wave is a downward deflection, followed by R wave, an upward deflection, and in turn followed by a second downward deflation, the S wave (see Fig. 6.6). The distance between two equal points in two repetitions of QRS complex provides the necessary information to calculate the HR [146]. Usually, the R wave is used for this purpose, thus we detected the RR interval on ECG signal using OSEA algorithm [147]. The algorithm returns the points where the QRS waves are located, thus we selected only the fiduciary points of R waves. Usually, during the ECG, the clinical staff ask to reduce the movement in the specific locations where the electrodes are placed; however, this is not applicable to our experiments because the participants should not have movements limitations in order to better interact with the games. A second and more in-depth analysis of the ECG was performed to find the motion artifacts (or signal zones where the QRS can not be detected correctly), and, in order to have an approximation of RR intervals, we applied a data interpolation placing a dummy RR peaks on motion artifacts.

Then, each physiological signal (except for the ECG) was sampled with a range of half second and for each subsample the values between 1Hz to 180Hz in the frequency domain (using FFT) were calculated. Lastly, the self-assessment data (30 data/sec) were subsampled in order to have the same length of the other signals (2 data/sec).

Summarizing, the features (583) used for predicting the arousal and valence values are: row data of each physiological signal (avoiding ECG), HR, and the magnitude information of each physiological signal considering only the frequencies between 1Hz to 180Hz. Albeit most of these features are not designed to be specific for our physiological data, they were selected in order to understand if the overall framework is able to predict the players' emotions.

6.2.3 Features Selection and Supervised Learning in Pilot Study

In order to identify the most informative features, we randomly split our sample into two different groups called TRs and TE. TRs group is composed of 8 participants (a matrix with 40954 instances and 583 features) selected to train the ML algorithm; this group was used to determine the most informative features. TE group is composed by only the remaining participant (a matrix with 6510 instances and 583 features), and it was used to test the features extracted by the ML algorithm trained on TRs. In order to have a robust feature selection, we randomly created 50 different permutations of a subset of 6 subjects extracted in TRs. The data of these participants were merged and we trained a Random Forest (RF) [148] on them in order to predict the self-assessment data of valence and arousal, labeled by the users as described in Sec. 6.1.4. The RF creates a bagging of decision trees [149], and it is able to extract an index of importance for each feature as described in [150]. After each training, the importance of each feature was stored obtaining a matrix $M = R \times C = 50 \times 583$. For each row, we extracted the most important k features, where k is an incremental iterator from 1 to C. Among these features, we selected those available in all permutations (R), and we extracted the features from the joined data of TRs (Algorithm 1).

$T \leftarrow user_1, user_2, ..., user_n$
$r \leftarrow randomIntegerbetween[0, n-1]$
$TRs \leftarrow T[-r]$
$TE \leftarrow T[r]$
$arousalFeatures \leftarrow Matrix(50, C)$
$valenceFeatures \leftarrow Matrix(50, C)$
for $i = 0$ *to* 49 **do**
  $aroTrain \leftarrow$ model of RF using arousal as target variable
  $arousalFeatures.addRow(aroTrain.FeaturesImportnace)$
  $sortedTRs \leftarrow shuffle(TRs)$
  $toTrain \leftarrow sortedTRs[1:6]$
  $valTrain \leftarrow$ model of RF using valence as target variable
  $valenceFeatures.addRow(valTrain.FeaturesImportnace)$
end

Algorithm 1: Robust Feature Selection Algorithm

For each k, we trained a RF using 1/3 of features subset for each decision split and with 128 trees as suggested in [151]. The obtained hypothesis was used to predict the self-assessment values of user TE. Thus, we saved the Root Mean Square Error (RMSE) index calculated between the estimated data and the original target variables. After all interactions, we acquired two sets of C RMSE values, one for arousal values and one for valence. On these vectors, we selected the features that minimize the RMSE value, as shown in Algorithm 2.

Using only the most important features, we trained two new RFs with the same parameters described before. The first model was used to understand if the framework is able to predict the players' emotion starting from the self-assessment data provided by the other participants. Thus, we trained a RF using a Holdout method: we performed the training on TRs and we tested the data on TE. The second model was developed to understand if the framework is able to predict the labeling data using the self-assessment information provided by the participant herself. In this case, the self-assessment data were more homogeneous, however, the number of instances used to train the ML algorithm are significantly reduced. Thus, we used a Cross Validation (CV) with 10 folds, considering the physiological data and the self-assessment information provided by the participant. It uses 9 folds for training and the last fold for evaluation. The process is repeated 10 times, leaving one different fold for evaluation each time. A summary of the methodology used for supervised learning is shown in Figure 6.7.

```
//dataFeatures can be both arousalFeatures and valenceFeatures
C ← numberOfColumn(dataFeatures)
RMSE ← float(C,1)
featuresList ← list(C,1)
for k = 1 to C do
    foreach row in dataFeatures do
        row ← { 1  for the most important k features
              { 0  otherwise
    end
    indexes ← bool(1,C)
    for c = 0 to C-1 do
        column ← dataFeatures.getColumn(c)
        if sum(column) == numberOfRow(dataFeatures) then
            indexes[c] ← true
        else
            indexes[c] ← false
        end
    end
    tr ← train RF on TRs data considering only indexes features
    pr ← predict using atr hypothesis on TE data
    rmse[k] ← RMSE(pr, TE.targetVariable)
    featuresList.append(indexes)
end
minRMSE ← min(rmse).getPos
features ← featuresList[minRMSE]
return features
```

Algorithm 2: Global Minimum Error Selection

6.3 Results

In order to consider the framework as a valid tool, we analyzed three different factors: the acquired self-assessment data, the most important features, and the ability to predict the players' emotions during game fruition.

To analyze the annotated data, we considered the emotion self-assessment mapped on the 2-dimensional space. In Fig 6.8a, we show the distribution of the emotion tagging values of each participant. As shown in the figure, the acquired self-assessment data are unbalanced, since the most of the assessed emotion data are distributed on positive arousal values. It may be due to the genre characteristics, since platform games require to be focused and precise during the game task. Furthermore, the self-assessed values of valence show an average tendency to neutrality, covering, however, almost all the values. As shown in Fig. 6.8b, these considerations are in-line with the values provided by user TE.

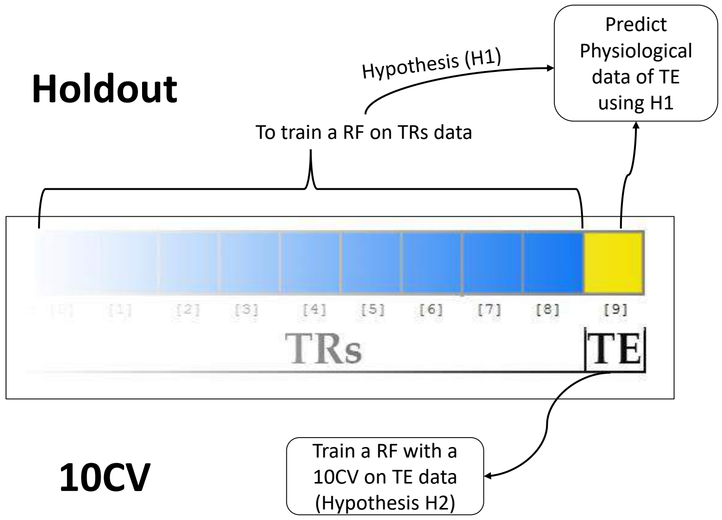
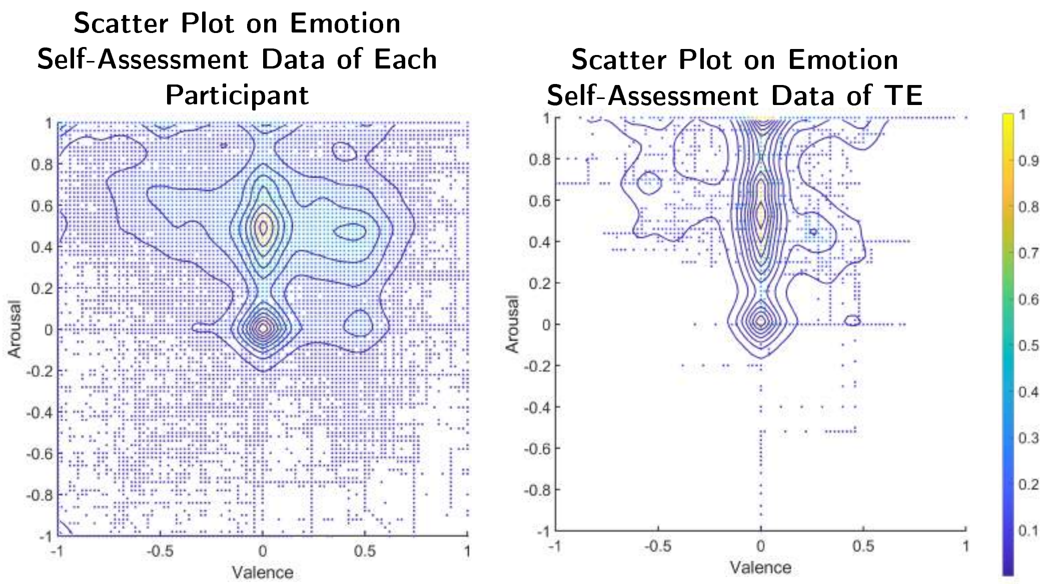


FIGURE 6.7: The general prediction model is shown in the top side of the figure. It uses the information of a set of users in order to predict the information provided by another participants (Holdout). The method of validation used to study the individualized hypothesis (10CV) is shown in the bottom area.



(A) Overall emotions tagging distribution on each participant of validation experiment

(B) Emotions tagging distribution of participant TE

FIGURE 6.8: The figures show the distribution in the 2-dimensional model of the participants' emotions identification during the validation experiment

The process of feature selection has obtained different results for arousal and valence (see Fig. 6.9). The number of features required to minimize the arousal RMSE value is quite large (i.e.,

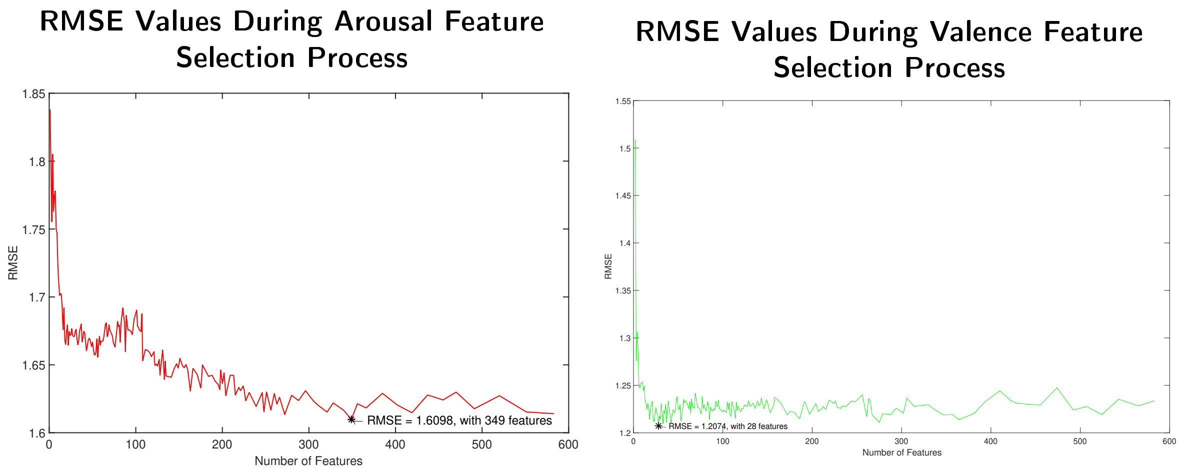


FIGURE 6.9: The figures show the number of features which minimize the prediction error. These results have been provided using algorithm 2

349). In contrast, valence uses only 28 features in order to obtain the lower prediction error.The variables used for the ML process are shown in Tab. 6.1, while a table with a complete overview of the feature selection process and, as a consequence, the selected variables at each step of the Algorithm 2, is available on GitHub[5].

Summarizing, our analysis produced two ML models:

- A general model, using a RF trained on the joined data of the 8 participants TRs (i.e., Holdout).
- An individualized model, using a RF trained using a 10CV on the TE.

Fig. 6.10 illustrates the predictions data trends on the self-assessment values during the TE game session. The Plots 1 and 2 show the predictions trends on the data trained on TRs. Albeit the RMSE values (1.6 for arousal and 1.2 for valence) do not indicate a statistical significance of the results, we can, however, observe from the graphical plots that the values remain in the emotional range provided by the user TE. Predictions represented in Plots 3 and 4, trained with the CV on TE, follow coherently the data labeled by the user TE through self-assessment (Sec. 6.1.4) obtaining an RMSE index, respectively, equal to 1.0 and 0.8. Also in this case, due to the users variability in the self-assessment methods, with the consequent introduction of several peaks in the evaluation data, the observed RMSE index does not provide a significant information. Although

[5]https://github.com/grano00/EmotionsAnalysisInVideogames

TABLE 6.1: This table summarizes the features that the algorithm revealed most important in order to predict the test user data. The rows with "F_*" contain the values on the frequencies of the corresponding signal. In valence column, the frequencies values considered as informative are only in EMG1 and EMG3. The other information in frequency domain were not considered informative according to the feature selection algorithm.

	Arousal	**Valence**
Heart Rate	yes	yes
GSR	yes	yes
Resp.	yes	yes
EMG1	no	no
EMG2	no	no
EMG3	no	no
EMG4	no	no
F_EMG1	All frequencies	1.9 - 11.2, 14.9 - 35.4, 39.2, 42.9, 57.9
F_EMG2	1.9 - 7.5, 39.2, 44.8, 57.9 - 61.6, 78.4, 80.3, 93.3 - 98.9, 106.4, 110.1, 115.7, 130.6, 145.6, 149.3, 153, 177.3	no
F_EMG3	1.9 - 11.2, 14.9 - 85.9, 89.6, 93.3 - 104.5, 108.3 - 113.9, 121.3, 123.2, 126.9 - 151.2, 154.9 - 169.9, 173.6, 175.5	1.9 - 7.5
F_EMG4	1.9 - 9.3, 14.9, 18.7, 20.5, 24.2, 28, 33.6, 35.5 - 41, 42.9, - 57.9, 63.5 - 69, 76.5 - 84, 87.7, 91.5, 93.3, 98.9, 102.7 - 108.3, 119.5, 132.5 - 136.3, 145.6, 154.9 - 160.5, 164.1, 171,7	no
F_GSR	1.9 - 5.6, 31.7 - 44.8, 50.4, 52.3, 56, 57.9, 61.6, 65.3 - 70.9, 84, 85.9, 91.5, 93.3, 98.9 - 102.7, 110.1 - 113.9, 119.5 - 127, 130.7 - 134.4, 138.1 - 145.6, 149.3,153, 162.4 - 166.2,169.9, 177.3, 192.2	no
F_Resp.	1.9 - 22.4, 26.1, 31.7, 35.5, 44.8, 46.6, 50.4 - 54.1, 57.9 - 61.6, 72.8, 78.4, 80.3, 100.8, 104.5, 106.4, 119.5, 125, 130.7 - 134,4, 138.1, 143.7, 154.9 - 166.1	no

the prediction data do not follow the peaks in the self-assessment values, they follow the pattern identified by the user TE. The prediction of the cross-validation analysis on TE is more accurate than the prediction of RF trained on TRs, since the self-assessment data can be considered user sensitive.

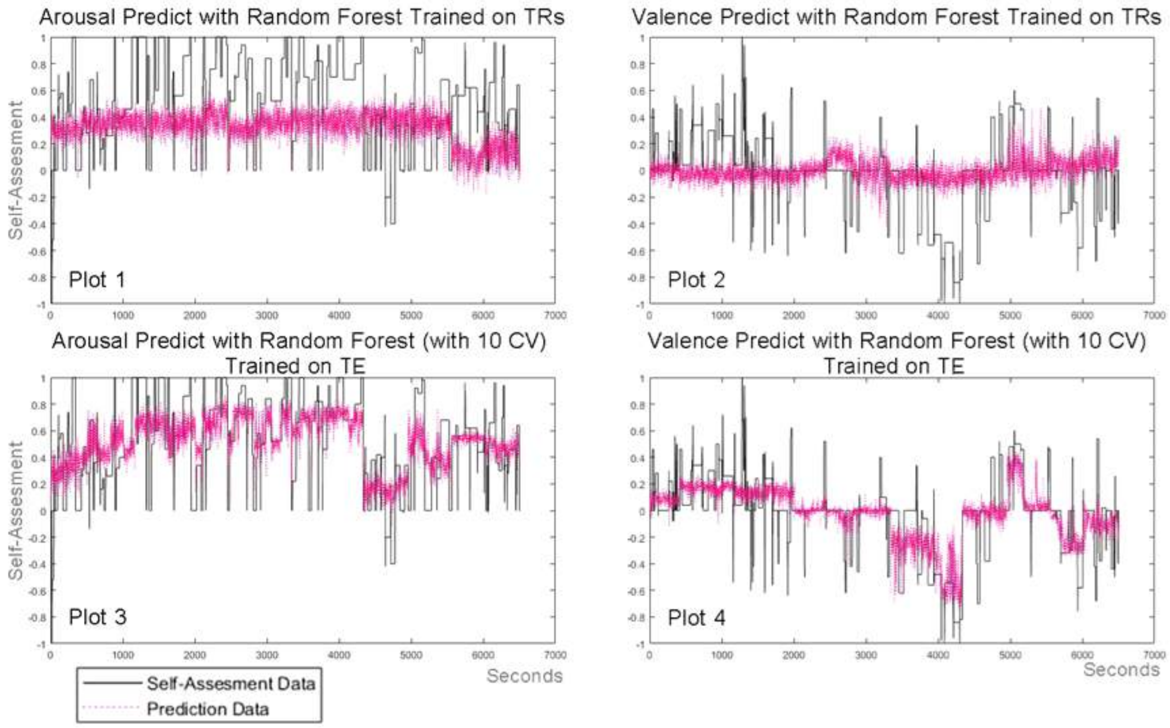


FIGURE 6.10: The figure represents the mapped value of TE on emotional assessment. The time of each game session (using half second as step unit) is shown on the x-axis, the emotional assessment values are shown on the y-axis. The prediction values of the first two plots are defined by a Random Forest trained on TRs, while, the remaining two plots illustrate the predictions of a Random Forest trained with a 10CV on TE.

6.4 Pilot Study Outcomes

Summarizing, in the pilot study, we performed a set of experiments on participants involved to play at 4 different platform games. Thus, we acquired the physiological data and the self-assessed emotions evaluation through the tools presented in the Ch. 5. We also analyzed the data, filtering the noise on physiological signals, considering a set of extracted features starting from RAW data with a precision of half-second, selecting the most important features, and predicting the emotions of a random participant. We also proposed two different methodologies to predict the players' emotions: the fist considers a model trained on a different group of subjects (Holdout), and the second applies a 10CV on the participant herself.

Albeit the results are not particularly significant, our study shows that is possible to predict the players' emotions during video games sessions using the considered types of physiological data and the emotions self-assessment values. Moreover, we hypothesize that collecting additional data on a larger sample of users, and through the support of external annotators, the difference between the two predictions can be attenuated.

During the experimental sessions, we found technical problems and a poor signal quality on the EMG placed on *Depressor Labii Inferioris* (*EMG6*), especially in men with beards. Moreover, the electrodes of *EMG6* were placed in an area not covered by gaming devices, like, e.g., a VR headset. For these reasons, we decided to no longer consider the electrical activity produced by this muscle in future analysis.

In conclusion, we consider the described approach as a potentially valid starting point for the developing of a framework to analyze the players' emotions. In order to develop a more effective framework, we designed a set of possible improvements, starting from a full validation with a larger number of subjects. This extension permits to have a large database of participants, which may improve the accuracy of the data analysis and, consequentially, the prediction outcome. Moreover, a study on a larger sample may improve the significance and the reliability of the overall experimental outcomes. The experiment results suggested critical aspects, which have been thoroughly investigated in the extension of this work. In particular:

- an extension of the experiment using a different game category. As shown in Sec. 6.3 the selected platform games have been not able to cover all the area of the 2-dimensional emotional space. Moreover, the acquired self-assessment values are strongly unbalanced
- an improvement of the overall data analysis, starting from the quality of the signals, the types of extracted features, the feature selection algorithm, and the supervised learning technique
- the support of a VR headset, in order to improve the players' immersion, which may be able to arouse different type of emotions. In addition, we can compare the two game fruition modalities
- a new design of the experimental procedure, reducing the required time of the overall experimental session, and, as consequence, improving the participants' performance, and minimizing the participants' bias during the emotion labeling.

Chapter 7

Creation of RAGA Dataset and Data Analysis

THE pilot study provided interesting results which seem to suggest the adequateness of our framework to infer the players' emotions during video game fruition. Moreover, it gave the opportunity to identify technical issues and to consider a set of improvements as discussed in Sec. 6.4. In particular, we decided to modify some characteristics of the signals acquisition, of the experimental procedure, and of the data analysis. Thus, through a larger set of experiments and considering a different game genre (i.e., racing games) in the experimental setup, we collected the RAcing GAme (RAGA) dataset, a dataset of physiological data acquired during video game fruition using both a standard monitor and a VR headset.

Part of the content of this chapter were used for the paper "Granato M., Gadia D., Maggiorini D., and Ripamonti L. A., An Empirical Study of Players' Emotions in VR Racing Games Based on a Dataset of Physiological Data", submitted to the journal "*Multimedia Tools and Applications*".

7.1 Experimental Setup Improvements

A common problem that affects the analogical signal acquisition is the noise provided by the electric hum of AC current. A solution to avoid this noise is to put a band-stop filter at the electric hum frequency in cascade to the Arduino, or to remove the noise in post-processing, as we have done in our previous data analysis. However, this approach deletes the information at (and often near) the hum frequency. Therefore, we decided to completely isolate the device and sensors from the AC current, feeding the board with an external battery. Moreover, we avoided the computer current implementing a wireless connection using an Arduino module to convert serial data to Bluetooth (HC-06). It simulates the serial communication using the Bluetooth protocol instead of the cable connection. This solution also provided a greater flexibility to the overall experimental setup, giving the ability to place the sensors not close to the computer.

Thus, for ECG and EMG signals acquisition, we involved six Olimex-EKG-EMG shields. One Olimex was used to acquire the ECG data connecting three disposable electrodes Fiab F9079/100 (36x40 mm) on clean skin (Fig. 7.3c). Like the previous experiment, we followed the Einthoven triangle guidelines [121], placing two electrodes on both wrists and one at the left ankle.

The other 5 Olimex sensors were used to collect the data on 5 different areas on the right side of participants' face using disposable electrodes of size 32x32 mm, (Fiab F9053N) as illustrated in Fig. 6.2: on *Zygomaticus Major* (EMG1), *Corrugator Supercilii* (EMG2), *Nasalis* (EMG3), *Orbicularis Oculi* (EMG4), and *Temporalis* (EMG5) muscles. We used again a common electrode connected to the forehead (near the hair border) as a reference. Even if the Fiab F9053N electrodes are designed for the pediatric use, their reduced dimension particularly fits with the experiment requirements. Using these electrodes, we were able to better cover better the participant's face, without artificially altering their shape (see Fig. 7.3b). Moreover, they were equipped with cables, avoiding the direct connection of the terminal part of the EMG cables to the electrodes. This solution reduced the vision occlusion and it improved the general comfort of the experimental setup. The terminal cables and the electrodes are shown in Fig. 7.3a.

We also changed the sensor used to acquire the respiration signal, adopting an analog thermistor (NTCLE203E3 SB0). The base of the sensor was isolated using insulating tape and it was placed avoiding contact with the user's skin in order to limit the noise involved with the epidermal temperature. Thus, when the user exhales, the temperature under the nose area rises and, as a consequence, the tension information acquired by the sensor is reduced (vice versa when she inhales). The thermistor has an accuracy of ± 0.5 C° in a range between 25 C° and 85 C°, as declared by the manufacturer[1]. Removing the digital input of the previous sensor, we stabilized the communication frequency.

In addition, we collected the information about the light presented in VR headset using a photoresistor (GL5516)[2]. This signal is actually not used during the data analysis, however, it can be an interesting variable for a future study.

The GSR signal was acquired by placing two small electrodes (Fiab F9053N) on two distal phalanxes of the left hand. Usually, the players control racing games with the left area of a

[1] Datasheet: `http://www.vishay.com/docs/29118/ntcle203.pdf`

[2] Datasheet: `http://en.nysenba.com/upfiles/file/LDR.pdf`

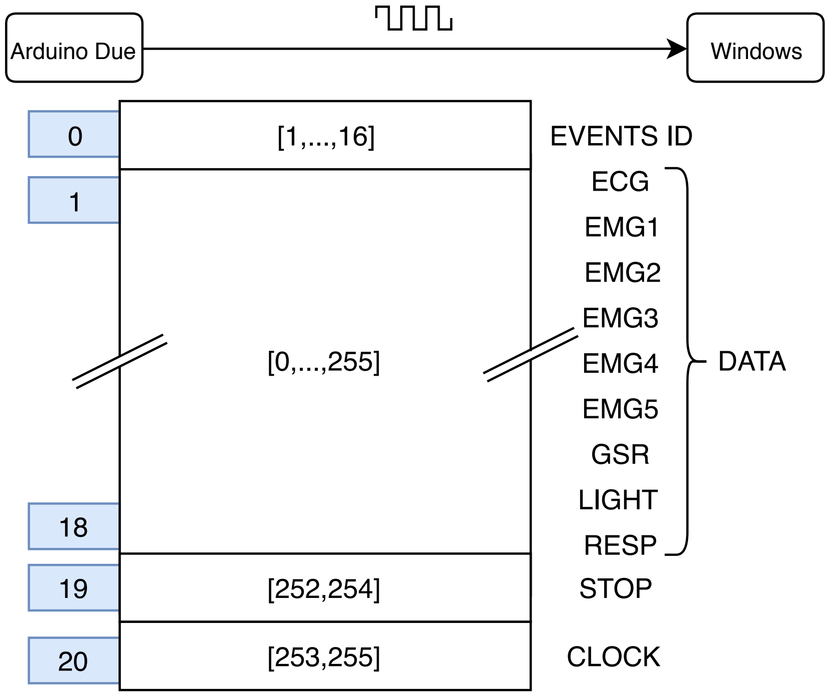


FIGURE 7.1: Bytes communicated by the Arduino to the Computer. It involves a buffer of 21 bytes

gamepad, using the left-hand thumb to steer through an analog joystick and the left index to brake using a trigger. Thus, we connected the electrodes on two fingers, middle and ring, which usually are not used to control racing games.

During the gaming session, the video of the player face and the gameplay was acquired at 60fps.

We structured a new buffer of 21 bytes (Fig. 7.1), each one transmitted at a frequency band of 115200, without parity control and with 1 stop bit (115200/8N1). The analogical data were collected using the Arduino Due that uses its built-in ADC to perform a 12-bit and the communication buffer was transmitted at a frequency of 556Hz. All the sensors received a tension of 3.3V directly by the Arduino Due, thus the ADC step is equal to 805.66 $\mu V/bit$. An architecture overview is presented in Fig. 7.2.

7.2 Motivation of Dataset Construction

Starting from the overview of the affective datasets provided in Sec. 3.1, we considered that the reason for the creation of a novel physiological dataset is given mainly to the lack of affective data regarding video game players with annotations in a dimensional space. In fact the most common

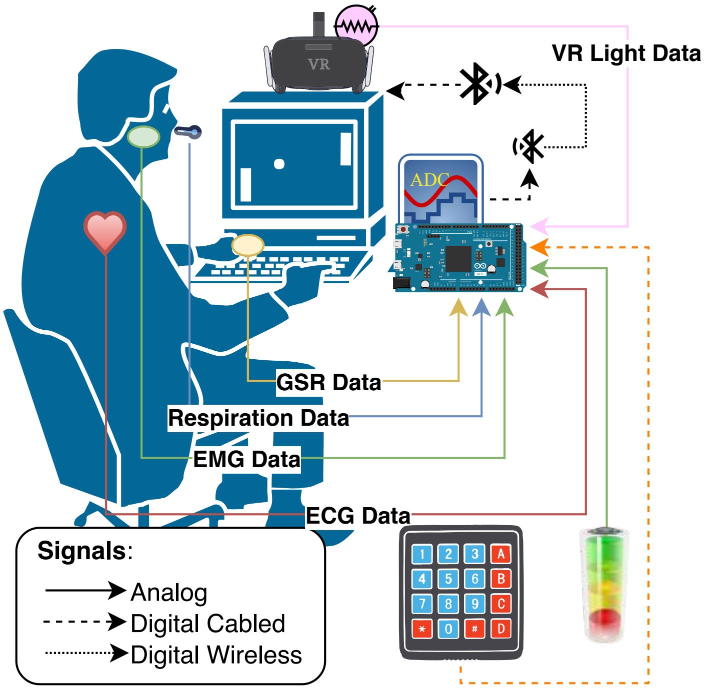
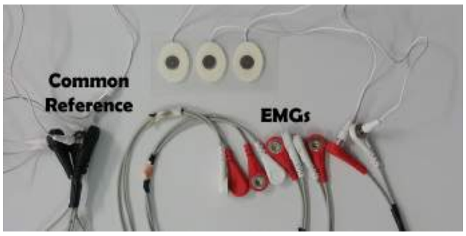


FIGURE 7.2: Improvements of the hardware architecture used to acquire physiological information for collected RAGA dataset

available datasets [75, 77, 81, 82, 87, 83, 89, 90, 61, 93] are mainly based on different kind of stimuli, like, e.g., video, images, etc. Summarizing, they provide information on the physiological signals acquired from the subjects, the method of the emotions annotation (self-assessed or using external annotators), and the representation of emotions in a 2D or 3D vector space. At the best of our knowledge, RAGA is the first freely available dataset, developed for academic research, based on the use of racing video games as stimuli, which provide annotation in a continuous dimension: in particular, we considered game sessions based on a standard monitor and a VR headset. Table 7.1 presents a comparison of the features (i.e., kind of stimuli, physiological data acquired, type of emotions identification) between RAGA and part of the above-mentioned datasets.

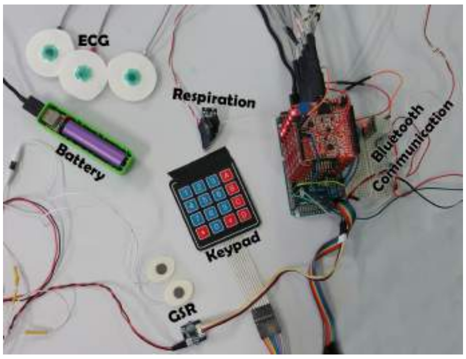
7.3 Dataset Acquisition

In this section, we describe the steps followed to acquire RAGA. In particular, we present some information of the participants, the methodology used to select the appropriate games, and the detailed procedure used during the experiments. We conclude the section providing some pratical consideration. The different steps used to acquire and analyze RAGA are summarized in Fig. 7.4.



(A) EMG terminal electrodes. The black *snap* of black wire was connected together at a common reference, while the other couples of *snaps* (red and white) were placed on the face skin over the muscles that you want to consider.

(B) Electrodes placement to acquire the facial EMGs, and respiration sensor placed under the participant's nose



(C) Overall hardware architecture and sensors used to infer physiological data during the experimental sessions.

(D) Photoresistor used to collect the light on the VR headset lens

FIGURE 7.3: Photos of the hardware architecture and example of electrodes/sensors placement used to collect the physiological and events data for RAGA

7.3.1 Participants

The group of participants was composed of 33 players (29 males and 4 females), between 18 and 40 ($\mu = 24.66$ and $\sigma = 5.15$) years old. The recruitment was performed in different classes of the Computer Science department at the University of Milan, through social networks, and through leafleting. They usually play video games 3.6 days per week in average ($\sigma = 2.16$). More than half of the participants reported their game sessions to be longer than 2 hours. The 60.00% of the participants had already tried a VR headset, while the 13.33% had already used it in a gaming application. Moreover, the 57.14% of the participants declared to reach better

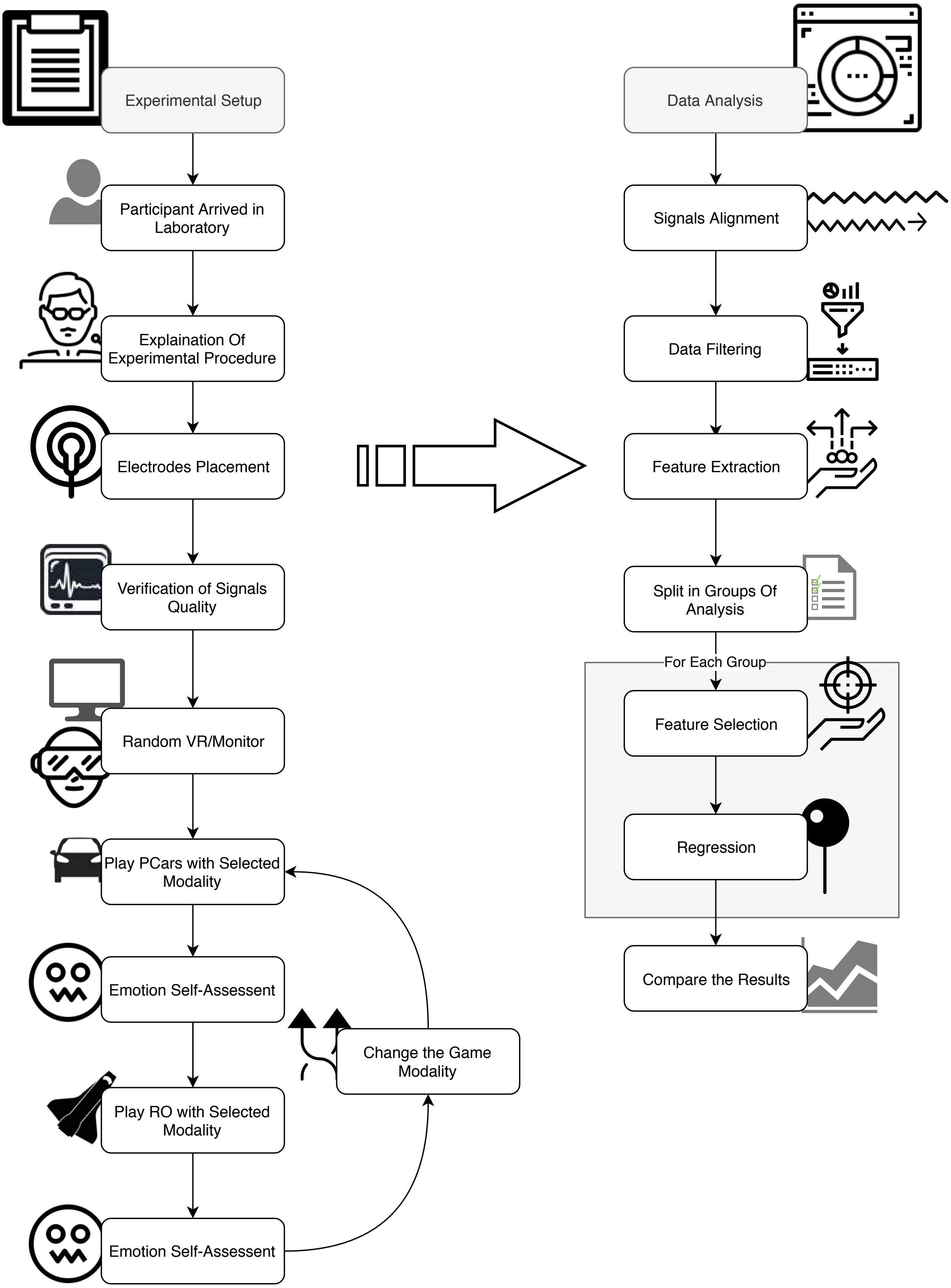


FIGURE 7.4: RAGA Flowchart. On the left, the figure shows the different steps involved during the experimental setup and the database acquisition, while, in the right, it illustrates the main steps involved in the data analysis.

performance during video game fruition in VR rather than standard monitor. All participants were Italian and volunteers, and they had not received any monetary or academic contribution for the experiment.

7.3.2 Considered Games

In the previous work, we used as stimuli four platform games. However, they resulted not adequate enough in order to cover all the VA dimensional space (see Sec. 6.3). It may be due to the characteristics of the considered media, which, usually, was designed to keep the player focused. Moreover, a level of a platform game may be too difficult and long for players without an appropriate skill. As a consequence, one of the first decision to take, designing the creation of a physiological dataset based on video games, is which kind of game is the more adequate for the final purpose. Focusing only on the influence of the game mechanics, and avoiding the emotional effect of narrative elements (like e.g., cinematic sequences), we decided to consider racing video games. In racing video games, the users start from a point A, and they must arrive at a point B, usually driving vehicles, in the shortest possible time or overcoming the opponents. Racing games usually have a set of possible events that can arouse players' emotions. In fact, players are involved in high-speed races where usually: accidents, overtaking, high-speed corners, etc. can occur. Therefore, we hypothesized a high variability of players emotions during a racing game and, as a consequence, of the corresponding VA values. This genre can be further divided mainly into three sub-categories:

- **Arcade Racing** games, where the priority is given to fun and feel of speed
- **Simulation** games, designed to guarantee a user experience similar to the reality (the goal of the game engine is to simulate a truthful vehicles physics)
- **Kart** games, based on a simplified driving mechanic characterized by features that usually do not appear in other racing sub-genres (e.g., obstacles, weapons, possibility to jump, power-up, etc).

We decided to consider 2 racing games from different sub-genres. We defined a set of constraints for the games' selection: they should have an intuitive and straightforward game mechanic and environment; they must have comparable input controls and level length; they can be played either in VR or using a standard monitor. Albeit the VR in the video game field is a novel technology, we are interested to study the emotional impact of these devices on players in order to provide also a novel contribution in this research field. Moreover, we considered to collect the data on both peripherals (i.e., VR headset and monitor) in order to have a comparison and, at the same time, to keep the generalizable information provided by the standard monitor. Thus, we selected a simulation driving game, Project Cars[3] (*PCars* from now on), and an arcade driving game, RedOut[4] (*RO*). PCars, a game released to the market in 2015, is developed by *Slightly Mad Studios* and published by *Bandai Namco*. It is a simulation game, where the drivers are involved in races on virtual reproductions of existing cars that compete on famous circuits. RO is a futuristic racing game developed by an Italian company (*34BigThings*) and released in September 2016. It is an independent game where futuristic shuttles compete in full acrobatics tracks. The main inputs for the above-mentioned games are quite simple and symmetrical: a steering input, an input to accelerate, and an input to brake. RedOut (RO), due to its arcade nature, has two other inputs: the first dedicated to control the shuttle inclination, and the second used to activate the turbo speed. Both games can be played from the driver point of view, with or without a VR headset, and they have received positive scores by the critic (i.e., 83 for PCars, and 81 for RO on Metacritic).

7.3.3 Experimental Procedure

Upon arrival in the laboratory, the participants were invited to sit on a comfortable chair. They were informed about the experimental procedure and they were invited to read and sign an informed consent, and a permission to use the data and the video recorded during the experiment for research and academic purposes. The acquired data were collected anonymously, and we assigned, at each participant, an incremental identifier for the analysis and for future experiments. The participants with ID lower than 10 are the same users who participated in our pilot study (see Ch. 6). Lastly, each user performed the test in a single daily session. As in the previous experiment, to attenuate the *observer-expectancy effect* [141], the staff member was positioned in a predefined area out of the participant's field of view in each phase of the experiment. She also avoided any kind of interaction or observation during the different experimental phases. Moreover, we asked the participant to reduce the contacts with the staff member, and to report only technical problems.

The experiment consisted of three stages: electrodes placement and test presentation, main test, and a final survey. We connected the electrodes and the thermometer on the participant's skin, taking care to not bother her view or attention. After the electrodes placement, we powered the Arduino Due and, as a consequence, all the sensors. After checking that all the sensors were working, we controlled the signal and communication quality asking the participants to perform facial movements, in order to evaluate the EMGs signals, and to have a deep breath to check the GSR signal.

[3]https://www.projectcarsgame.com
[4]https://34bigthings.com/portfolio/redout/

TABLE 7.1:
Comparison among available affective datasets and RAGA dataset. The physiological signals considered are: EEG = electroencephalography (with the number of channels), ECG = electrocardiogram, BVP = blood volume pulse, GSR = galvanic skin response, Facial EMG = electromyography placed on participant face (with the number of muscles considered), Resp = respiration, Temp = temperature, Gaze = eye gaze tracking. The last 3 columns define the type of emotion identification, where: ES = Emotion Space, in type column D = Discrete, C = Continuous, and in Annotator column S = Self Report, E = External Report with, in brackets, the number of annotators.

Dataset	Stimulus	Subj.	EEG	ECG	BVP	GSR	Facial EMG	Resp	Temp	Gaze	ES	Type	Annotator
Eight-Emotion Sentics Data [77]	Sentic [80]	1	-	-	✓	✓	1	✓	-	-	8-D emot.	D	S
MAHNOB-HCI [83]	Video, Images	30	32 Ch.	✓	-	✓	-	✓	✓	✓	PAD	D	S
DEAP [87]°	Music Video	33	32 Ch.	✓	✓	✓	1	✓	✓	-	PAD + Liking	D	S
RECOLA [89]	Collaborative Work	16	-	✓	-	✓	-	-	-	-	VA	C	6(E) + S
DECAF [90]*	Movie Clips/Music Video	30	-	✓	-	-	-	-	-	-	PAD	D	S
OPEN_EmoRec_II [61]	Mental Puzzles	30	-	-	✓	✓	2	✓	-	-	VA	D	4(E) + S
AMHUSE [93]	Movie Clips	36	-	-	✓	✓	-	-	✓	-	VA	C + D	4(E) + S
RAGA	Racing Video Games in VR	33	-	✓	-	✓	5	✓	-	-	VA	C	S

° *The dataset provides also the signals of: EMG on Trapezius muscle, and EOG to investigate the eye movements*

* *The dataset provides also the signals of: EMG on Trapezius muscle, MEG in order to measure the brain activity, and EOG to investigate the eye movements*

(A) Screenshot of a game session with PCars

(B) Screenshot of a game session with RO

FIGURE 7.5: Screenshots of different game sessions with both games.

A video with the demo of the games, the tracks, and the vehicles were also shown to the participants. This preliminary phase were performed in order to provide at each participant a common knowledge of the game environments, and a brief overview of the structure of the level. For all the gaming session, we selected a *McLaren, 12C* in the *California Highway Stage 2* track for PCars (Fig. 7.5a), while we used *Asera, Yoshinobu* shuttle on *Alaska, Airbone* on RO (Fig. 7.5b). Furthermore, a member of the laboratory staff explained the game mechanics for both games and he made sure that the participant understood how to interact with the games.

ESAT (see Sec. 5.3.2) were presented to each participant, and a short training to familiarize with the input system have been conducted. The assessment software shows to the participant a video with the information of the player's face, a video of the gameplay, and a data synchronization graph. We asked the participants to identify their emotions during the entire video playback using the SAM/AS tools described in Sec. 4.3.

We modified the placement in the ESAT GUI of the window with the recorded videos, placing the video of the participant's face in a more isolated and bigger area. Following the structure of Fig. 5.5 presented in Sec. 5.3.2, the three different areas of the video are structured as follows:

(c) In the bottom right area, the player's face during the game session is shown: we asked the participant to focus on this area in order to re-evoke the emotions experienced during the game session

(d) In the top right area, the gameplay useful to support the emotions recall was shown

(e) In the top left area, the acquired physiological information is shown. It was used only to synchronize, through its top and bottom bars, the collected data with the VA values acquired by DAPIS.

The *main test* were structured in two stages, each repeated two times (i.e., with and without VR): game session and emotion annotation. Randomly, the participants started the game session with or without VR: 14 participants have started with VR, while 19 have performed the first game session without VR. Each player, whether she used VR or not, played to PCars as the first game. The beginning and end of each race were synchronized to physiological and assessment data by pressing two different buttons, connected to the Arduino Due. The synchronization starts/ends at the beginning of the Arduino clock cycle. As in the pilot study, the first button pressure inserts in a specific column of the physiological dataset the value 14 (a number used only to identify the session beginning), and, at the same time, it switches the color of the two bars (e) to green. In the same way, the second button pressure was used to insert the value 15 and to switch the color of the bars to red. Thus, the two buttons were used to synchronize the physiological data with the first emotion tagging performed immediately after the game session. However, in this experiment, a member of the staff was in charge to define the start and the end of the level, permitting the participants to focus only on the game.

The second game session was performed right after the first, followed by another emotion tagging stage. Before the VR stage, the laboratory staff explained the potential risks related to the device. The possible motion sickness deriving from the VR device may arise due to a not accurate settings of the VR parameters [152, 153], or to physiological issues of the players with stereoscopic vision [154].

FIGURE 7.6: Photo acquired during the experimental session in the VR gaming stage

Lastly, we asked the participants to fill out a survey with questions about their habits and game skills, and regarding the overall experiment considerations.

Practical Consideration

Albeit the experiments were conducted on 36 people, we had not considered in the analysis the data of participants with ID 11, 12 and 15 due to technical issues. Furthermore, the participants 20, 23, and 40 experienced sickness during the VR session, although they have completed the experiment. The acquired RAGA dataset and the data analysis source code are freely available to the academic community[5].

7.4 Data Analysis

The data analysis consisted in four main steps: data filtering, extraction of features from raw data, selection of the most informative features, and application of an algorithm of supervised learning, in order to understand if these data can be used to predict the players' emotions during the game sessions.

Since the physiological data and the emotions self-assessment information were acquired at two different frequencies, it was necessary to uniform the number of instances. Thus, let a_f and v_f be respectively the values of Arousal and Valence at the instance $f \in \{1, 2, 3, .., F\}$, where F is the length of VA arrays, and let N the number of instances of their corresponding physiological data, we calculated the new points in the interval $[1, F]$ to have a length equal to N. Consequentially, let $n1$ be a set of integers $\{0, 1, 2, .., N-1\}$, the new set of points $nf = [1, F]$ is:

$$nf = 1 + n1 \odot ((F-1) \oslash (N-1)) \tag{7.1}$$

Where $\odot$ defines the *Hadamard product*, and $\oslash$ means the *Hadamard division*. As a consequence, we obtained two arrays: $f \in \mathbb{N}$ of length F, and $nf \in \mathbb{Q}$ of length N, with values in $[1, F]$. Thus, we applied a linear interpolation on each Arousal and Valence data in nf. For each element in nf, we considered the integer neighbors in f such that $f_j \leq nf_i \leq f_k$. Lastly, we generated the new Arousal (an) and Valence (vn) points for every i in $\{1, 2, 3, ...N\}$.

$$an_i = \frac{a_j(f_k - nf_i) + a_k(nf_i - f_j)}{f_k - f_j} \tag{7.2}$$

$$vn_i = \frac{v_j(f_k - nf_i) + v_k(nf_i - f_j)}{f_k - f_j} \tag{7.3}$$

[5] https://github.com/grano00/GameVRRacingPhysioDB

As a consequence, we calculated two new sets of values with the same length, and frequency, of the physiological data.

7.4.1 Data Filtering

All the physiological and emotional tagging data were separated for each game session, thus getting 4 sets of data for each participant: RO, PCars, RO in VR, PCars in VR, Thus, we filtered the groups of data separately in order to minimize the probability to introduce data patterns that may alter the future analysis. For each type of physiological information, we centered and scaled, using the standard deviation, the acquired data [155]. Let x be a set of physiological data (e.g., ECG signal):

$$cs(x) = \frac{x - \mu(x)}{\sqrt{\frac{\sum_{i=1} N(x_i - \mu(x))^2}{N-1}}} \tag{7.4}$$

where $\mu(x) = \frac{1}{N}\sum_{i=1}^{N} x_i; x \in [1, N]$.

As in the previous experiment, the emotional tagging data was initially expressed as an integer value from 0 to 100. Thus, the value 50 underlines a neutral emotion and, as a consequence, we centered the data at this value, scaling them in order to create a signal in the interval [-1,1].

Thanks to the use of an external battery and the removal of a direct connection between the sensors and the laboratory electrical system, we were able to considerably remove most of the noise present in the signal. However, we filtered some frequencies of the physiological data in order to remove the noise generated by other sources of data acquisition. Starting from ECG data, we were interested to collect the HR of the players during video games fruition. As suggested by Fedotov [156], the most informative frequencies to understand the HR are between 5 to 30 Hz. We considered an upper band frequency of 35 Hz, filtering the data using an Equiripple FIR band-pass [157], in order to consider also a possible excessive increase of heartbeats. All the EMG signals were filtered using a high-pass Equiripple FIR with the cut-off band at 20Hz as suggested in [158].

As breathing produces a low-frequency signal alteration, we applied a moving average filter [159] with a coefficient equal to $1/_{SampleRate}$. As a consequence, we obtained the average temperature, under the participants' nose, over a period of 1 second. For the GSR, we applied a 1^{st} order Butterworth low-pass filter with 5Hz cutoff. Through Ledalab [146], we also extracted the SCL and SCR information [160], and we added them to the set of physiological signals in RAGA.

Each filter was designed to be applied as zero-phase in order to minimize the differences (i.e., the phase) in the signal ends. In order to obtain a filtering with zero phase, the applications of our filters were performed in both directions, forward and backward [161]. Albeit this type of data filtering is not applicable in real time, it was applied to better synchronize the labeling data with the physiological data. In a future real-time application, where the hypothesis of a ML algorithm was already defined, a filter with linear phase can be applied without excessively affecting the overall goodness.

In order to improve the emotion recall, we suggested to the users to not stop or rollback the video. However, in some cases, some participants performed a wrong assessment, and they have quickly corrected the mistake by moving back the SAM/AS slider pointer. This action produced high-frequency noise in the final assessment data, i.e., in the VA vectors. Consequentially, in this experiment, we removed the high-frequency information applying a moving average filter with a coefficient equal to $1/Samplerate$ on both vectors.

7.4.2 Overview of Features Extraction

After the data filtering, we proceeded to extract features from the physiological data. In order to preserve the *ground truth*, we collected numerous variables following the approaches of different papers in literature . Once all the features were extracted, an automatic features selection algorithm was applied to the extensive number of variables in order to remove the noisy data (see Sec. 7.4.3).

We used a 1-second precision in order to analyze the physiological data and to predict the players' emotional states, thus the features were structured considering this constraint.

The first extracted feature was the HR, collected on the ECG signal. In order to acquire the position of the QRS complexes, we leaned on the same algorithm used during the pilot study. As mentioned in Sec. 6.2, the algorithm is not able to find correctly the QRS points in the noisy areas (e.g., on motion artifacts). Thus, we developed an algorithm able to create or to remove the fiduciary points on these signal segments, according to the length of each RR interval. As a consequence, for each RR interval (rr), we defined a score defined as:

$$rrscore = \|rr/\mu(rr)\| - 1 \tag{7.5}$$

where $rrscore = 0$ indicates a correct interval dimension, $rrscore > 0$ means an underestimation, and $rrscore < 0$ underlines an overestimation (Fig. 7.7). According to $rrscore$, we used the Algorithm 3 to generate new points in the underestimated areas, and to remove overestimated points.

foreach *rr* **do**
  $r1 \leftarrow$ r point at start to rr interval
  $r2 \leftarrow$ r point at end to rr interval
  if *rrscore >0* **then**
    for *i = 0 to rr* **do**
      add new R position as follow: $r1 + (r2 - r1) * \dfrac{i+1}{rr+1}$
    end
  else
    if *rrscore <0* **then**
      delete point at start of rrscore interval
    end
  end
end

Algorithm 3: The algorithm checks the overestimated and underestimated r points, and it corrects their positions.

In order to have RR data in a more understandable measure unit, we converted RR distances in the time domain (seconds), and we multiplied the result by 60 in order to obtain Beats Per Minute (BPM): $bpm = {}^{60 * SampleRate}/_{rr}$. Lastly, we applied a moving average filter with a coefficient equal to $1/4$ on the BPM data, and we acquired the information at second precision extracting, for each second, the average BPM value.

Considering the breath signal, we collected the information on the time passed between the inhalation and the exhalation. Thus, we calculated all the slopes (upper and lower), and the frequency between the peaks, which returns the time spent between the breaths (Fig. 7.8).

Finally, on the EMG, GSRs (Raw, SCL, and SCR), and respiration data, we extracted a set of features using a floating window. Usually, these types of analysis were directed on the central area of the approximation window, which considers also the past (left area) and the future (right area) information. However, our goal is to predict the players' emotions in real-time during game sessions, and, as a consequence, we placed the area of analysis on the right side of the 3 seconds floating window, joining the data of the present (the last second) with those of the past (the previous 2 seconds), as shown in Fig. 7.9. In literature [162, 163, 164, 165, 166, 167], some features were usually suggested for the EMG signal analysis, however, most of them can be considered reliable also for other physiological signals. Table 7.2 summarizes the overall extracted features.

Time Domain Features

Huang and Chen [164] suggested to estimate the signal power computing the integration of the raw data available in the sliding window. For example, considering an EMG signal, a resting muscle

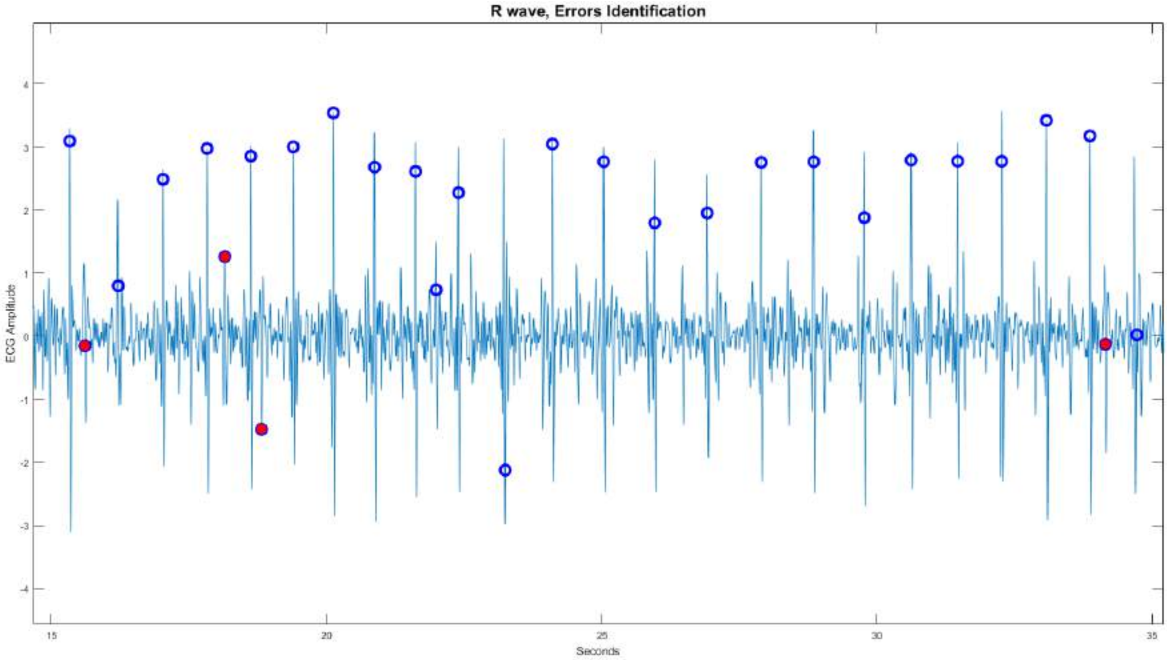


FIGURE 7.7: Example of overestimation points due to noise in ECG signal. The red points were considered as overestimated.

does not show any particular change over the time, while a contracting muscle provides changes in potential. Since the acquired signals were digitalized, we can approximate the integral as follows:

$$IntEMG_w = \sum_{i=1}^{N} |x_i| \tag{7.6}$$

where x_i considers each data of the floating window $w \in 1, 2, ..., N$.

Englehart and Hudgins [165] proposed to calculate the Mean Absolute Value (MAV) as follows:

$$MAV_w = 1/N \sum_{i=1}^{N} |x_i| \tag{7.7}$$

Starting from MAV, we also calculated two different Modified MAV (ModMAV) functions, applying a pre-processing on the original data. These modified versions smooth the signal keeping the data of the last second unchanged, in order to focus the analysis on the current second. As a consequence, $ModMAV1_w(x1_n)$ implements a *fade in*, where:

$$x1_n = \begin{cases} x_n/2 & \text{if } n < 0.75N \\ x_n & \text{otherwise} \end{cases} \tag{7.8}$$

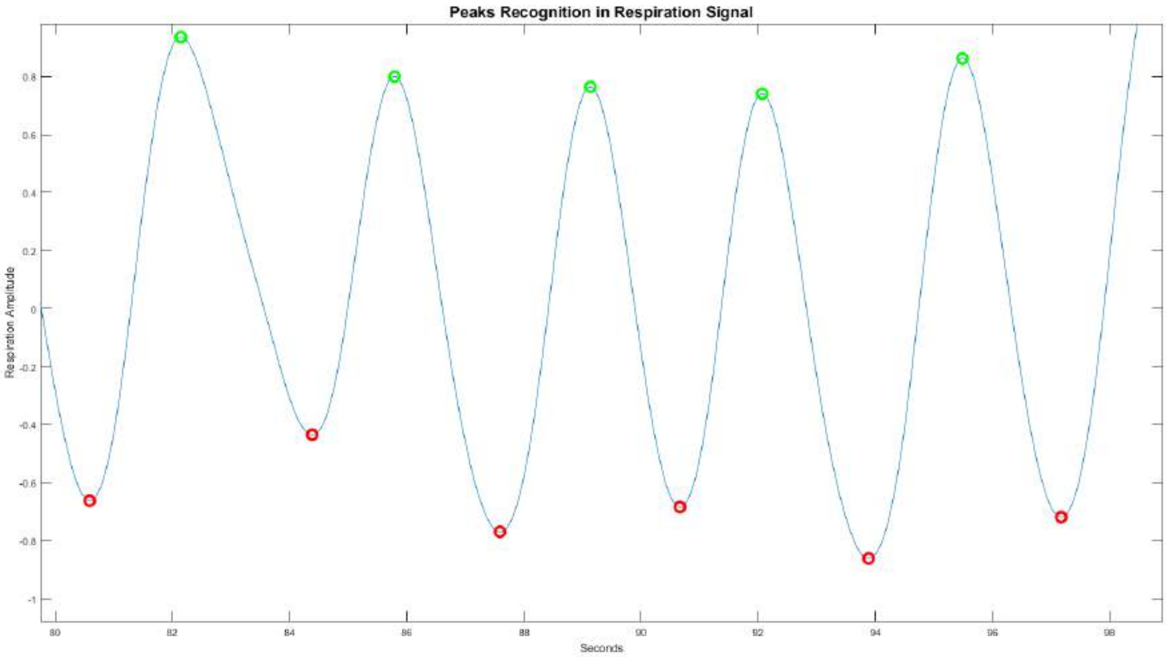


FIGURE 7.8: Respiration peaks detection, where the green circles indicate the inhalation end, and the red circles show the exhalation end.

while, the second modified function ($ModMAV2(x2_n)$) uses a more gradual *fade in*, applying an incremental weight between 0 to 1, with:

$$x2_n = \begin{cases} x_n * n/0.75N & \text{if } n < 0.75N \\ x & \text{otherwise} \end{cases} \tag{7.9}$$

Moreover, we calculated the average value only of the floating window last second. Let sr be the value of the sample rate:

$$PMAV_w = {}^1/_N \sum_{i=N-sr}^{N} x_i \tag{7.10}$$

The Waveform Length (WL) is a feature able to summarize the measures of waveform amplitude, frequency, and duration in a single parameter. Its value indicates the cumulative length of the waveform over the considered floating window. It is defined as follows:

$$WL_w = \sum_{i=1}^{N} |\Delta x_i| \quad \text{where } \Delta x_i = x_i - x_{i-1} \tag{7.11}$$

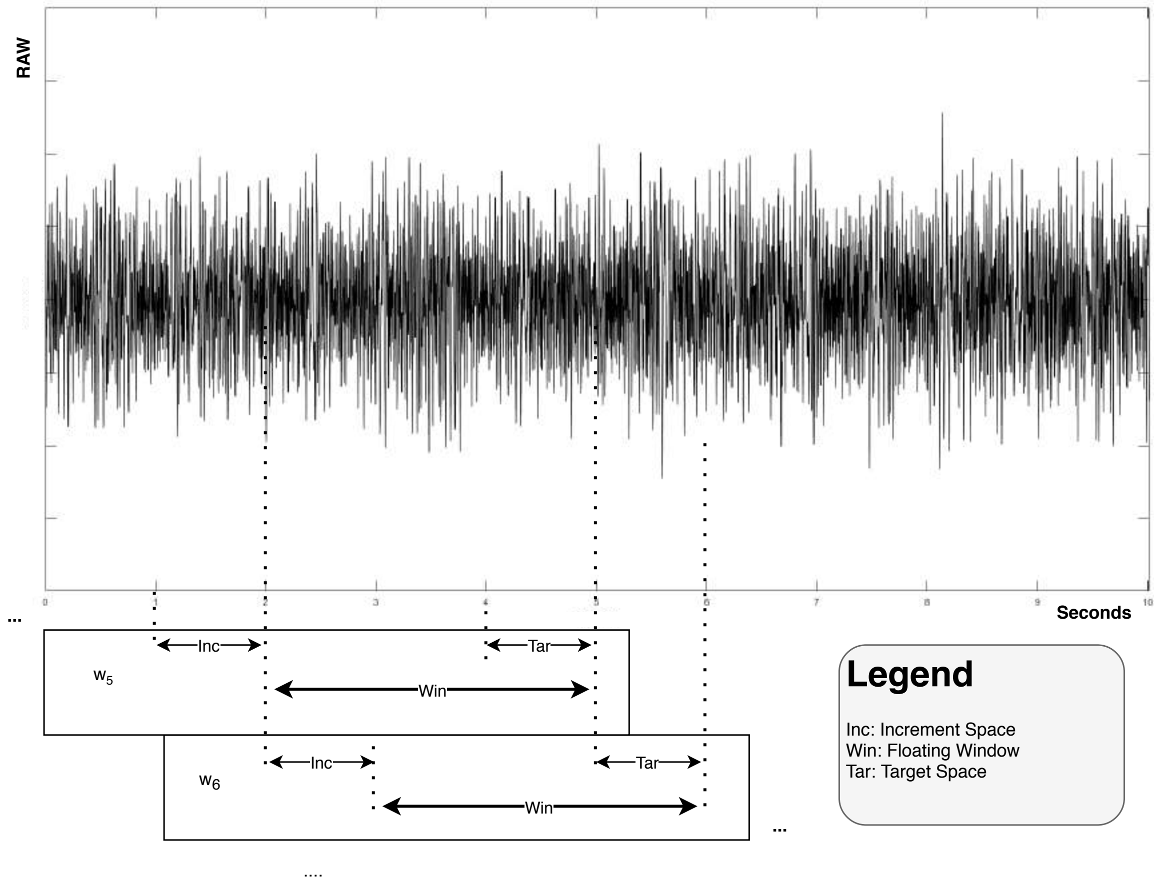


FIGURE 7.9: Floating Window, the sub-area of the window (*Win*) *Tar* is the actual second of analysis.

The Zero Crossing (ZC) is a simple measure of the frequency obtained in the time domain. Its output value was calculated counting, in the floating window, the number of times the signal crosses the zero axis. Due to an eventual signal noise, it is useful to introduce a threshold (ϵ_{zc}). The result is provided as follows:

$$ZC_w = \sum_{i=1}^{N-1} g(x_i, x_{i+1})$$
$$\text{where } g(x,y) = \begin{cases} 1 & \text{if } \{x > 0 \text{ AND } y < 0\} \text{ OR} \\ & \quad \{x < 0 \text{ AND } y > 0\} \text{ AND} \\ & \quad \{|x - y| \geq \epsilon_{zc}\} \\ 0 & \textit{otherwise} \end{cases} \tag{7.12}$$

For each type of signals, we defined specific thresholds:

- $\epsilon_{zc} = 0.5$ for the EMGs signals
- $\epsilon_{zc} = 0.05$ for the respiration signal
- $\epsilon_{zc} = 0.001$ for the GSRs signals

Another feature that provides a measure of frequency, calculated in time domain, is *Slope Sign Changes* (SSC). It calculates the number of times the signal slope changes its sign. A specific threshold (ϵ_{ssc}) is required also for this feature, in order to not accidentally increment the SSC counter due to an eventual signal noise. It is defined as follows:

$$SSC_w = \sum_{i=1}^{N-1} h(x_i, x_{i+1}, x_{i-1})$$
$$\text{where } h(x,y,z) = \begin{cases} 1 & \text{if } \{x > y \text{ AND } x > z\} \text{ OR} \\ & \quad \{x < y \text{ AND } y < y\}) \text{ AND} \\ & \quad (\{|x-y| \geq \epsilon_{zc}\} \text{ OR} \\ & \quad \{|x-z| \geq \epsilon_{zc}\}) \\ 0 & otherwise \end{cases} \tag{7.13}$$

For SSC, we defined the following thresholds:

- $\epsilon_{ssc} = 0.5$ for the EMGs signals
- $\epsilon_{ssc} = 0.005$ for the respiration signal
- $\epsilon_{ssc} = 0.001$ for the GSRs signals

In order to have another measure of the signal power, in [163] the authors suggested to include in the data analysis the signal variance (VAR):

$$VAR_w = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2 \tag{7.14}$$

where $\bar{x}$ is the mean value of the floating window.

In their paper, Phinyomark et al. [166] extended the proposed features, adding other interesting variables. Starting from the MAV, they computed the difference between the next window ($w+1$) and the current window (w). However, it does not fit with our goal to provide a real-time prediction of the players' emotions. Thus, we calculated the slopes between the windows as follows:

$$MAVSLP_w = MAV_w - MAV_{w-1} \tag{7.15}$$

In the same paper, the authors also suggested to acquire the Root Mean Square (RMS). It is a measure of signal power (comparable to MAV [168, 169]), however, "the measured index of power property that remained in RMS feature is more advantage than MAV feature" [166]. It is expressed as follows:

$$RMS_w = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2} \tag{7.16}$$

The Willison Amplitude (WAMP) calculates the number of times that the absolute value of the difference between the amplitudes of two consecutive signal samples exceeds a predetermined threshold (ϵ_{wamp}). We set the ϵ_{wamp} equal to 0.5 for each physiological signal. The WAMP function is implemented as follows:

$$\begin{gathered} WAMP_w = \sum_{i=1}^{N-1} f(|x_i - x_i + 1|) \\ \text{where } f(x) = \begin{cases} 1 & \text{if } x \geq \epsilon_{wamp} \\ 0 & \text{otherwise} \end{cases} \end{gathered} \tag{7.17}$$

SCC, ZC, and WAMP features require the definition of threshold values. As suggested in [166], the threshold of EMGs signals should be chosen between 10 and 100mV, through an accurate observation and analysis of the overall signals. In our specific case, due to the data centering and scaling performed separately for each experiment, we were not able to identify uniform voltage. This issue will be corrected in next versions, acquiring the corresponding threshold before the data filtering. Moreover, we used these features also for different types of data (i.e., GSRs and respiration), and, to the best of our knowledge, there is not a defined range of values for the threshold definition in these kinds of signals. However, we defined the threshold values through an accurate observation and analysis of the overall filtered signals.

An additional feature, proposed by the same authors, able to consider the current window energy of a signal, is the Simple Square Integral (SSI). It is calculated as follows:

$$SSI_w = \sum_{i=1}^{N} |x_i|^2 \tag{7.18}$$

Lastly, only for GSRs signals, we acquired the minimum and maximum amplitudes, the number of peaks, and the peaks average amplitudes as suggested in [170].

Frequency Domain Features

For each approximation window of the EMGs, GSRs, and Respiration signals, we collected the average spectrum power, and we also acquired the average power in different subsets of the frequency range, considering steps of 5Hz regarding EMGs, and 0.05Hz for GSRs and Respiration.

Oskoei and Hu [163] proposed also to consider the *Frequency Median* (FMD) and the *Frequency Mean* (FMN). These features were both calculated through the *Power Spectral Density* (PSD) analysis. The first splits the density of the power spectrum in two equal parts. Let M be the power spectrum length, and d_i the i^{th} value of the PSD:

$$FMD_w = \frac{1}{2}\sum_{i=1}^{M} d_i \tag{7.19}$$

The FMN indicates the mean frequency value:

$$FMN_w = \frac{\sum_{i=1}^{M} d_i \dfrac{i * sr}{2 * M}}{\sum_{i=1}^{M} d_i} \tag{7.20}$$

Two modified versions of FMD and FMN were proposed in [166], respectively named *Modified FMD* (MFMD), and *Modified FMN* (MFMN). These features consider the amplitude instead of the PSD. As a consequence, let A_i be the spectrum amplitude at frequency f:

$$MFMD_w = \frac{1}{2}\sum_{f=1}^{M} A_f \tag{7.21}$$

and

$$MFMN_w = \frac{\sum_{f=1}^{M} A_f \dfrac{f * sr}{2 * M}}{\sum_{f=1}^{M} A_f} \tag{7.22}$$

The last variable is the *Frequency Ratio* (FR) [167], it was calculated applying Fast Fourier Transform (FFT). It computes the ratio between the amplitude of low and high frequencies components:

$$FR_w = \frac{|F(.)|_{\text{w low freq.}}}{|F(.)|_{\text{w high freq.}}} \tag{7.23}$$

The authors suggested to define the two ranges through the observation of the acquired signals. Thus, we considered:

- for **EMGs** signals
  - *low freq.* [20 - 100]
  - *high freq.* [100 - sample_rate/2]
- for **GSRs** signals
  - *low freq.* [0.01 - 0.2]
  - *high freq.* [0.2 - 2]
- for **respiration** signal
  - *low freq.* [0.01 - 0.2]
  - *high freq.* [0.2 - 1]

Table 7.2: Features Extraction of Raw Data

Feature Collected without Approximation Window

ECG	BPM
Respiration	Respiration Rate (RR)
EMGs, GSRs, and Respiration	Raw Data

Features Collected with 3 sec Approximation Window

Common Features for EMGs, GSRs, and Respiration data	Band Power (BP), Power, Integral [164], Mean Amplitude (MA), Mean Absolute Value (MAV) [165], Precise MAV (PMAV), Mod MAV1 [166], Mod MAV2 [166], MAV Slope (MAVSLP) [166], Root Mean Square (RMS) [166], Variance σ^2 (VAR) [163], Waveform Length (WL) [165], Zero Crossings (ZC) [165], Slope Sign Changes (SSC) [165], Willison Amplitude (WAMP) [166], Simple Square Integral (SSI) [166], Frequency Median (FMD) [163], Frequency Mean (FMN) [163], Modified Frequency Median (MFMD) [166], Modified Frequency Mean (MFMN) [166], Frequency Ratio (FR) [167]
Features of GSRs data	MIN, MAX, # of Peaks (NP), Mean Amplitude of Peaks (PA)

After the feature extraction process, we collected: 1 feature for ECG, 38 features for Respiration, 62 features for each GSR signals, and 77 features for each EMG signal. Therefore, the overall number of features is equal to 610.

Part of the content of this section were presented in UBIO workshop and it will be published in " Granato, M., Gadia, D., Maggiorini, D., and Ripamonti, L.A., "Feature Extraction and Selection for

Real-Time Emotion Recognition in Video Games Players", *Proceedings of International Workshop on Ubiquitous implicit BIOmetrics and health signals monitoring for person-centric applications, IEEE*" [74].

7.4.3 Feature Selection

For each experiment, we considered 9 groups of analysis:

RONOVR : RO game session with a standard monitor

PCarsNOVR : PCars game session with a standard monitor

ROVR : RO game session in VR

PCarsVR : PCars game session in VR

NOVR : the merged data of game sessions with a standard monitor

VR : the merged data of game sessions in VR

RO : the merged data of RO in both configurations

PCars : the merged data of PCars in both configurations

Player : all the data collected on each participant

The final goal of each analysis is to select a restricted and informative subgroup of features able to support the definition of an accurate regression hypothesis through a ML algorithm. We used as target variables the values of emotion self-assessment (VA) provided by participants during the experiment.

Moreover, we reduced the number of features in order to alleviate the *curse of dimensionality* [171]. In addition, we removed the redundant or irrelevant variables in order to improve the training performance of the ML algorithm. In order to not falsify or manipulate the data, the feature selection was applied through an automatic procedure.

The first filter aims to remove the features not relevant for the prediction (with ML methods) of the target variables. It is applied to calculate the Pearson linear correlation between each variable and the VA arrays. Thus, we tested the hypothesis of no correlation and we have stored only the features rejecting the hypothesis (*p-value* < 0.05). On the resultant subgroup, we applied an algorithm which identifies only few variables considered as most informative. It is a modified version of Sequential Floating Forward Selection (SFFS) [172] algorithm. It returns a set of features that should be able to minimize the regression error provided by the ML hypothesis. We show the pseudo-code of the method in Algorithm 4. Considering a generic predictor method, the ERR(x) function returns the predictor error index. In our specific case, it is calculated using a 10CV in which a set of RFs [149] were trained. Each RF used 100 trees, as suggested in [173], and with 1/3 of features for each decision split. For each fold, we acquired the RMSE index divided by the number of the elements available in the test set.

```
Y ← {∅}
F ← {ExtractedFeatures with corr. p-value < 0.05 }
oldY ← emptysetofY
while length(Y)  < length(F) do
    ERRListOne ← {∅}
    for each V ∈ (F − Y) do
        ADD ERR({Y ∪ V}) to ERRListOne
    end
    Y ← FEATURES that MIN(ERRListOne)
    if  Y ∈ oldY OR
    ERR(Y) = 0 OR
    ERR(Y) is a local minimum then
        return Y
    end
    ADD new line in oldY with Y features

    ERRListTwo ← {∅}
    for each V = element ∈ Y do
        ADD ERR({Y − V}) to ERRListTwo
    end
    if MIN(ERRListTwo) < MIN(ERRListOne) then
        Y ← FEATURES that MIN(ERRListTwo)
    end
end
```

Algorithm 4: Modified version of SFFS. F is the subset of features selected in the first step of feature selection.

Part of the content of this section were presented in UBIO workshop and it will be published in " Granato, M., Gadia, D., Maggiorini, D., and Ripamonti, L.A., "Feature Extraction and Selection for Real-Time Emotion Recognition in Video Games Players", *Proceedings of International Workshop on Ubiquitous implicit BIOmetrics and health signals monitoring for person-centric applications, IEEE*" [74].

7.4.4 Emotional State Prediction

Considering only the selected features, acquired through the feature selection process presented in Sec. 7.4.3, we tested different supervised learning techniques in order to verify which one performs better on our dataset. As a consequence, we listed a set of potential regression algorithms (further details on the parameters used to train these algorithms can be found in appendix B):

- **Gaussian Process Regression** (GPR) [174]
- **Random Forest** (RF) [148]
- **Gradient Boosting** [175] of trees (GBoT)
- **Support Vector Machines** [176] with *Linear* kernel
- **Support Vector Machines** [176] with *Gaussian* kernel

Each algorithm was tested using a 5CV on each group of analysis. Thus, we used the algorithm which, in average, provided a model reaching a better accuracy on the depended variables prediction (i.e., VA self-assessment values) to infer the players' emotions. If the error levels of the subgroup of models that provide the better accuracy do not present a significant difference, all the models were trained *n* times, and the algorithm with the average better accuracy were selected.

In our specific case, the Gaussian Process Regression (GPR), also known as *Kriging*, provided the better results. Moreover, it is a different method than the ML algorithm used in the second step of feature selection: this should minimize the probability to obtain biased results. Although the GPR computational cost is quite high ($O(n^3)$), the limited number of instances for each experiment does not affect excessively the overall performances of the method. Moreover, GPR has a non-parametric approach, which only assumes that similar data points, defined by a covariance function, are close in the output space. As suggested in [177], GPR is robust to errors in input sources (e.g., loss of an electrode contact), since it depends directly on data, and not on the features' relationship. In our approach, to reduce computational time, we lowered the number of cross-validation folds from 10, used for the feature selection, to 5. Reducing the number of folds, we increased the number of elements in the test set and, consequentially, the algorithm returns a more pessimistic error. However, as we discuss in Sec. 7.5.3, the final results show how this choice does not affect excessively the accuracy of the prediction of *Kriging*. For each experiment, the average time required for a single core to train the *Kriging* algorithm, on the computer configuration presented in Ch. 5, is ≈ 4 seconds, and the 5CV, computed in multi-thread, requires in average ≈ 12 seconds (Fig. 7.10). Moreover, we tested, for each training, a

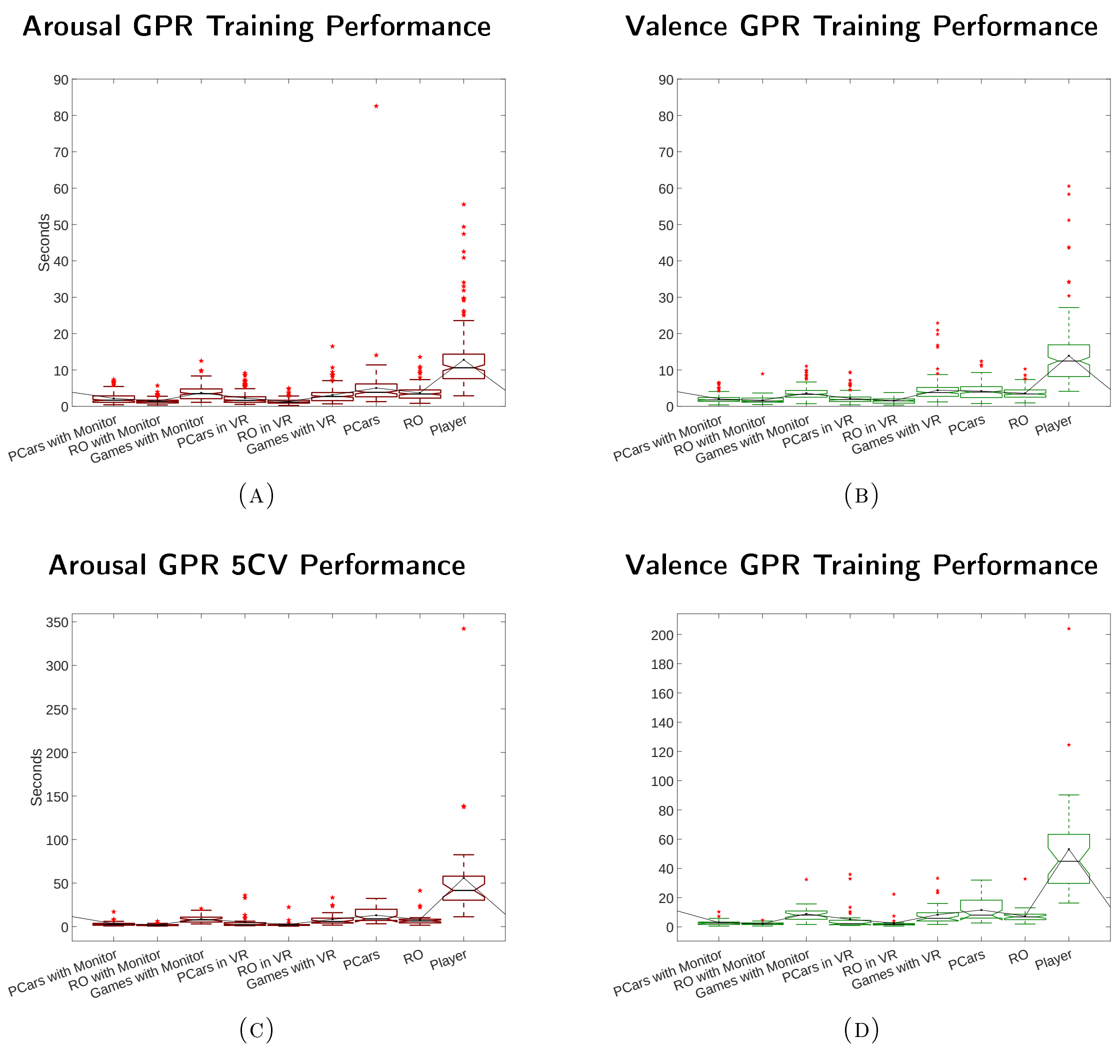


FIGURE 7.10: Box Plots of the GPR algorithm training. In (A) and (B) the algorithm were trained on a single core, while in (C) and (D) each experiment has distributed the 5CV training on 4 cores. The cross inside the boxes underlines the average value, while the outliers are represented with the stars.

set of σ_{gpr} values in a range $[10^{-3}, \sigma(x)]$ and we collected the results which minimize the internal CV using Bayesian optimization as suggested in [178].

7.5 Results

In the following subsections, we present the data analysis outcomes. In Sec. 7.5.1, we focus on the data acquired by the survey, and on the evaluation of the emotion self-assessment provided by the participants. In Sec. 7.5.2, we compare the importance of the different signals, and their related features. Lastly, in Sec. 7.5.3, we present the results of the different prediction models, with a particular attention on the results of the final model.

7.5.1 Overall Experiment Outcome

After each session, we asked at each participant to evaluate the overall experience of the experiment. Almost all participants had not reported a significant discomfort due to the motion sickness ($\mu = 2.85$, $\sigma = 2.59$, in a rank between 1 to 10), and the 82% claimed they were not disturbed by the sensors used in the experiment. One of the crucial hypothesis at the basis of the proposed method is that the affective self-assessment performed by the participants, using the video tagging procedure described in Sec. 7.3.3, is reliable. To validate this, we performed a post-experiment evaluation of the participants' self-assessment accuracy. Thus, we asked the participants to fill out a survey with a set of questions aimed at evaluating how they think they were accurate during the emotion tagging stage. In particular, a subset of the survey questions were designed to evaluate the average emotion assessment in VA vectors, and to evaluate the accuracy during the self-assessment stage. The main hypothesis is that the users are able to express their emotional state using ESAT software, and, as a consequence, that the video annotation procedure can be considered reliable. The participants declared an average precision in emotion tagging equal to 7.48 (in a rank between 1 to 10) for Arousal, and 7.30 for Valence. Moreover, they ranked their focus during the game session (Arousal) equal to 0.56 (in a range between -1 to 1), while the ability of the game to arouse positive or negative emotions (Valence) was equal to 0.28. In particular, the ability to arouse a positive emotion was ranked equal to 0.34, while 0.14 was the rank in case of a negative emotion.

Considering the data acquired by ESAT, the average values of Arousal and Valence were, respectively, 0.41 ($\sigma = 0.44$), and 0.18 ($\sigma = 0.49$). Considering individual participant outcome, the Mean Square Error (MSE) index between the average value acquired by ESAT and the survey answer was, for Arousal, equal to 0.2655, while for Valence was 0.2858. These data seem to validate ESAT, although the survey answers are slightly overestimated.

The results seem to suggest that ESAT is a valid tool to self-assess the emotions of players during video game fruition, and that the proposed hardware setup is an effective solution for the detection of physiological data in the video game research field. In addition, we can consider the answers in line with the information given by emotion tagging, validating thus the reliability of the self-assessment stage. Summarizing, this seems to suggest that the participants were able to report their emotions through the self-assessment phase.

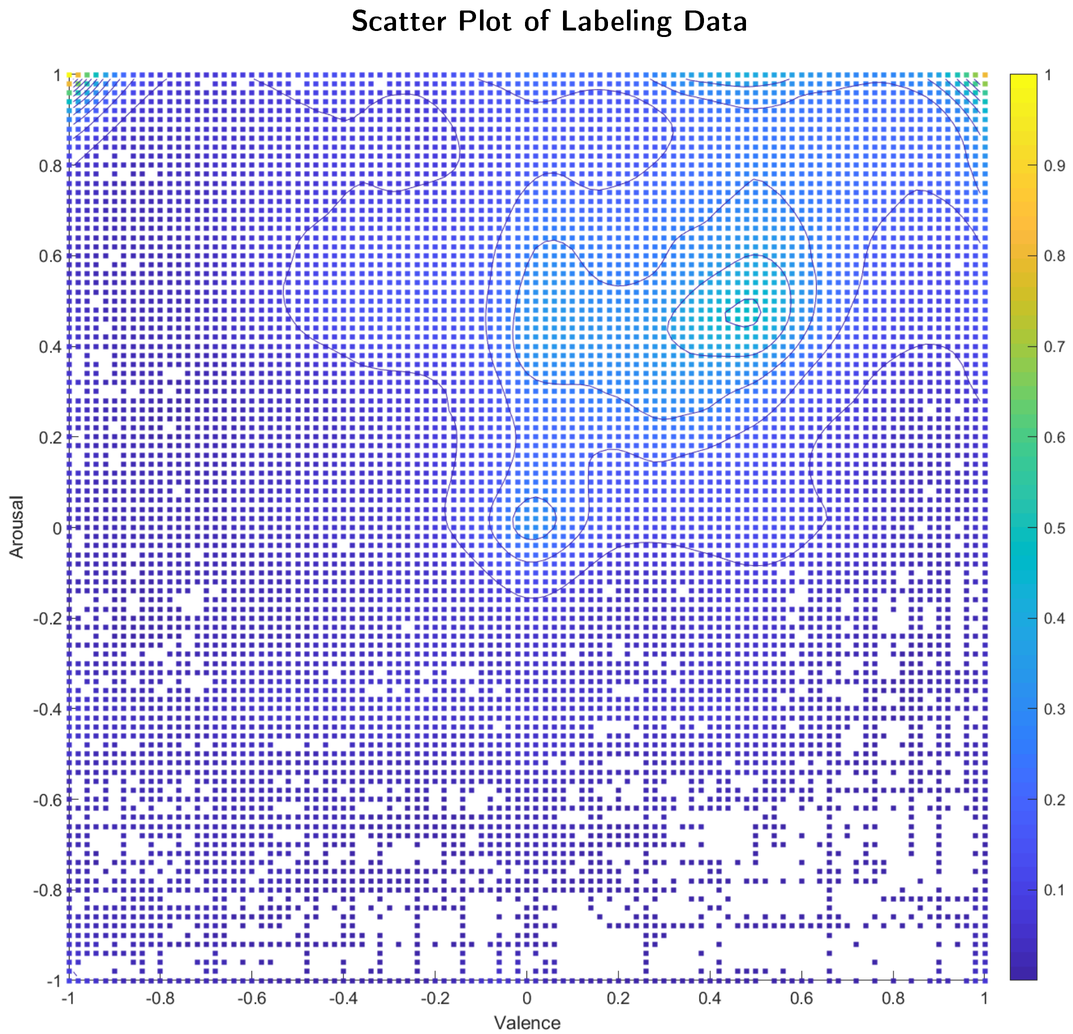


FIGURE 7.11: Distribution of self-assessment evaluations collected on RAGA participants

Considering the self-assessment data distribution (Fig. 7.11), the reported values on the current set of experiments can be considered balanced in the 2-dimensional space, albeit, as in the pilot study (see Ch. 6), the arousal data tend more on the positive values. Moreover, the self-assessment data are more contracted at the ends of valence and top of arousal, and in positive values for both type of emotions.

7.5.2 Features for Players' Emotions Prediction

The feature selection through the Pearson linear correlation has the main goal to store only the informative variables. In our case, this procedure reduced the number of variables, for each analysis, to less than half. In particular, the average number of features considered for arousal and valence

TABLE 7.3: The table provides an overview of the features that have not provided a correlation with the target variables in any experiments

Signal	Arousal	Valence
RESP	MAVS, WL, MFMN, Breath Rate	InterpData, MAVS, WL, MFMN, BP 0-70Hz
SCR	FR, NPeaks	InterpData, FMN, NPeaks, BP 1-1.65Hz
SCL	PeaksMean	InterpData, VAR, PeaksMean, BP 0.7-2Hz
GSR	MAVS, VAR, MFMN, FR, PeaksMean, BP 0-0.65HZ	InterpData, FMN, PeaksMean, BP 1-1.65Hz
EMG5	Mean, PMean, MAVS, WL, FR, BP 25-30HZ, BP 85-90HZ, BP 110-120HZ, BP 125-130Hz , BP135-145HZ, BP 165-175Hz, BP 180-185Hz, BP 225-235Hz, BP 255-260Hz, BP 275-280Hz	InterpData, PMean, MAVS, WL, FR, BP 0-25Hz, BP 30-45Hz, BP 50-55Hz, BP 100-105Hz, BP 150-155Hz, BP 160-165Hz, BP 175-180Hz, BP 205-215Hz, BP 250-255Hz
EMG4	Mean, PMean, WL, MFMN, FR	InterpData, PMean, MAVS, WL, FR, BP 0-20Hz, BP 30-40Hz, BP 60-65Hz, BP 95-105Hz, BP 150-155Hz, BP 195-205Hz
EMG3	Mean, PMean, WL, FR	Mean, PMean, MAVS, WL, MFMN, BP 0-15Hz, BP 70-75Hz, BP 85-90Hz, BP 95-105Hz, BP 130-135Hz, BP 140-145Hz, BP 160-165Hz, BP 170-175Hz, BP 200-205Hz, BP 220-230Hz
EMG2	InterpData, Mean, PMean, WL, FR	Mean, PMean, MAVS, WL, FR, BP 0-15Hz, BP 190-195Hz, BP 230-235Hz
EMG1	Mean, PMean, MAVS, WL, MFMN, FR	InterpData, Mean, PMean, MAVS, WL, MFMN, FR, BP 0-15Hz

analysis is, respectively, 303.6 (49.77%) and 286.2 (46.92%). In the overall experiments, the variables never considered were presented in Table 7.3. This set of features was considered not informative in our research, and, as a consequence, they will no longer be calculated in future analysis.

In literature, there are some feature selection models able to define which variable is more informative than others (e.g., [179]). For example, RF provides an index for each feature, considering which split of the tree will be the most effective to distinguish the classes, and reporting the importance through a standard index, like e.g., Gini index [179]. Unfortunately, Algorithm 4 does not provide a direct measure, using a standard index, of the importance ranking of each feature.

As a consequence, we were not able to identify accurately which feature is the most important. However, considering the number of times in which a feature were selected by our algorithm, we were able to obtain the information on the most involved features in the definition of different kind of players' emotions during racing video games. In particular, the proposed method involved 154 features to analyze the Arousal self-assessment values, and 206 for the Valence, for an average of, respectively, 6.46 and 6.89 variables considered in each group of analysis. Moreover, starting from the features filtered through Pearson correlation, our model selects (in average) 2.50% of the *arousal* and 2.88% of the *valence* variables. In Fig. 7.12 the number of features occurrences is shown, grouped by their origin signal, while an overview of the importance of each extracted features is shown in Fig 7.13. The tonic component of GSR (SCL) can be considered as the most informative signal: its features have been selected 882 times for Arousal and 852 times for Valence analysis. Moreover, its minimum and maximum amplitudes are the features most involved during the feature selection process. These results seem to be in line with racing game design, as they are designed to maintain the player's attention over the time, increasing stress levels as the player approaches to the end of the race. Thus, it is coherent with the behavior of the SCL, which provides information on the emotional status in the medium period (over a second).

Focusing, for example, on Player analysis which considers all the data of each participant during the overall experiment (see point **Player** in the list presented in Sec. 7.4.3), the selected variables used to predict the VA target values are in line with the features occurrences distribution above mentioned (Fig. A.9).

The proposed feature selection method can be also used to validate novel features, since it compares all the features and it extracts only the most interesting. It is also designed to work autonomously, looking for the best set of features that maximizes the data prediction, and, as a consequence, the ground truth of the hypothesis. Lastly, it provides a history of the selected variables, structured in the order in which the features are selected. As a consequence, during the test of a novel feature, the algorithm may provide (with a certain approximation), beside the boolean information of feature importance (i.e., considered or not considered), an index of informative level which contains the variable, according to its position in the history.

Part of the content of this section were presented in UBIO workshop and it will be published in " Granato, M., Gadia, D., Maggiorini, D., and Ripamonti, L.A., "Feature Extraction and Selection for Real-Time Emotion Recognition in Video Games Players", *Proceedings of International Workshop*

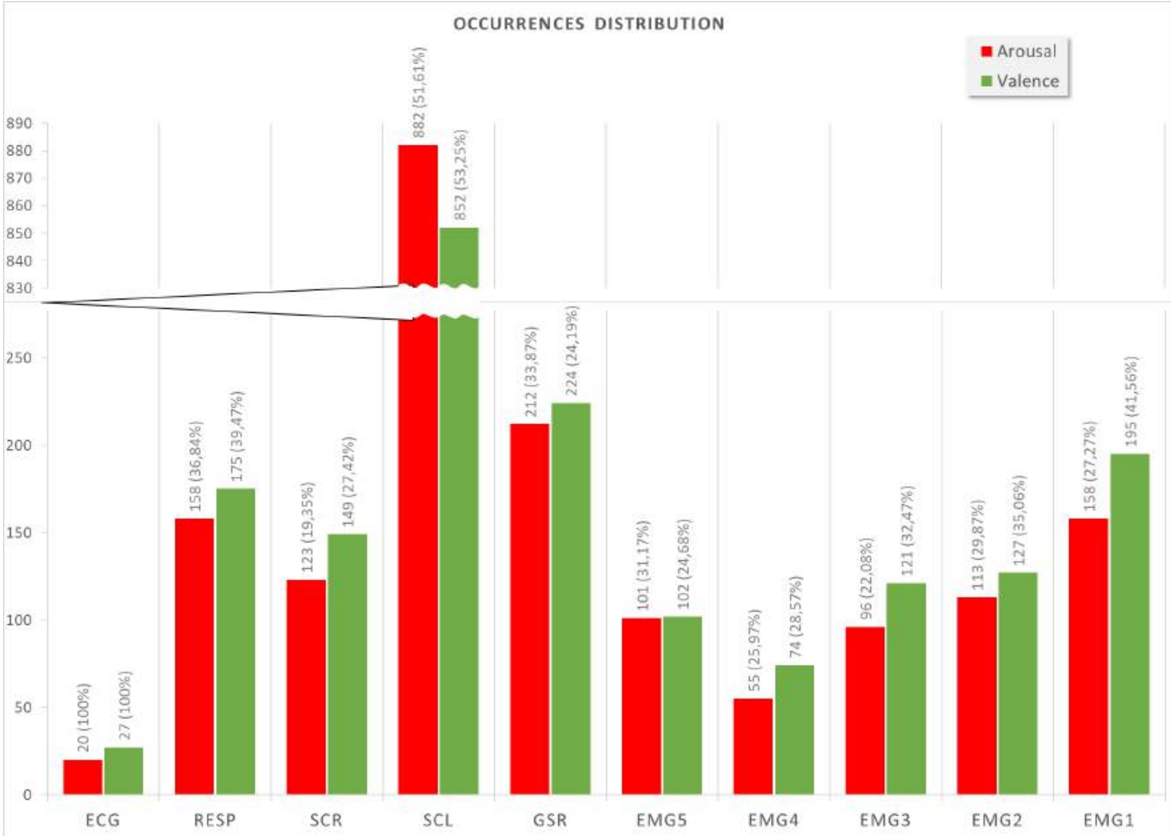


FIGURE 7.12: The figure presents, for each feature, the number of occurrences, grouped by their original signals. On the top of each bar the percentage of features involved in the process is presented. The correspondence of EMGs number are shown in Fig. 6.2.

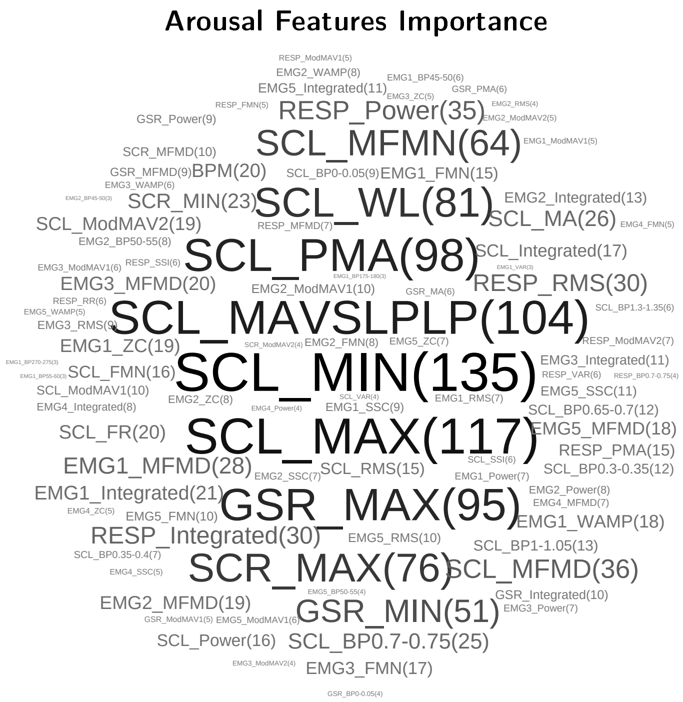
on Ubiquitous implicit BIOmetrics and health signals monitoring for person-centric applications, IEEE" [74].

7.5.3 Emotions Prediction Outcomes

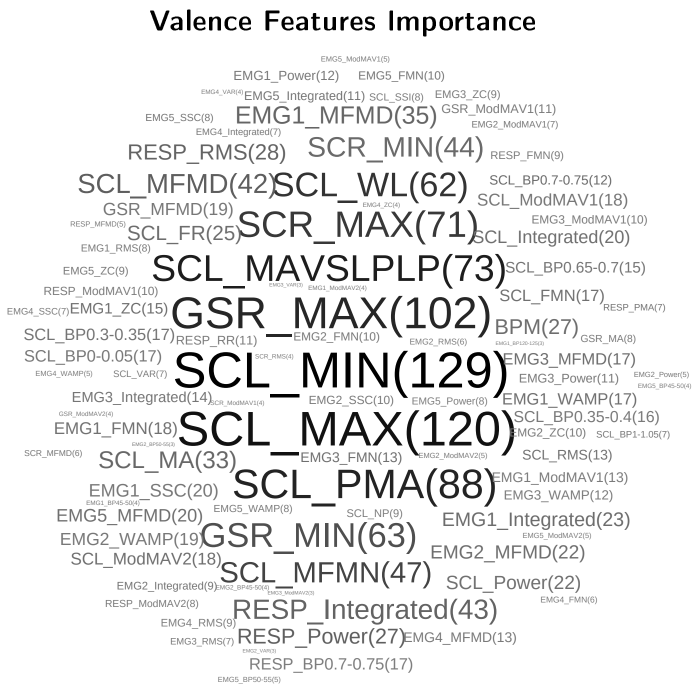
We also evaluated the efficacy of the ML algorithm to predict the emotions, in the Valence and Arousal space, during the video games sessions. We extracted the results of 5CV, for each participant, and we acquired the RMSE[6] between observed and estimated data. Thus, we calculated the Normalized Root Mean Square Error (NRMSE), presented in Eq. 7.24, in order to have scale-free results.

$$NRMSE(y, \hat{y}) = \frac{RMSE(y, \hat{y})}{max(\hat{y}) - min(\hat{y})} \tag{7.24}$$

[6]for the RMSE formula see Appendix C



(A) Word Cloud of selected features for Arousal



(B) Word Cloud of selected features for Valence

FIGURE 7.13: The Word Clouds indicate the number of times each feature has been selected as informative

The NRMSE may vary between 0.0, that indicates a perfect overlap among the estimator ($\hat{y}$) and observed (y) sets (i.e., a perfect prediction), and 1.0 that indicates two divergent sets of data. This range is respected only under two constraints: if $max(\hat{y}) \geq max(y)$ and $min(\hat{y}) \leq min(y)$, which were satisfied in our case study.

In Table 7.4 and in appendix B the accuracy values comparison of each ML methods used for the preliminary analysis are shown. As it can be seen from the results, and as already mentioned in section 7.4.4, the GPR algorithm holds a higher accuracy in the prediction of the results. Thanks to its significant ability to provide a better accuracy than the others models, it was not necessary to repeat the train several times. For example, the higher outcome of *valence* data using SVML prediction model is an outlier. It were calculated on a participant which suffered of a strong motion sickness during the VR session, albeit she completed the experiment (see Appendix B). In particular, the NRMSE value of the overall VR session is 29.695. Instead, GPR handled this particular case better, predicting the *valence* observed data with a good accuracy (NRMSE = 0.078).

Starting from the previous outcomes, we provided a Hyperparameter (i.e., σ_{gpr}) tuning as described in section 7.4.4. This procedure has the main goal to improve the model accuracy, and,

TABLE 7.4: Average NRMSE between observed and estimated for each ML method

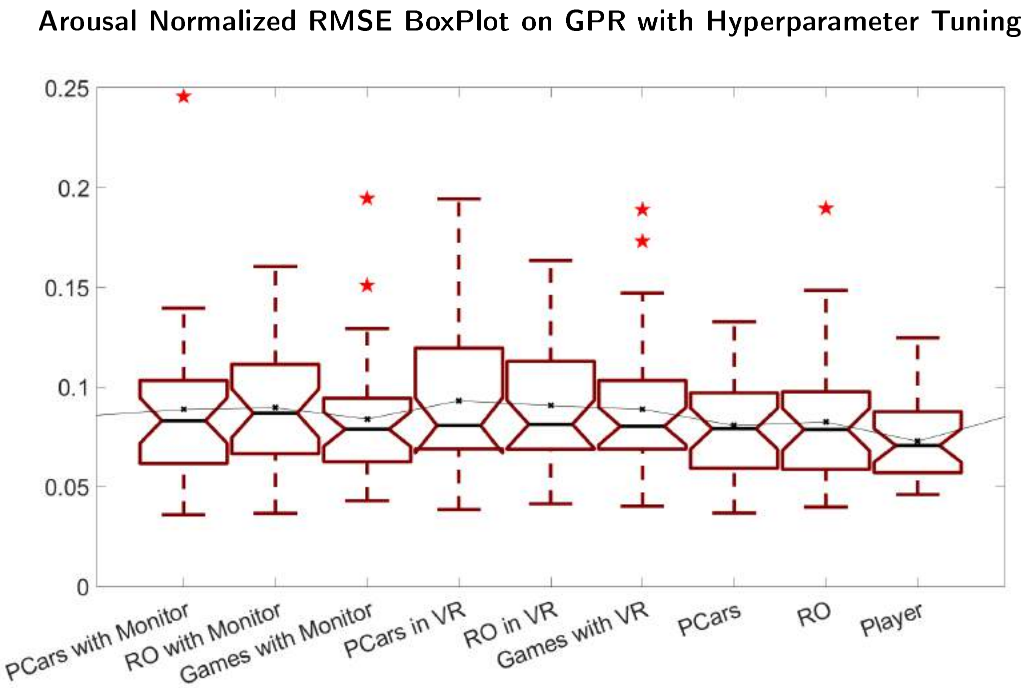
Method	Arousal				Valence			
	Mean	*STD*	*MIN*	*MAX*	*Mean*	*STD*	*MIN*	*MAX*
GPR	0.086	0.031	0.033	0.242	0.089	0.031	0.03	0.233
RF	0.113	0.031	0.039	0.239	0.113	0.0283	0.043	0.233
GBoT	0.143	0.039	0.034	0.301	0.142	0.035	0.045	0.284
SVML	0.220	0.064	0.078	0.698	0.316	1.714	0.114	29.695
SVMG	0.181	0.053	0.056	0.348	0.176	0.050	0.072	0.330

TABLE 7.5: NRMSE of GPR algorithm (with tuning hyperparameters) results for each experiment.

Experiment	Arousal				Valence			
	Mean	*STD*	*MIN*	*MAX*	*Mean*	*STD*	*MIN*	*MAX*
PCars in Classic Enviroment	0.090	0.037	0.036	0.241	0.095	0.036	0.041	0.236
RO in Classic Enviroment	0.089	0.032	0.034	0.152	0.097	0.033	0.037	0.162
Classic Enviroment	0.086	0.030	0.045	0.195	0.090	0.024	0.050	0.158
PCars in VR	0.091	0.035	0.036	0.189	0.096	0.045	0.028	0.195
RO in VR	0.091	0.031	0.041	0.167	0.093	0.029	0.045	0.165
VR	0.087	0.034	0.042	0.187	0.080	0.028	0.051	0.176
PCars	0.079	0.026	0.035	0.135	0.082	0.025	0.044	0.147
RO	0.082	0.032	0.034	0.192	0.089	0.028	0.037	0.161
Player	0.073	0.020	0.046	0.125	0.078	0.021	0.050	0.121
AVERAGE	0.085	0.031	0.039	0.176	0.089	0.030	0.043	0.169
MIN	0.073	0.020	0.039	0.125	0.078	0.021	0.028	0.121
MAX	0.091	0.037	0.046	0.241	0.097	0.045	0.051	0.236

as a consequence, minimize the prediction errors [180]. In Fig. 7.14 and in Tab. 7.5, respectively, the box plots that indicate the NRMSE dispersion of Valence and Arousal, and the experiments numerical results are shown. The average regression error is quite low, which indicates the ability to design a hypothesis that is able to predict the values of Arousal and Valence during video game session with a precision of 1 second. In Fig. 7.14, we present the box plots of the acquired NRMSE data for each kind of experiment. Plots related to each single prediction are available on the RAGA homepage[7] presented in Sec. 7.3.3, while box plots of the different indexes can be found in appendixes B and C.

[7] https://github.com/grano00/GameVRRacingPhysioDB



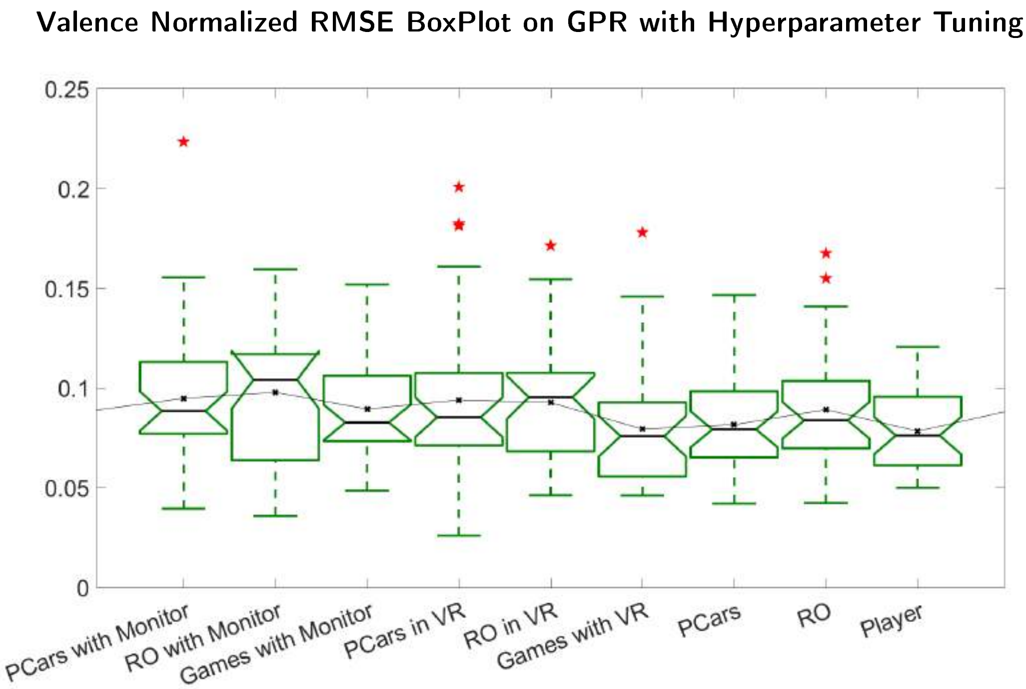


FIGURE 7.14: The box plots present the NRMSE distribution between the estimated data collected using the GPR (trained with different hyperparameters) and the observed self-assessment data across the experiments. The cross inside the boxes underlines the average value, while the outliers are represented with the stars.

7.6 Discussion

In this chapter, we presented RAGA, an affective dataset based on the acquisition of physiological signals from video game players. We provided an overview of the considered physiological data and of the improvements in hardware setup and procedure used to acquire it. We collected the physiological signals and self-assessment information from a set of participants playing racing games. The players played in two different environments, using a monitor, and a VR headset. Furthermore, we provided an analysis of the relevance of each signal and of their contribution to predict the players' emotional state. Lastly, we described the ML algorithm used to design a hypothesis able to predict the emotions of participants at second precision. The results seem to confirm the validity of the experimental framework, since the ML model obtain low errors in each experiment. Thus, the considered algorithm is able to predict, with a certain precision, the self-assessment emotion into the VA 2-dimensional space. Moreover, to the best of our knowledge, there are not freely available databases of affective annotation using video games as stimuli.

Lastly, we can propose a potential interpretation between some of the acquired results, and the *Spatial Presence* and theory of *Flow* concepts, introduced in Sec. 1. In the final survey, the 88% of participants reported more intense emotions during the sessions with VR games; however, we had not found a significant difference in the Valence and Arousal values between the sessions with and without VR. However, some researches hypothesized higher level of *Spatial Presence* condition when using immersive devices. Our results seem to suggest that the evaluation of the *Spatial Presence* condition can not be defined using the 2-dimensional Valence/Arousal space used in our experiment. As a consequence, the future analysis may investigate the use of a more complex system for the identification of emotions (like e.g., PAD). Moreover, analyzing the Valence and Arousal values for each game, we can notice how these values are, in average, quite positive. In fact, PCars collected $\mu = 0.38$ for Arousal and $\mu = 0.15$ for Valence, while RO obtained $\mu = 0.44$ for Arousal and $\mu = 0.21$ for Valence. Considering the average absence of emotions tending to averseness, and also the high level of positive feedback provided by the players through the post-experiment survey, we can hypothesize that the players were relevantly absorbed (thus, in the *Flow* state) during the game sessions.

Chapter 8

Conclusion

THE purpose of this dissertation was to provide a novel contribute to the Affective Computing and video game research field. In particular, our aim was to explore the emotions of the players during video game fruition. We used techniques inspired at different studies in Affective Computing field. Since this thesis has as a transversal approach, which includes different disciplines, we provided in Ch. 2 a background on the relation between video games and emotions, and how to measure the emotional state of the players. In particular, we analyzed how the video games can arouse the emotions in players, and we described some examples of methods useful to study and design games able to elicit emotions in players. Then, we described the relationship between the human physiology and the emotions, and how to assess them. In addition, in Ch. 3 we analyzed the related work, considering the available affective datasets, the researchers that inferred the emotions, and the studies that use physiological information as game input.

In Ch. 4, we described the methodology used to design a novel framework able to infer, in real-time, the players' emotions. Thus, we presented a high-level framework architecture, focusing on the emotion recognition node of the Affective Loop. We also justified the different types of physiological data acquired on the players, and the methods used to assess the emotions.

In Ch. 5, we presented an open source architecture able to reveal the humans' physiological data. It is based on Arduino Due, and it is designed to provide a robust data synchronization. Two software were developed to support the data acquisition. DAPIS is a software able to store the data from the Arduino. It also visualizes the physiological signals in order to detect eventual noises. ESAT is a software used to acquire the emotions self-assessment values by the players. It is designed to acquire the VA vectors values in a dimensional space, and it provides the annotations continuous on the overall game sessions.

In Ch. 6, we described a pilot study conducted on 10 participants. We invited the players to play 4 platform games. During the game sessions, we acquired a set of physiological data. Hence, we asked the players to assess their emotions using ESAT. Lastly, we performed a data analysis, and we developed a regression model, based on RF, in order to predict observed emotional data. The analyzes were performed in two ways: a general model, and an individualized model. The pilot results are not particularly significant, however, they shown some design criticism that we fixed in the main experiment.

Finally, in Ch. 7, we performed an extensive experiment on 33 participants. We corrected the design and architecture issues revealed during the pilot study, and we improved the overall experimental setup. The participants played at two racing games on a standard monitor and with a VR headset. Thus, we collected a dataset, available to the scientific community, which provides the physiological information of ECG, GSR, EMGs on 5 facial muscles, and respiration. These data are well synchronized with emotions labeling provided by the participants after each game sessions. Thus, we performed an analysis which shown significant and interesting results.

Summarizing, the dissertation contributed to introduce a modality to study the players' emotions, widely used in *Affective Computing*, in the video game research field. This dissertation can also contribute to improve the research in GUR [17], since one of the methods applied in this research field is the biometric measure of the physiological activity of the players. It could be used to understand which part of a game can generate a specific physiological state, and, therefore, a corresponding emotion. For example, during the beta test of a game, the developers can identify the game areas which unwittingly induce unwanted emotions, and, consequentially, redesign them. In addition, the developers can handle and design an algorithm for a real-time adaptation of some game mechanics (e.g., the mechanics that affect the difficulty), in order to avoid specific unwanted emotions (e.g., boredom). These techniques may help the players' engagement, and, therefore, the possibilities to transmit the desired message.

In particular, the dissertation contributes are:

- To provide a framework and a set of tools able to investigate the players' emotions during video game fruition. The framework is coherent with the Affective Loop approach, since it is able to infer the players' emotions in real-time
- The creation of an affective dataset named RAGA. It is freely available to the scientific community. It contains synchronized physiological and emotional data
- To propose a novel algorithm to select the most important features in a large set of variables
- To investigate the different physiological signals and features, considering the more effective in order to infer the players' emotions

- To propose a method for the real-time players' emotions prediction, and to show a comparison of different ML algorithm accuracy

8.1 Future Work

In a future extension of this work, we will design a racing game (as case study) based on the prediction features introduced in the current research: this game will adapt its difficulty on the basis of the players' emotions, trying to keep the players' entertainment at a qualitative standard required by the video game industry.

Moreover, future works will address two challenges: to design an integrated set of gaming devices able to reveal the physiological data of the players, and to identify game features able to provide the evaluation of players mental state, avoiding a specific self-assessment for the players. The former can be achieved by the integration of the sensors in the game devices (see Fig. 8.1). Almost all the electrodes used in the experiments are located near the hands and face of the players; therefore, it is possible to design their integration into, respectively, a gamepad and a VR headset.

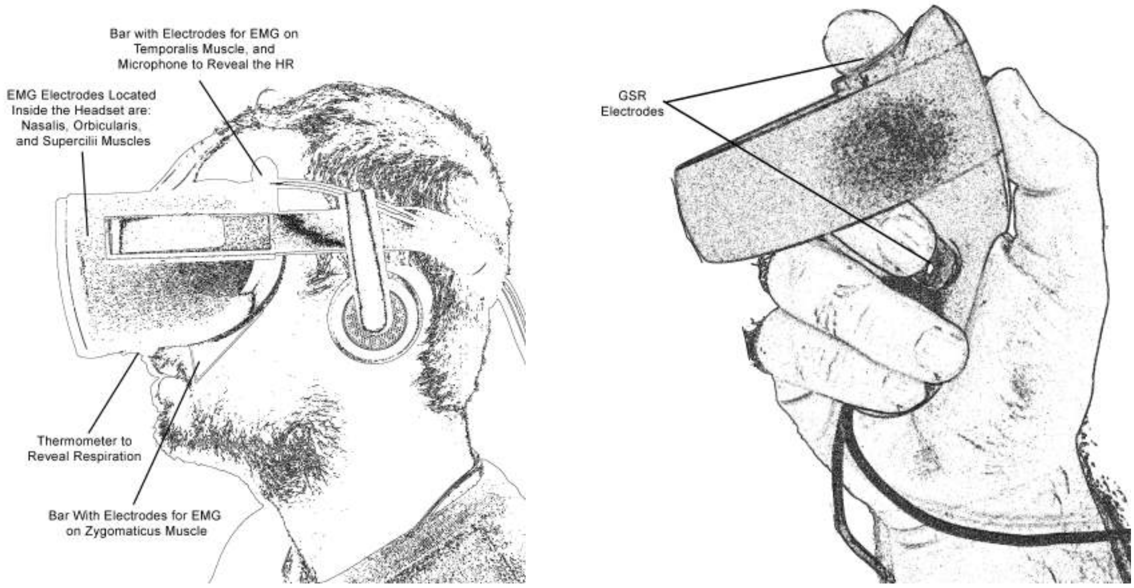


FIGURE 8.1: The figure shows an architecture to embed the sensors used during the experiment in a VR environment. In this example, the architecture is based on Oculus Rift Suite. In particular, in the left image, the headset embeds almost all sensors. The electrodes for acquiring the GRS data are placed in the gamepad. This environment may also support other sensors, like, e.g., a thermometer placed in the bottom area of the gamepad, under the pinkie.

Considering the second challenge, currently, the prediction model is linked to the individual player, as it has to be trained on data collected by each player. As a consequence, our algorithm needs to acquire, for each subject, the self-assessment data in order to provide the target variables to the ML algorithm. However, during the game design process, different options aimed at the acquisition of the emotion labeling can be considered, for example borrowing gamification elements in the final game [181]. A possible approach could be the design of a mini-game that asks to evaluate the emotions on the game highlights, providing, after the identification, an in-game reward. Another solution can be represented by the adoption of a different ML algorithm, with a general hypothesis aimed at predicting the emotion on a wide range of players avoiding the limitation to request the self-assessment information by each player. It could be supported by external annotators in order to decrease the noise due to the emotion tagging variability. Albeit it is commonly used in affective computing research, the time required to provide all the evaluations may not comply with industry production standards. Furthermore, the VR headset may be an obstacle in the consideration of external annotators. During the experiment, we curbed the problem by asking the participants to evoke their emotional state immediately after the game session. However, for an external annotator, it could be a challenge to identify the emotions of players with part of the face covered. As a consequence, a thorough investigation of the methodologies to acquire the players' emotions have to be performed. This generalization can contribute to the investigation on common traits of emotional response to certain game events, in order to build a common model.

The developing of these steps will complete the cycle of the Affective Loop, providing also more data, and, consequently, improving the accuracy in the players' emotions prediction.

Other minor research goals will be:

- The design of a novel framework addressed to study different and specific video game aspects, such as the narrative or the social aspect
- To consider different types of measures (e.g., the log of the gamepad buttons) in order to apply a mixed-method approach and to compare the effectiveness with the dissertation outcomes
- To compare the actual experiments results, and the results achieved without implementing the algorithm of Feature Selection (see Sec. 6.2.3 and 7.4.3). It will provide a measure of the different regression errors using both approaches. Unfortunately, this comparison requires the availability of dedicated hardware with high computational specifics.

- To collect a greater and more dedicated set of participants. It will permit to investigate differences of emotional responses according to different personal parameters, e.g., age or gender. Moreover, increasing the number of users familiar with VR technologies can provide a better comparison of the device emotional impact over the headset usage time.

Appendix A

Plots of Features Selection in RAGA

Starting from the overall representation of the most important features, grouped from the original signal and presented in Sec. 7.5.2, we provide details of each selected features and the number of occurrences in the different experiments (see Sec. 7.4.3). Albeit we have separated the selected features in the different considered experiment, the bar plots represent the cumulative information of each participant. The bar plots have been divided in colors, in order to visually separate the original signals. Moreover, the signals of the same type have the same color, but with different tonalities (e.g., all the EMGs are blue).

The x-axis of the plots has a range between [0,24], and the vertical dotted lines, used as a reference, are located every 3 integers.

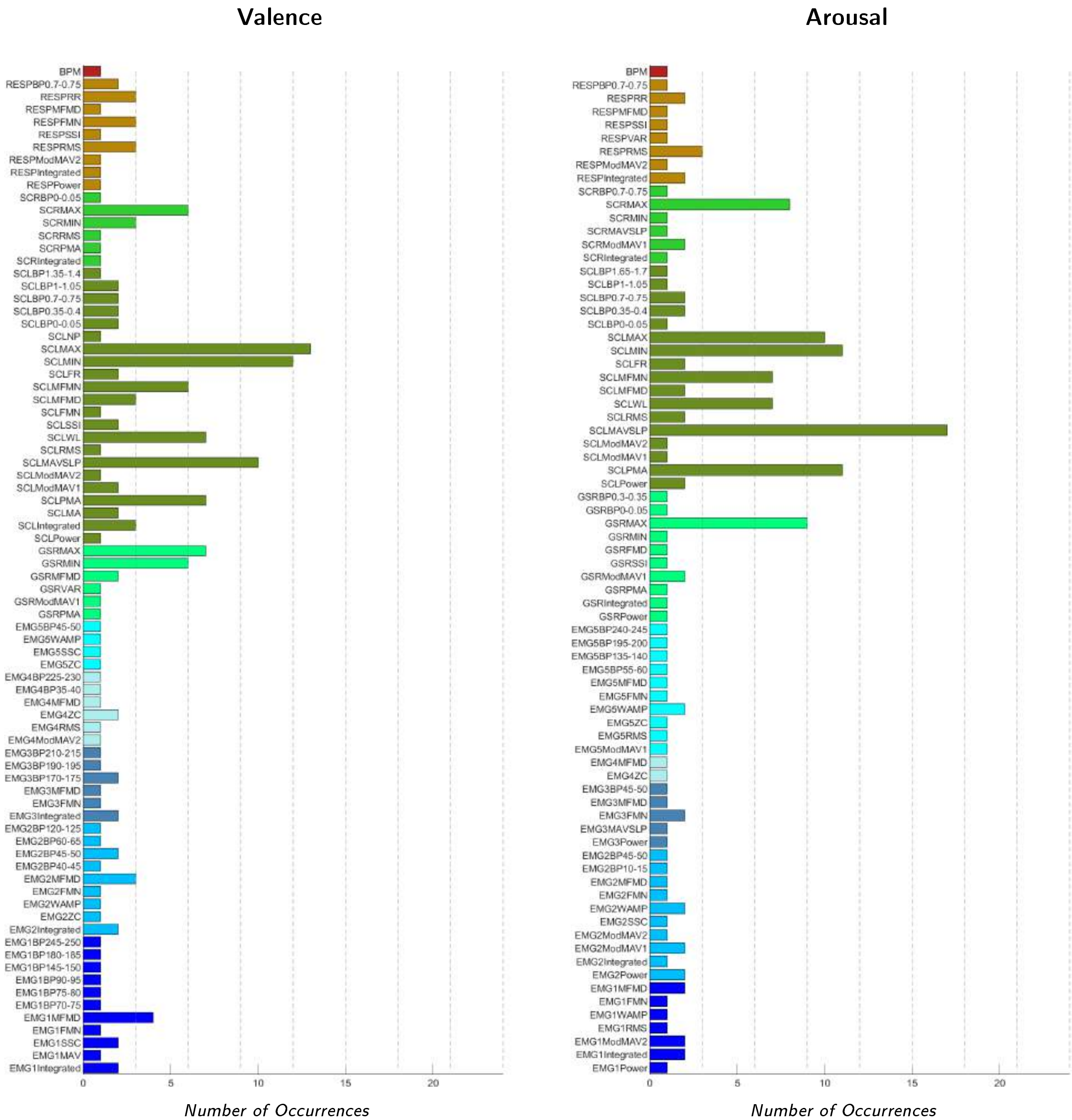


FIGURE A.1: VA Features Occurrences on RONOVR Analysis

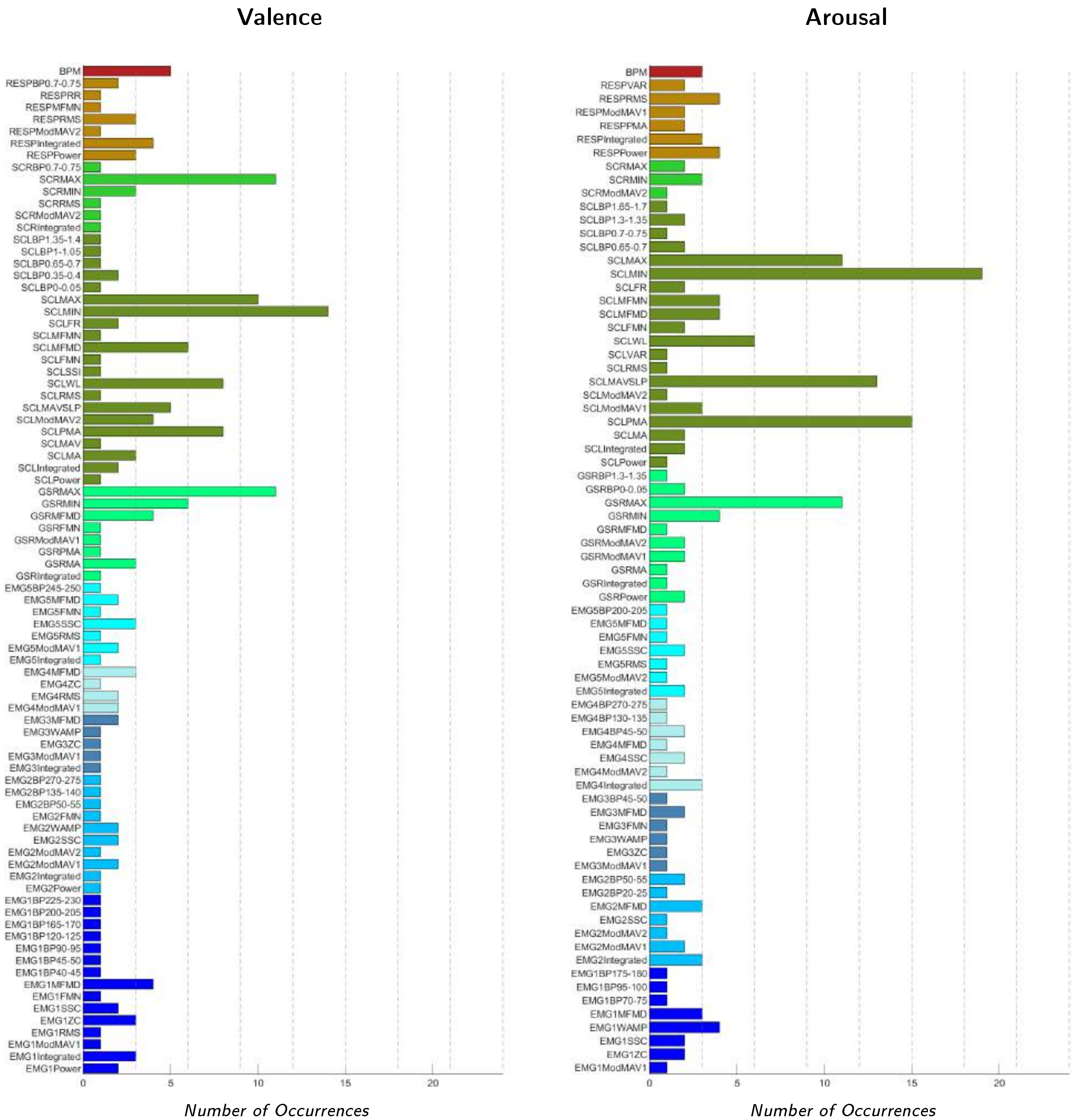


FIGURE A.2: VA Features Occurrences on PCarsNOVR Analysis

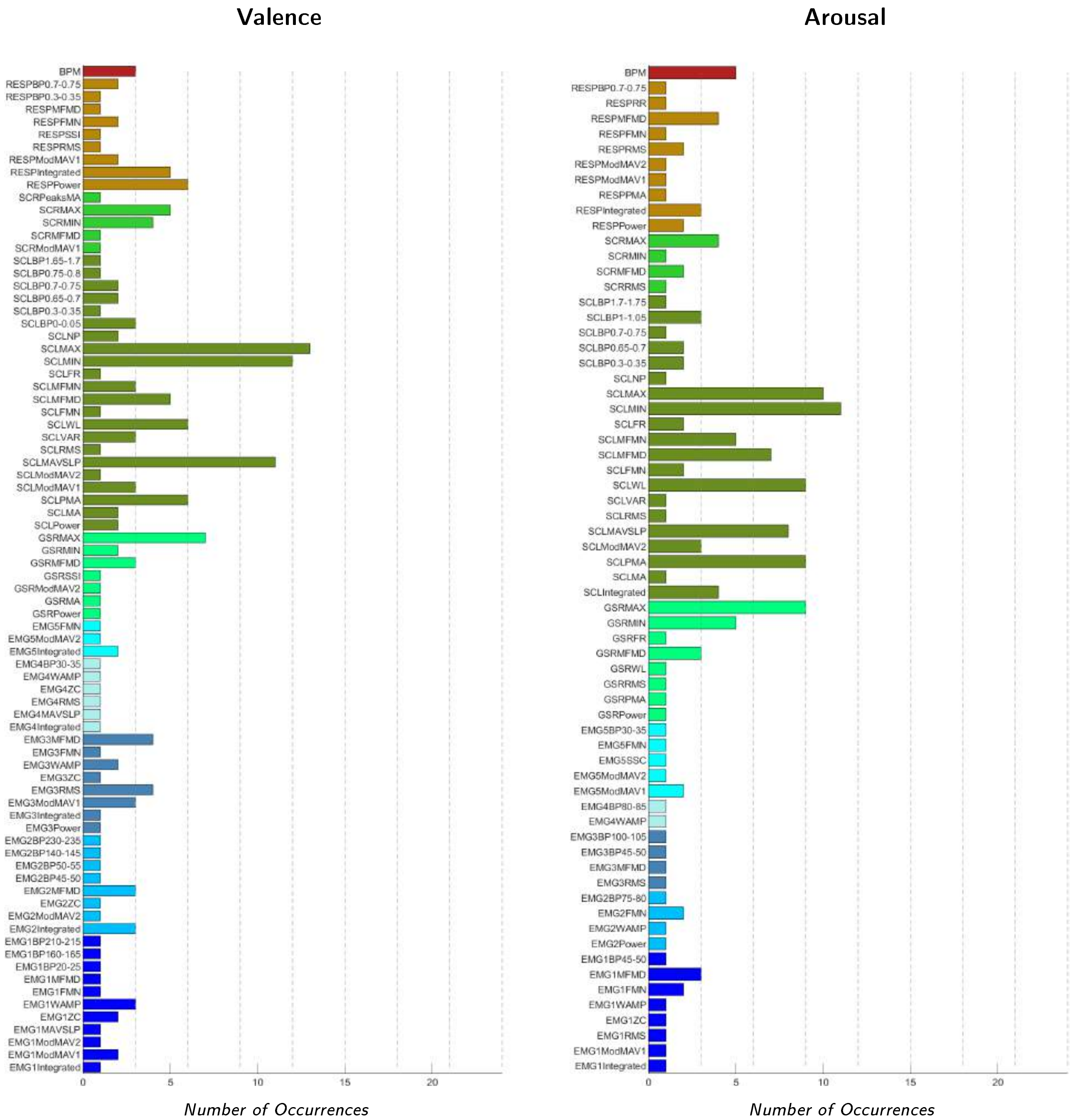


FIGURE A.3: VA Features Occurrences on ROVR Analysis

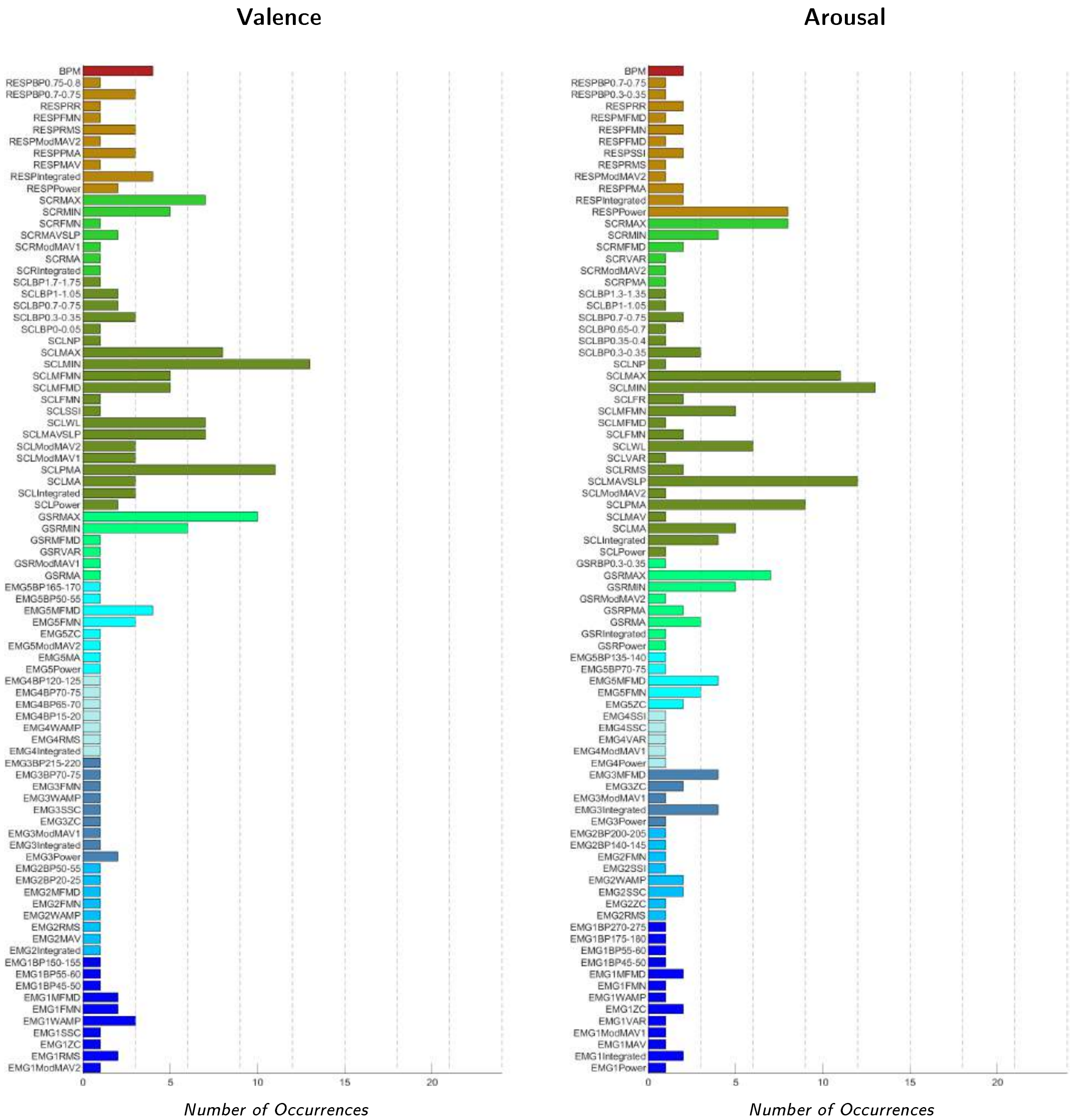


FIGURE A.4: VA Features Occurrences on PCarsVR Analysis

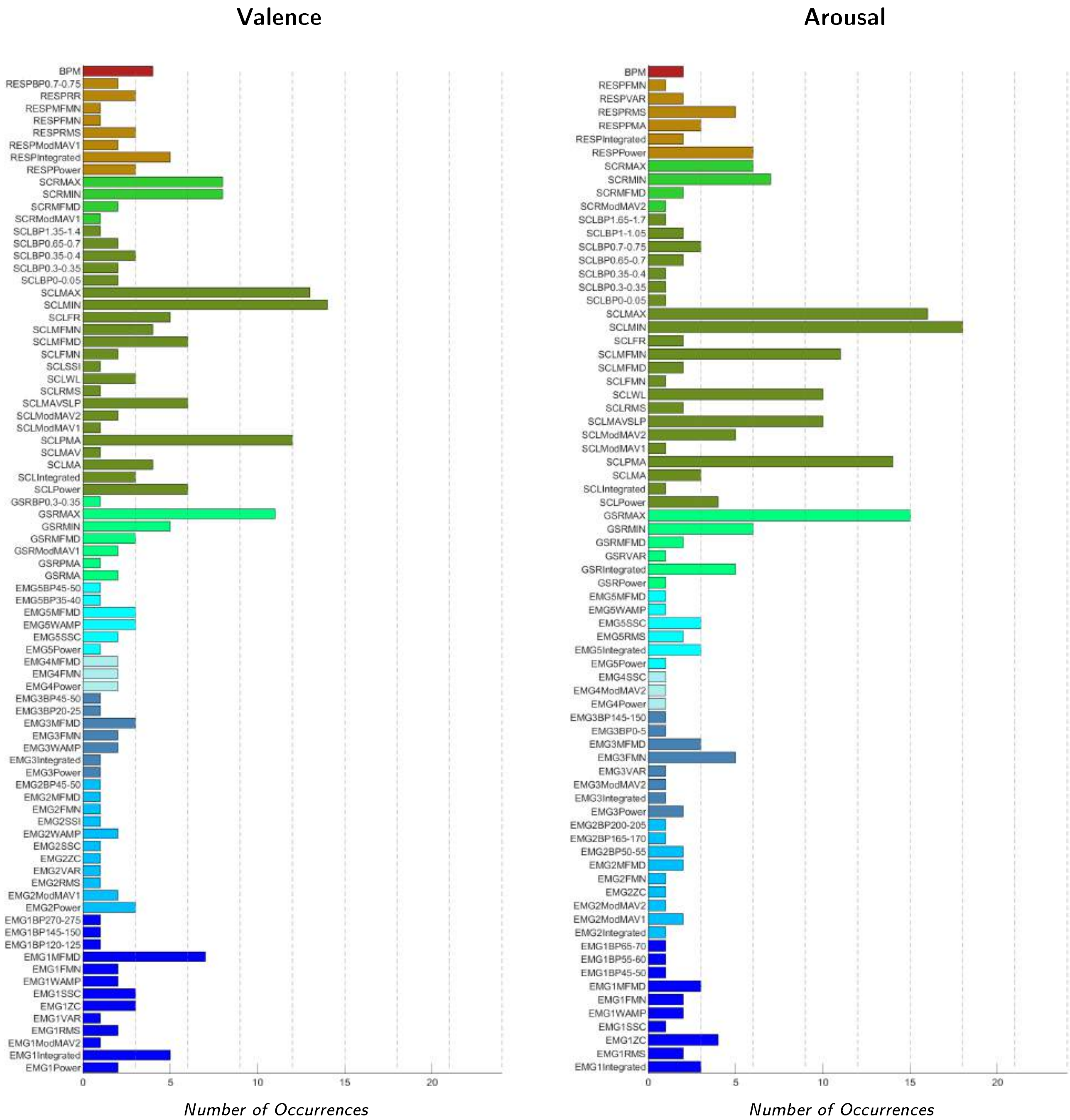


FIGURE A.5: VA Features Occurrences on NOVR Analysis

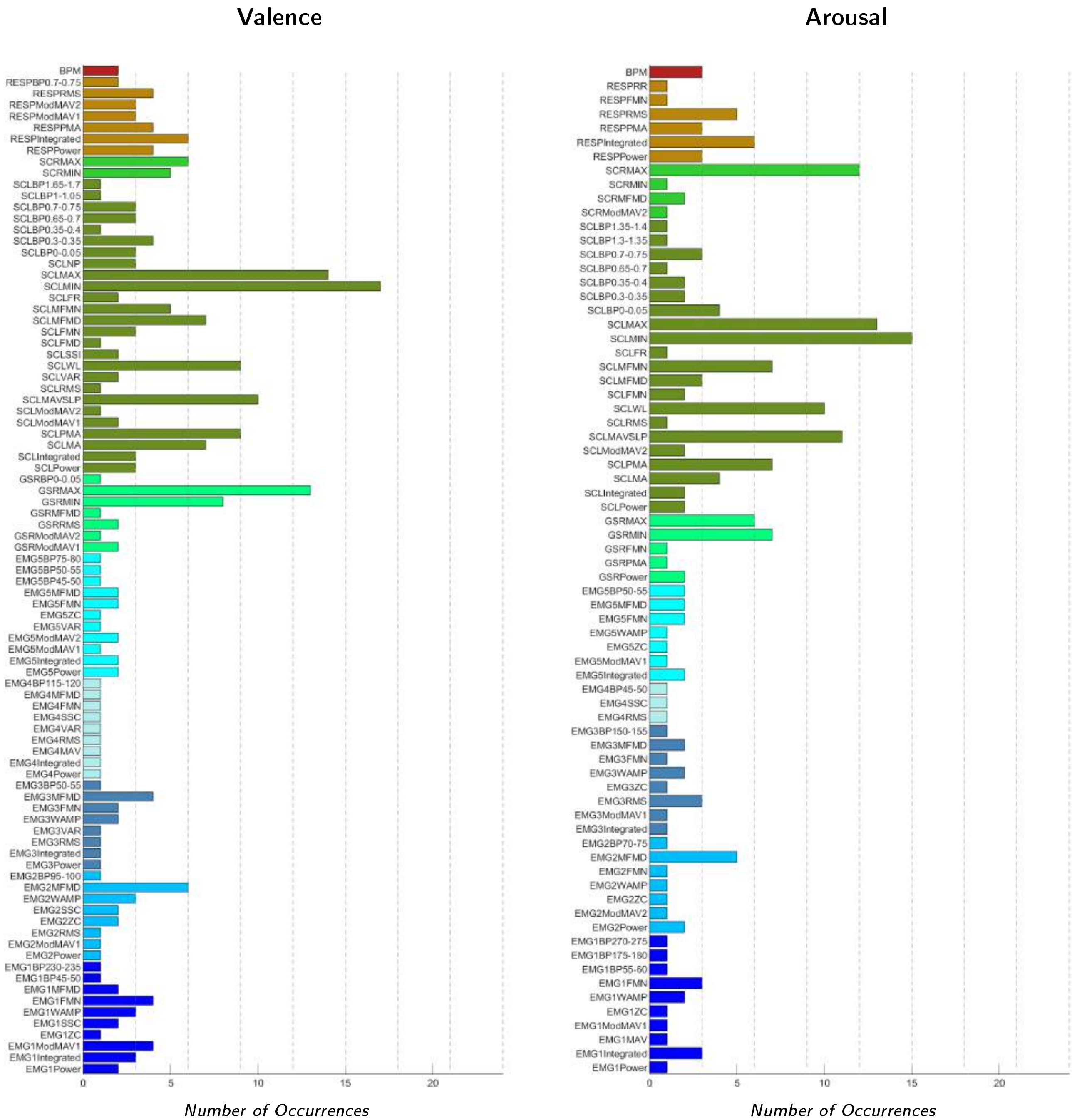


FIGURE A.6: VA Features Occurrences on VR Analysis

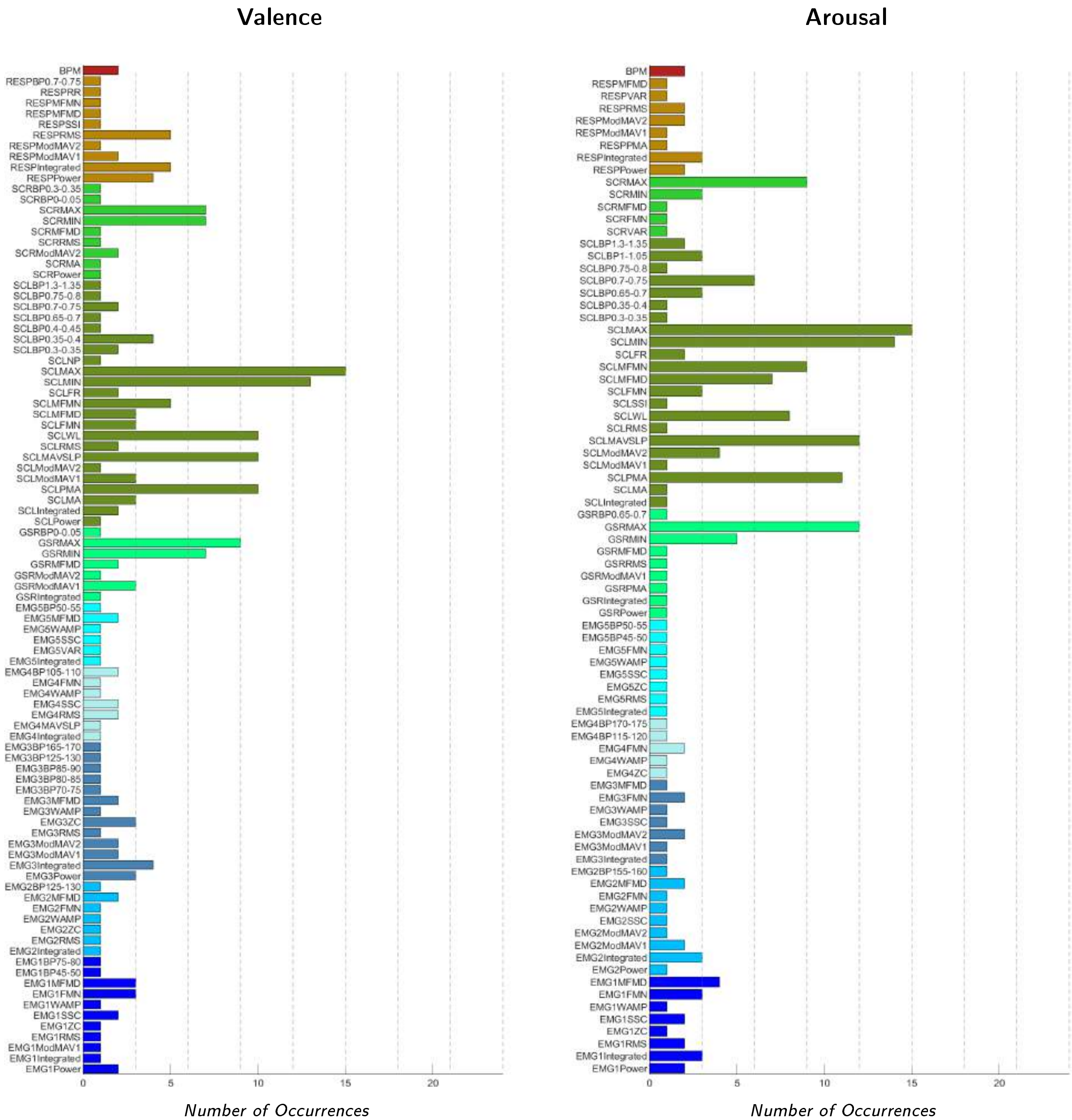


FIGURE A.7: VA Features Occurrences on RO Analysis

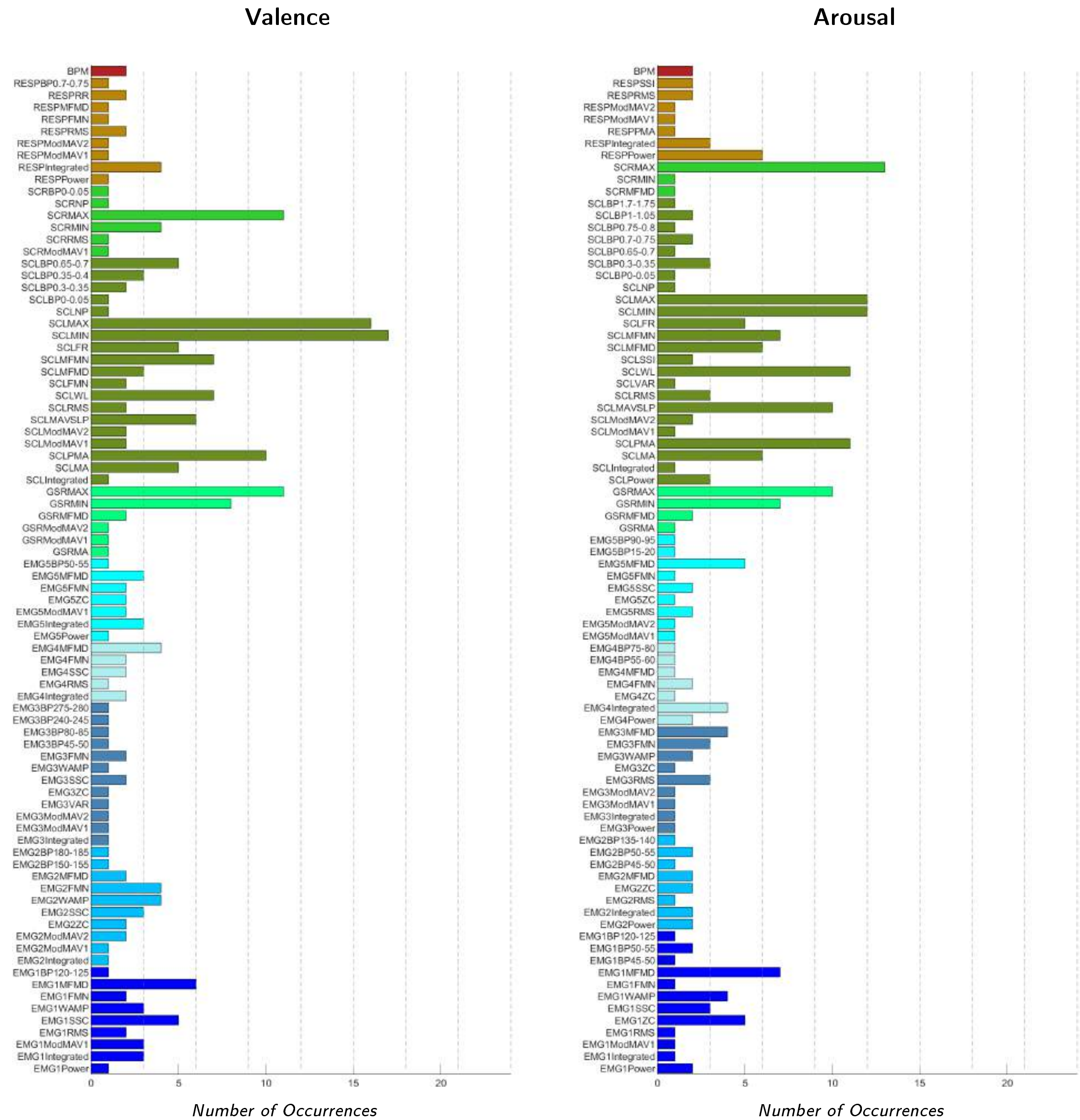


FIGURE A.8: VA Features Occurrences on PCars Analysis

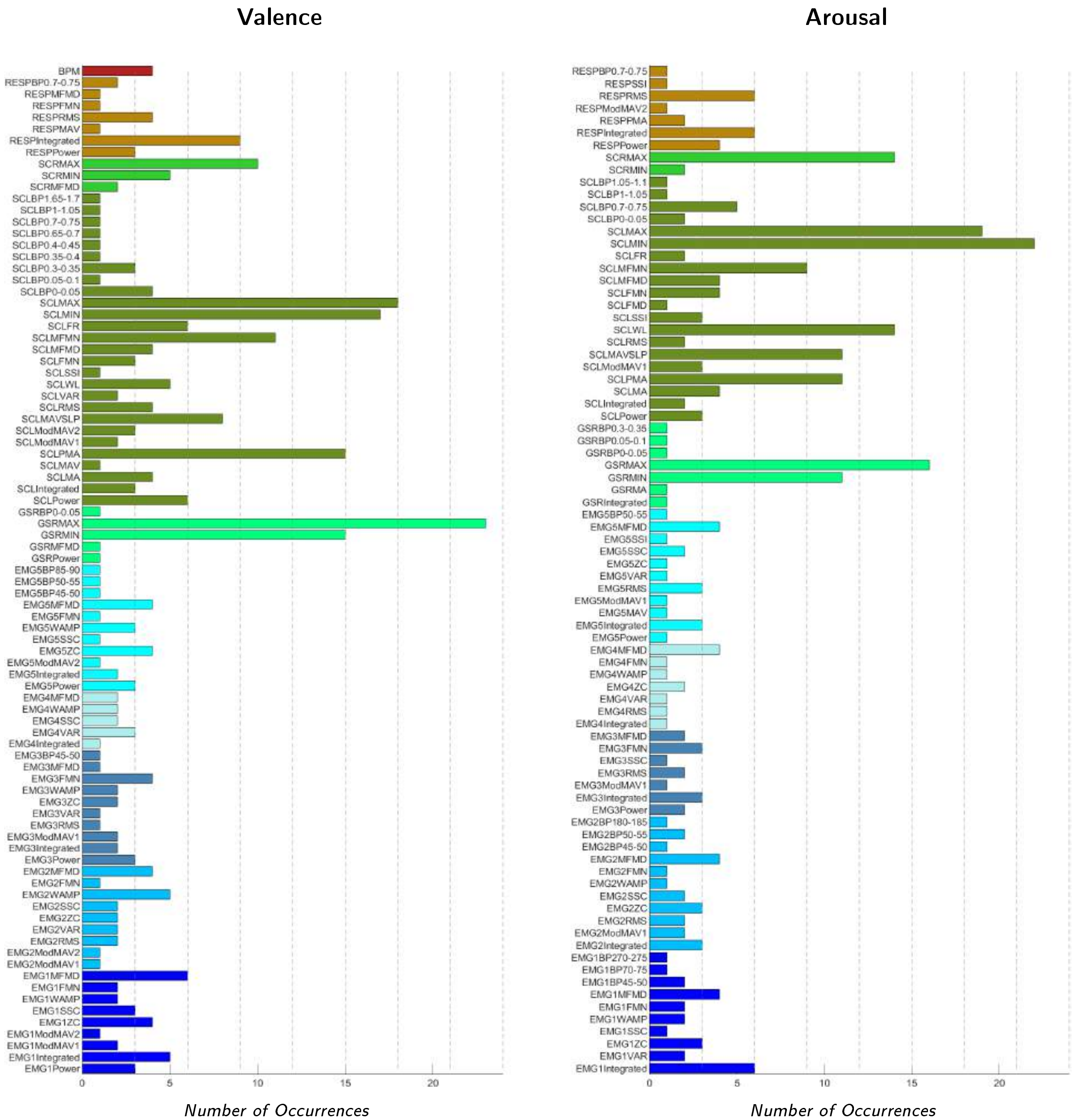


FIGURE A.9: VA Features Occurrences on Player Analysis

Appendix B

Plots of Different Machine Learning Methods

In this appendix, we compare the results, grouped for each experiment, of the different ML methods. The box plots present the NRMSE distribution between the estimated data collected using different models and the observed self-assessment data across the experiments. The cross inside the boxes underlines the average value. The considered algorithms are:

- **Gaussian Process Regression** (GPR) [174], which uses a rational quadratic kernel function. The starting sigma value is $\sigma_{gpr} = {}^{\sigma(x)}/\sqrt{2}$
- **Random Forest** (RF) [148] using 1/3 of features subset for each decision split and with 128 trees
- **Gradient Boosting** [175] of trees (GBoT), with the number of ensemble learning cycles equal to 100
- **Support Vector Machines** [176] with *Linear* (SVML), and with *Gaussian* (SVMG) kernels. In both SVMs methods, $\epsilon = iqr(y)/iqr(N(\mu, \sigma))$, where iqr is the *interquartile range*, y is the target variable (i.e., VA), and N is the normal distribution, with $\mu = 0$ and $\sigma = 1$. It is a robust estimator of the standard deviation [182]

As it can be noticed, the SVM with Linear Kernel has been not able to predict correctly the valence data provided by the user 20. She was the only participant to have suffered of a strong motion sickness in both games (and in particular in RO) during the VR sessions. Moreover, in this case, the NRMSE condition, presented in Sec. 7.5.2, is not respected, since the estimated data have a range greater than the observed data. However, it seems that this particular participant's condition has not influenced the performances of the other algorithms.

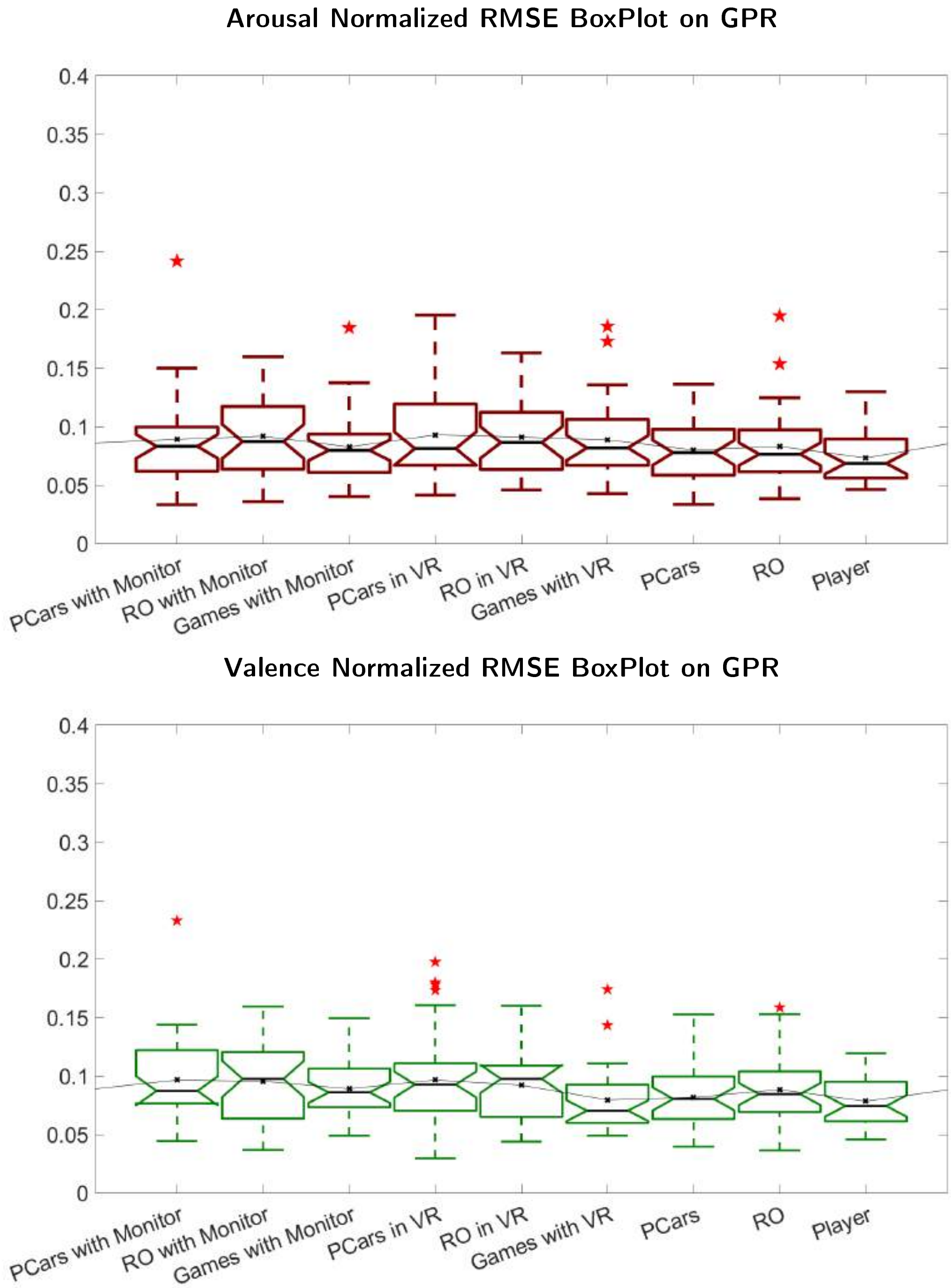
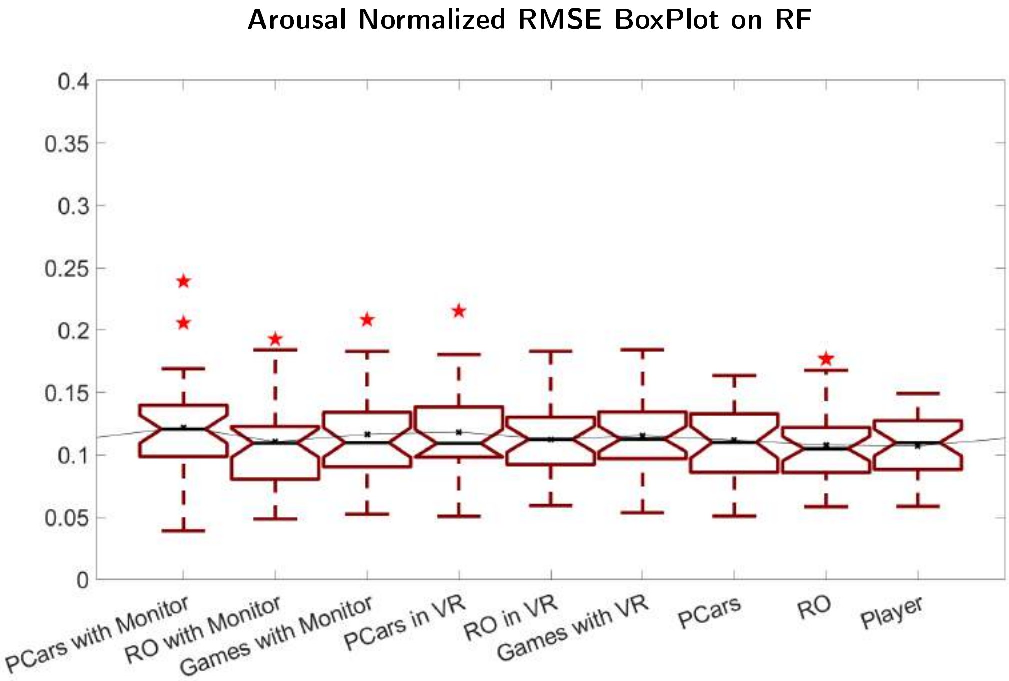


FIGURE B.1: Accuracy of the GPR algorithm on the observed data. We have used the NRMSE index as error metric. The cross inside the boxes underlines the average value, while the outliers are represented with the stars.



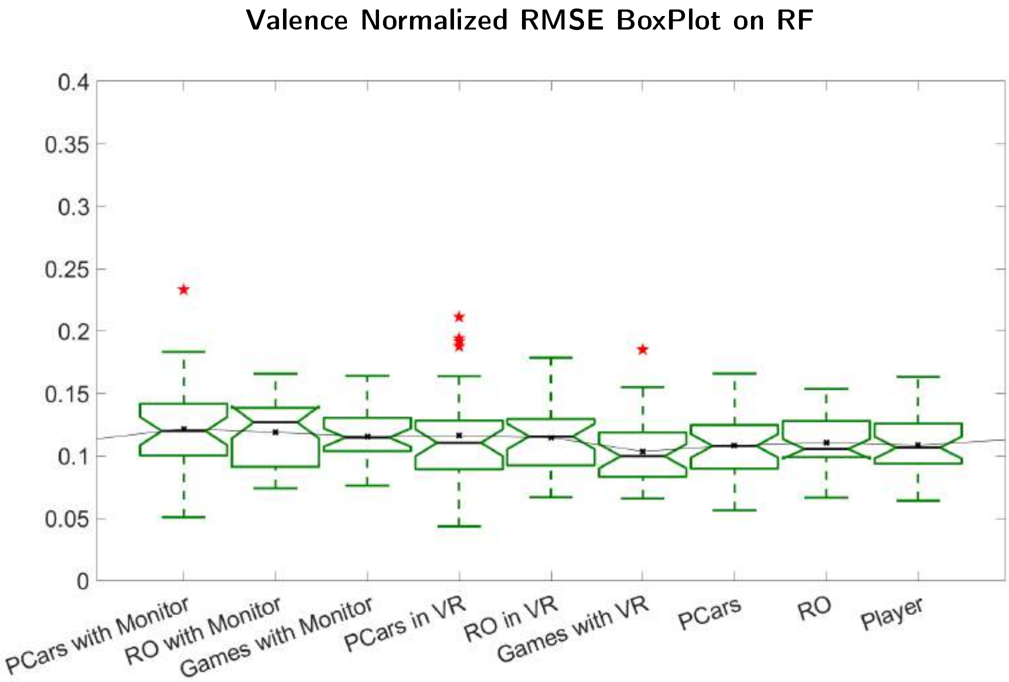
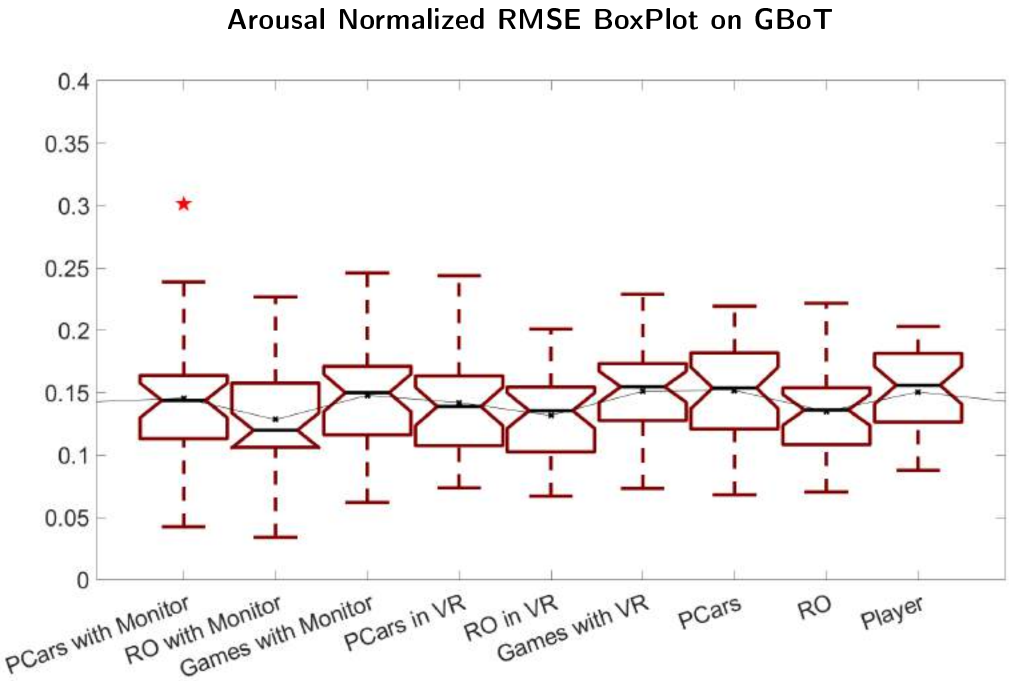


FIGURE B.2: Accuracy of the RF algorithm on the observed data. We have used the NRMSE index as error metric. The cross inside the boxes underlines the average value, while the outliers are represented with the stars.



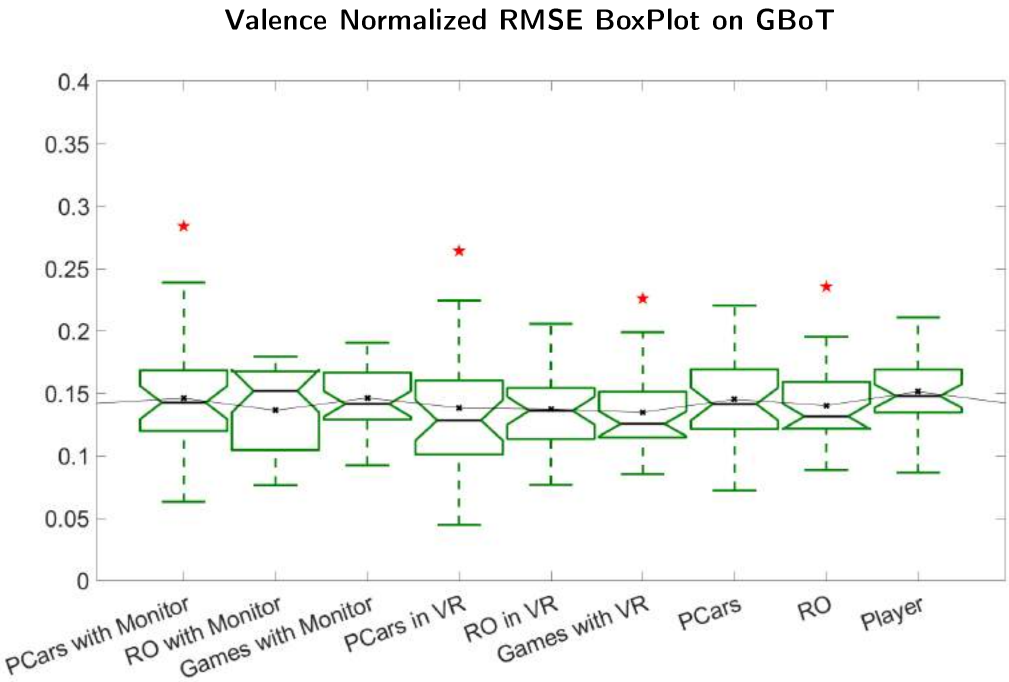


FIGURE B.3: Accuracy of the GBoT algorithm on the observed data. We have used the NRMSE index as error metric. The cross inside the boxes underlines the average value, while the outliers are represented with the stars.

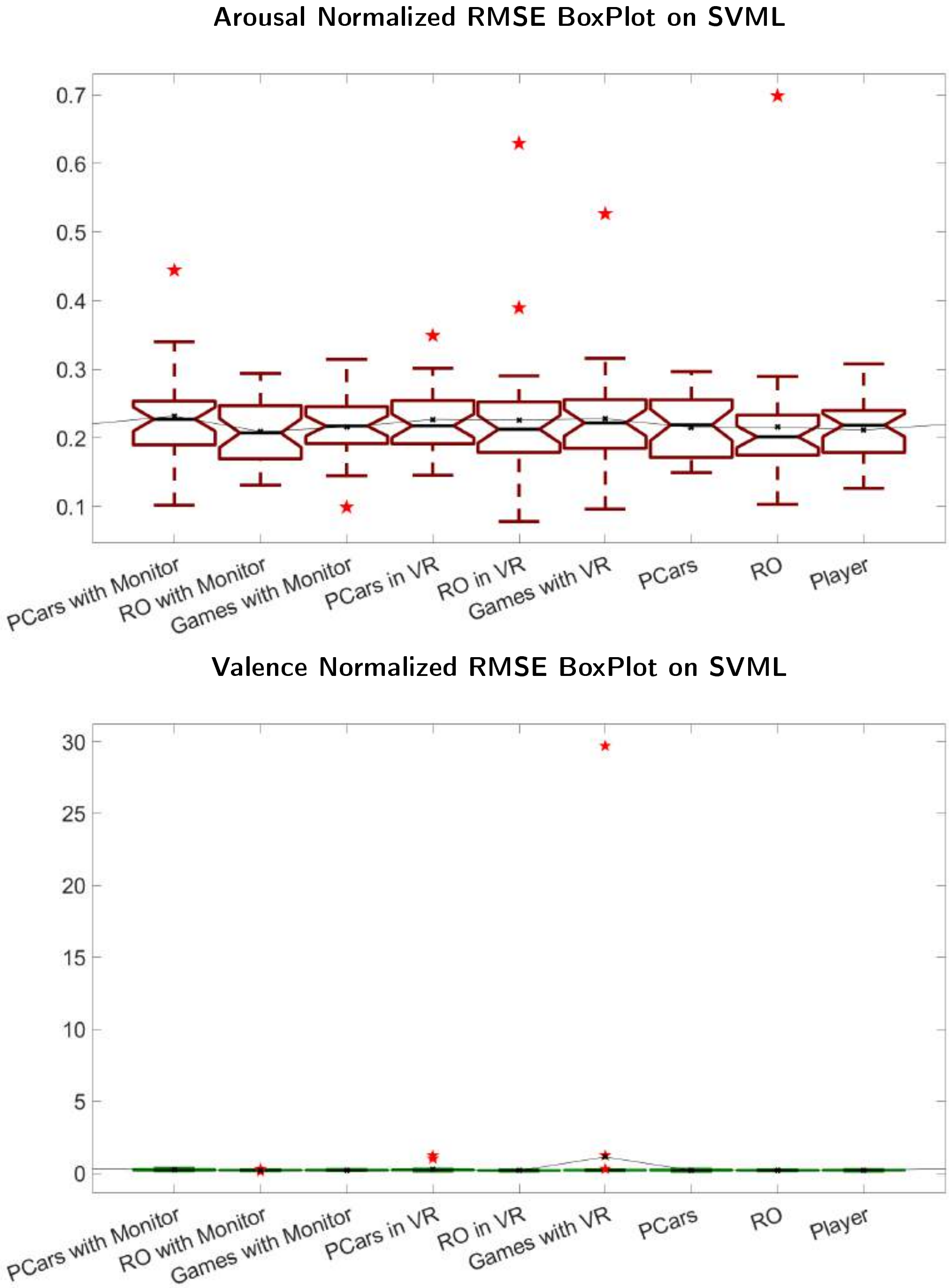
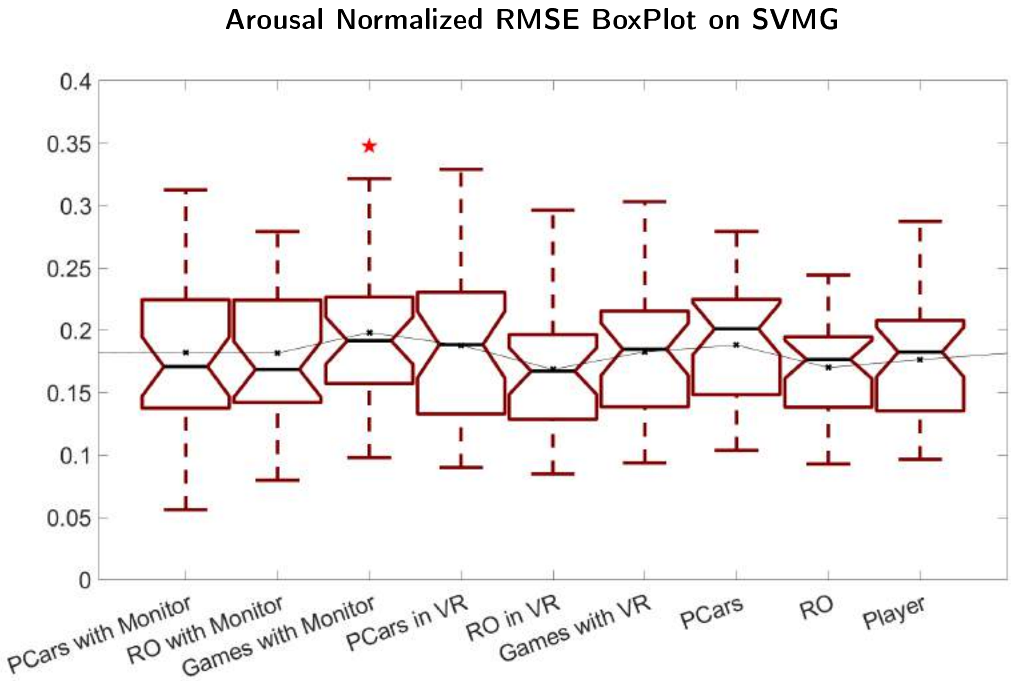


FIGURE B.4: Accuracy of the SVML algorithm on the observed data. We have used the NRMSE index as error metric. The outlier with the larger NRMSE is the user 20, which has suffered of a strong motion sickness. The cross inside the boxes underlines the average value, while the outliers are represented with the stars.



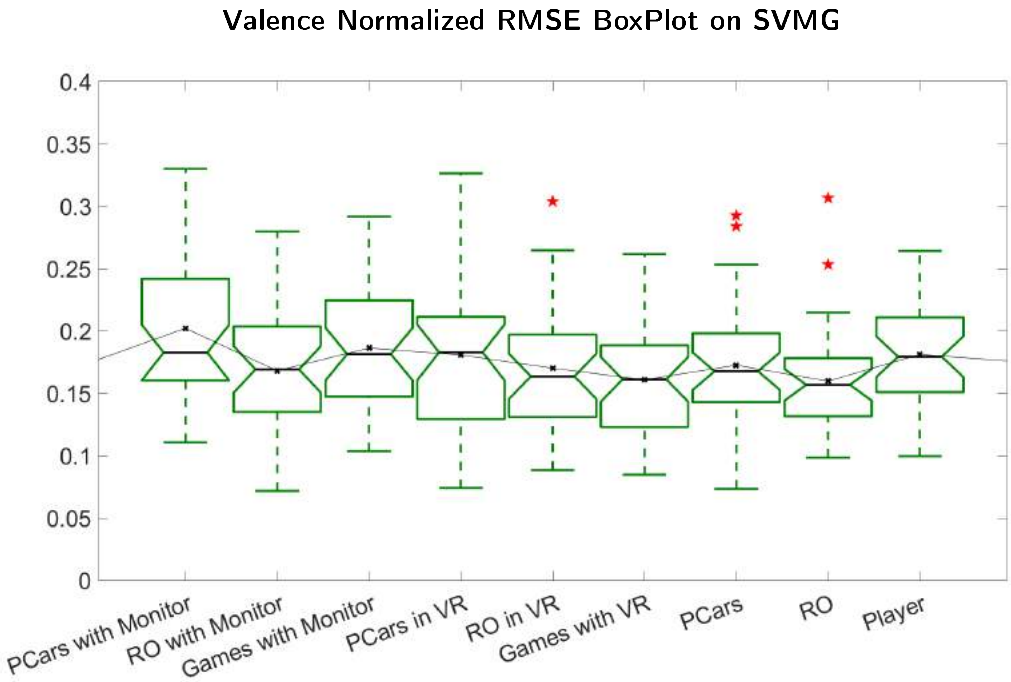


FIGURE B.5: Accuracy of the SVMG algorithm on the observed data. We have used the NRMSE index as error metric. The cross inside the boxes underlines the average value, while the outliers are represented with the stars.

Appendix C

Extended Results of GPR with Hyperparameters Tuning

In the following pages, we present the errors produced by the final estimator considering different indexes. For further details of the hypothesis design and the discussion of the outcomes see, respectively, Sec. 7.4.4 and Sec. 7.5.3. The considered error indexes are (where y are the observed data, $\hat{y}$ are the estimated data, and $\bar{y}$ is the average value of the observed data):

- **RMSE**: it is an error index frequently used in several researches. It measures the differences between 2 set of discrete values of the same length. It has always a value greater than 0, where 0 indicates a perfect fit of the predictor with the observed data. The result is dependent to the observed data scale (*scale-dependent*). Its formula is:
$RMSE(y,\hat{y}) = \sqrt{\frac{1}{N}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$
- **MAE**: it is the average vertical distance between the points and the identity line. The result is dependent to the observed data scale (*scale-dependent*). Its formula is:
$MAE(y,\hat{y}) = \sum_{i=1}^{N}\frac{|y_i - \hat{y}_i|}{N}$
- $\mathbf{R^2}$: it indicates the percentage (in a range between 0 and 1) of the variance in the observed data, which the estimated data explain collectively. The result is not dependent to the observed data scale (*scale-independent*). Its formula is $R^2(y,\hat{y}) = \frac{\sum_{i=1}^{N}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$
- **NRMSE**: it has the same meaning of RMSE, but it facilitates the comparison between data with different scales (e.g., results provided by further researches). Thus, the result is not dependent to the observed data scale (*scale-independent*). For more details see Sec. 7.5.2. Its formula is $NRMSE(y,\hat{y}) = \frac{RMSE(y,\hat{y})}{max(\hat{y}) - min(\hat{y})}$

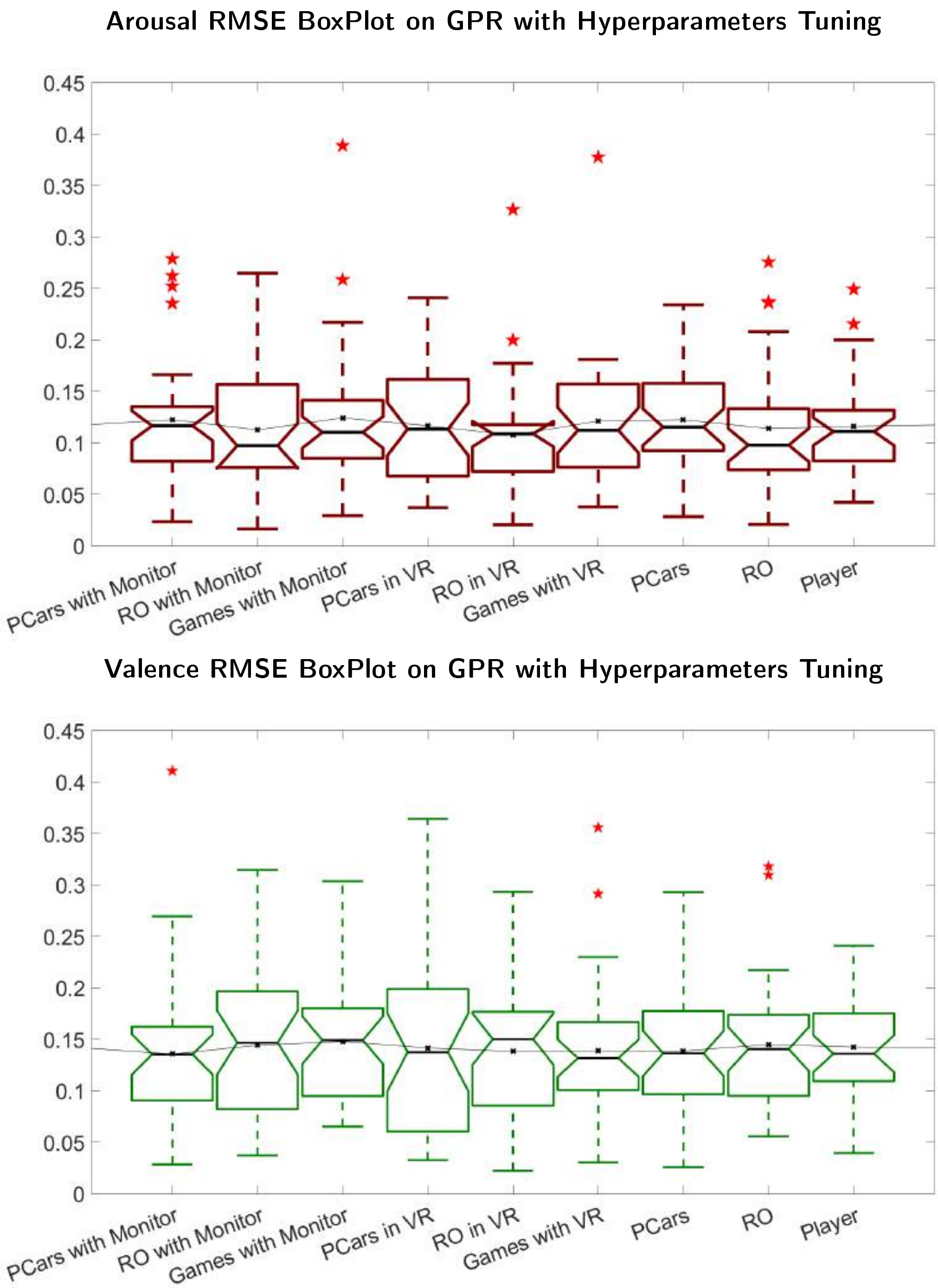


FIGURE C.1: The box plots present the RMSE distribution between the estimated data collected using the GPR (trained with different hyperparameters) and the observed self-assessment data across the experiments. The cross inside the boxes underlines the average value, while the outliers are represented with the stars.

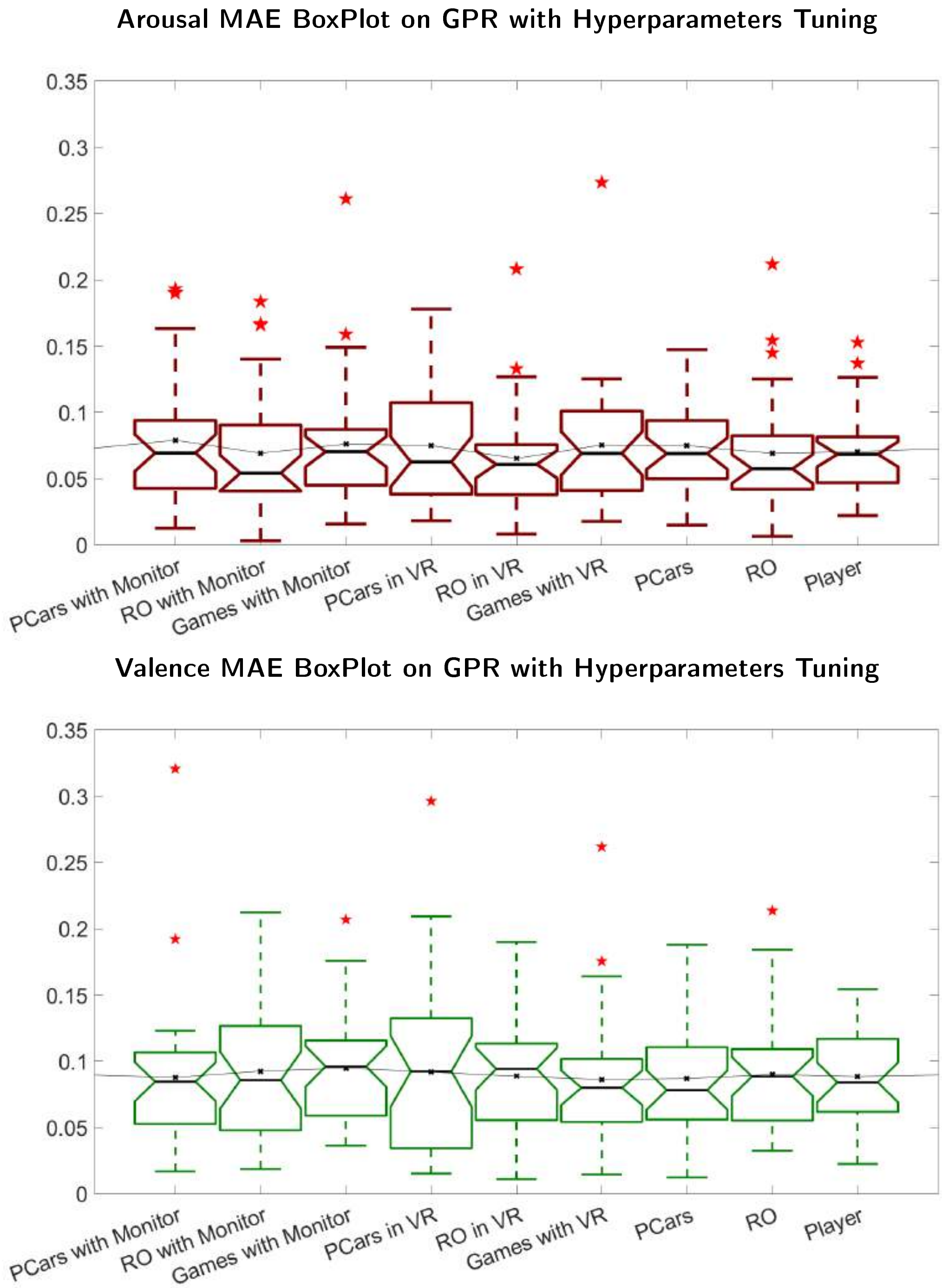


FIGURE C.2: The box plots present the MAE distribution between the estimated data collected using the GPR (trained with different hyperparameters) and the observed self-assessment data across the experiments. The cross inside the boxes underlines the average value, while the outliers are represented with the stars.

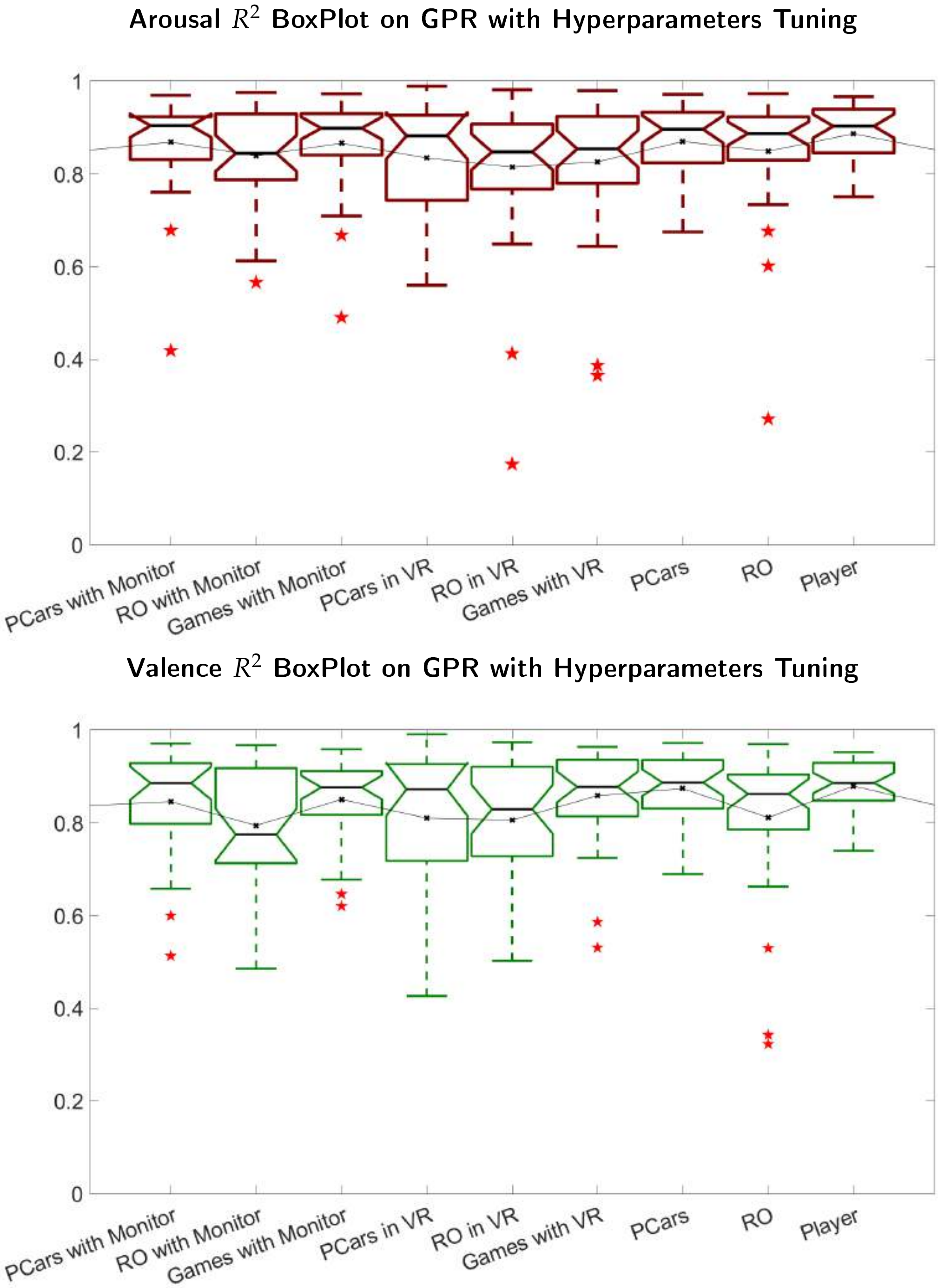


FIGURE C.3: The box plots present the R^2 distribution between the estimated data collected using the GPR (trained with different hyperparameters) and the observed self-assessment data across the experiments. The cross inside the boxes underlines the average value, while the outliers are represented with the stars.

Bibliography

[1] J. Huizinga, *Homo Ludens: A Study of the Play-Element in Culture*. Martino Fine Books, 2014.

[2] S. Tettegah and W. D. Huang, *Emotions, technology, and digital games*. Academic Press, 2015.

[3] A. Rollings and D. Morris, *Game architecture and design: a new edition*. New Riders Indianapolis, 2004.

[4] A. R. Damasio, D. Tranel, and H. C. Damasio, "Behavior: theory and preliminary testing," *Frontal lobe function and dysfunction*, vol. 217, 1991.

[5] A. R. Damásio, *Descartes' error: emotion, reason, and the human brain*. Quill, 1994.

[6] T. Grodal, "Video games and the pleasures of control," *Media entertainment: The psychology of its appeal*, pp. 197–213, 2000.

[7] I. Granic, A. Lobel, and R. C. Engels, "The benefits of playing video games.," *American Psychologist*, vol. 69, no. 1, p. 66, 2014.

[8] K. Isbister, *How games move us: Emotion by design*. Mit Press, 2016.

[9] C. Bateman, *Beyond game design: Nine steps toward creating better videogames*. Cengage Learning, 2009.

[10] K. Höök, "Affective loop experiences-what are they?," in *PERSUASIVE*, pp. 1–12, Springer, 2008.

[11] P. Sundström, *Exploring the affective loop*. PhD thesis, Department of Computer and Systems Sciences at the RISE, Swedish ICT, SICS, 2005.

[12] J. V. Draper, D. B. Kaber, and J. M. Usher, "Telepresence," *Human factors*, vol. 40, no. 3, pp. 354–375, 1998.

[13] M. Csikszentmihalyi and J. Nakamura, "The concept of flow," *Play and learning*, pp. 257–274, 1979.

[14] N. Ravaja, M. Salminen, J. Holopainen, T. Saari, J. Laarni, and A. Järvinen, "Emotional response patterns and sense of presence during video games: Potential criterion variables for game design," in *Proceedings of the third Nordic conference on Human-computer interaction*, pp. 339–347, ACM, 2004.

[15] R. Koster, *Theory of fun for game design*. " O'Reilly Media, Inc.", 2013.

[16] R. W. Picard, *Affective Computing (The MIT Press)*. The MIT Press, 2000.

[17] A. Drachen, P. Mirza-Babaei, and L. Nacke, *Games User Research*. Oxford University Press, 2018.

[18] D. R. Michael and S. L. Chen, *Serious games: Games that educate, train, and inform*. Muska & Lipman/Premier-Trade, 2005.

[19] R. Kremers, *Level Design: Concept, Theory, and Practice*. A K Peters/CRC Press, 2009.

[20] M. Spinka, R. C. Newberry, and M. Bekoff, "Mammalian play: training for the unexpected," *The Quarterly review of biology*, vol. 76, no. 2, pp. 141–168, 2001.

[21] G. Frasca, *Play the message: Play, game and videogame rhetoric*. PhD thesis, Center for Computer Games Research at the IT University of Copenhagen, 2007.

[22] K. Salen and E. Zimmerman, *Rules of play: Game design fundamentals*. MIT press, 2004.

[23] T. Fullerton, *Game design workshop: a playcentric approach to creating innovative games*. CRC press, 2014.

[24] J. Arjoranta, "Game definitions: A wittgensteinian approach," *Game Studies: the international journal of computer game research*, vol. 14, 2014.

[25] L. Wittgenstein, *Philosophische Untersuchungen = Philosophical investigations*. Chichester, West Sussex, U.K. Malden, MA: Wiley-Blackwell, 2009.

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

[27] N. Lazzaro, "Why we play games: Four keys to more emotion without story," *Games Developer Conference*, 2004.

[28] D. E. Freeman, *Creating Emotion in Games: The Craft and Art of Emotioneering*. New Riders, 2003.

[29] G. N. Yannakakis and A. Paiva, “Emotion in games,” *Handbook on affective computing*, pp. 459–471, 2014.

[30] G. N. Yannakakis and J. Togelius, “Experience-driven procedural content generation,” *IEEE Transactions on Affective Computing*, vol. 2, no. 3, pp. 147–161, 2011.

[31] A. Ortony, G. L. Clore, and A. Collins, *The cognitive structure of emotions*. Cambridge university press, 1990.

[32] D. De Felice, M. Granato, L. Ripamonti, M. Trubian, D. Gadia, and D. Maggiorini, “Effect of different looting systems on the behavior of players in a MMOG: Simulation with real data,” *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST*, vol. 181 LNICST, pp. 110–118, 2017.

[33] L. Ripamonti, M. Granato, M. Trubian, A. Knutas, D. Gadia, and D. Maggiorini, “Multi-agent simulations for the evaluation of looting systems design in MMOG and MOBA games,” *Simulation Modelling Practice and Theory*, vol. 83, pp. 124–148, 2018.

[34] R. A. Bartle, *Designing virtual worlds*. New Riders, 2003.

[35] S. N. Stahlke and P. Mirza-Babaei, “Usertesting without the user: Opportunities and challenges of an ai-driven approach in games user research,” *Computers in Entertainment (CIE)*, vol. 16, no. 2, p. 9, 2018.

[36] P. Mirza-Babaei, L. E. Nacke, J. Gregory, N. Collins, and G. Fitzpatrick, “How does it play better?: exploring user testing and biometric storyboards in games user research,” in *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 1499–1508, ACM, 2013.

[37] C. Darwin, *The expression of the emotions in man and animals*. Chicago University Press, USA, 1872.

[38] W. James, “What is an emotion?,” *Mind*, vol. 9, no. 34, pp. 188–205, 1884.

[39] C. G. Lange and W. James, *The emotions, Vol. 1*. Baltimore, Williams & Wilkins Company, 1922.

[40] W. B. Cannon, "The james-lange theory of emotions: A critical examination and an alternative theory," *The American journal of psychology*, vol. 39, no. 1/4, pp. 106–124, 1927.

[41] P. Bard, "A diencephalic mechanism for the expression of rage with special reference to the sympathetic nervous system," *American Journal of Physiology-Legacy Content*, vol. 84, no. 3, pp. 490–515, 1928.

[42] P. Ekman, "Universals and cultural differences in facial expressions of emotion.," in *Nebraska symposium on motivation*, pp. 207–282, University of Nebraska Press, 1971.

[43] P. Ekman, "Basic emotions," *Handbook of cognition and emotion*, pp. 45–60, 1999.

[44] T. Dalgleish, "The emotional brain," *Nature Reviews Neuroscience*, vol. 5, no. 7, p. 583, 20za04.

[45] I. B. Mauss and M. D. Robinson, "Measures of emotion: A review," *Cognition and emotion*, vol. 23, no. 2, pp. 209–237, 2009.

[46] J. A. Russell and L. F. Barrett, "Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant.," *Journal of personality and social psychology*, vol. 76, no. 5, p. 805, 1999.

[47] K. R. Scherer, "What are emotions? and how can they be measured?," *Social science information*, vol. 44, no. 4, pp. 695–729, 2005.

[48] C. E. Izard, *The psychology of emotions*. Springer Science & Business Media, 1991.

[49] P. Gable and E. Harmon-Jones, "The motivational dimensional model of affect: Implications for breadth of attention, memory, and cognitive categorisation," *Cognition and Emotion*, vol. 24, no. 2, pp. 322–337, 2010.

[50] H. Hoffmann, A. Scheck, T. Schuster, S. Walter, K. Limbrecht, H. C. Traue, and H. Kessler, "Mapping discrete emotions into the dimensional space: An empirical approach," in *Systems, Man, and Cybernetics (SMC), 2012 IEEE International Conference on*, pp. 3316–3320, IEEE, 2012.

[51] O. Villon and C. Lisetti, "A user model of psycho-physiological measure of emotion," in *International Conference on User Modeling*, pp. 319–323, Springer, 2007.

[52] M. M. Bradley and P. J. Lang, “Measuring emotion: the self-assessment manikin and the semantic differential,” *Journal of behavior therapy and experimental psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.

[53] A. Betella and P. Verschure, “The affective slider: A digital self-assessment scale for the measurement of human emotions,” *PLoS ONE*, vol. 11, no. 2, 2016.

[54] M. D. Robinson and G. L. Clore, “Episodic and semantic knowledge in emotional self-report: evidence for two judgment processes.,” *Journal of personality and social psychology*, vol. 83, no. 1, p. 198, 2002.

[55] S. Basu, J. Chakraborty, A. Bag, and M. Aftabuddin, “A review on emotion recognition using speech,” in *Inventive Communication and Computational Technologies (ICICCT), 2017 International Conference on*, pp. 109–114, IEEE, 2017.

[56] J. L. Tracy and R. W. Robins, “The nonverbal expression of pride: evidence for cross-cultural recognition.,” *Journal of personality and social psychology*, vol. 94, no. 3, p. 516, 2008.

[57] J. A. Russell, “Is there universal recognition of emotion from facial expression? a review of the cross-cultural studies.,” *Psychological bulletin*, vol. 115, no. 1, p. 102, 1994.

[58] M. N. Shiota, B. Campos, C. Oveis, M. J. Hertenstein, E. Simon-Thomas, and D. Keltner, “Beyond happiness: Building a science of discrete positive emotions.,” *American Psychologist*, vol. 72, no. 7, p. 617, 2017.

[59] P. Ekman and E. L. Rosenberg, *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, USA, 1997.

[60] E. Vural, M. Çetin, A. Erçil, G. Littlewort, M. Bartlett, and J. Movellan, “Machine learning systems for detecting driver drowsiness,” in *In-vehicle corpus and signal processing for driver behavior*, pp. 97–110, Springer, 2009.

[61] S. Rukavina, S. Gruss, S. Walter, H. Hoffmann, and H. C. Traue, “Open_emorec_ii-a multimodal corpus of human-computer interaction,” *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, vol. 9, no. 5, pp. 977–983, 2015.

[62] F. Shaffer, D. Combatalade, E. Peper, and Z. M. Meehan, “A guide to cleaner electrodermal activity measurements,” *Biofeedback*, vol. 44, no. 2, pp. 90–100, 2016.

[63] H. D. Critchley, "Electrodermal responses: what happens in the brain," *The Neuroscientist*, vol. 8, no. 2, pp. 132–142, 2002.

[64] E.-H. Jang, B.-J. Park, M.-S. Park, S.-H. Kim, and J.-H. Sohn, "Analysis of physiological signals for recognition of boredom, pain, and surprise emotions," *Journal of physiological anthropology*, vol. 34, no. 1, p. 25, 2015.

[65] P. H. Venables and M. J. Christie, "Electrodermal activity," *Techniques in psychophysiology*, vol. 54, no. 3, 1980.

[66] J. De Houwer and D. Hermans, *Cognition and emotion: Reviews of current research and theories*. Psychology Press, 2010.

[67] Q. Zhang, X. Chen, Q. Zhan, T. Yang, and S. Xia, "Respiration-based emotion recognition with deep learning," *Computers in Industry*, vol. 92, pp. 84–90, 2017.

[68] S. E. Rimm-Kaufman and J. Kagan, "The psychological significance of changes in skin temperature," *Motivation and Emotion*, vol. 20, no. 1, pp. 63–78, 1996.

[69] T. Partala and V. Surakka, "Pupil size variation as an indication of affective processing," *International journal of human-computer studies*, vol. 59, no. 1-2, pp. 185–198, 2003.

[70] S. D. Kreibig, "Autonomic nervous system activity in emotion: A review," *Biological psychology*, vol. 84, no. 3, pp. 394–421, 2010.

[71] R. M. Chapman and H. R. Bragdon, "Evoked responses to numerical and non-numerical visual stimuli while problem solving," *Nature*, vol. 203, no. 4950, p. 1155, 1964.

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

[73] 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, 2014.

[74] M. Granato, D. Gadia, D. Maggiorini, and L. Ripamonti, "Feature extraction and selection for real-time emotion recognition in video games players," in *Proceedings of the 14th International Conference on Signal-Image Technology and Internet-Based Systems, SITIS 2018*, 2018.

[75] R. Sinha, "Multivariate response patterning of fear and anger," *Cognition & Emotion*, vol. 10, no. 2, pp. 173–198, 1996.

[76] G. J. Boyle, "Reliability and validity of izard's differential emotions scale," *Personality and Individual Differences*, vol. 5, no. 6, pp. 747–750, 1984.

[77] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," *IEEE transactions on pattern analysis and machine intelligence*, vol. 23, no. 10, pp. 1175–1191, 2001.

[78] J. Healey and R. W. Picard, "Eight-emotion sentics data," 2002.

[79] A. McStay, *Emotional AI: The Rise of Empathic Media*. Sage, 2018.

[80] M. Clynes, *Sentics: The Touch of the Emotions*. Anchor Press/Doubleday, 1978.

[81] C. L. Lisetti and F. Nasoz, "Using noninvasive wearable computers to recognize human emotions from physiological signals," *EURASIP Journal on Advances in Signal Processing*, vol. 2004, pp. 1672–1687, Sep 2004.

[82] J. Kim and E. André, "Emotion recognition based on physiological changes in music listening," *IEEE transactions on pattern analysis and machine intelligence*, vol. 30, no. 12, pp. 2067–2083, 2008.

[83] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 42–55, 2012.

[84] M. Soleymani, J. Davis, and T. Pun, "A collaborative personalized affective video retrieval system," in *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, pp. 1–2, IEEE, 2009.

[85] J. R. Fontaine, K. R. Scherer, E. B. Roesch, and P. C. Ellsworth, "The world of emotions is not two-dimensional," *Psychological science*, vol. 18, no. 12, pp. 1050–1057, 2007.

[86] J. Jiao and M. Pantic, "Implicit image tagging via facial information," in *Proceedings of the 2nd international workshop on Social signal processing*, pp. 59–64, ACM, 2010.

[87] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for emotion analysis; using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012.

[88] S. Koelstra, A. Yazdani, M. Soleymani, C. Mühl, J.-S. Lee, A. Nijholt, T. Pun, T. Ebrahimi, and I. Patras, "Single trial classification of eeg and peripheral physiological signals for recognition of emotions induced by music videos," in *International Conference on Brain Informatics*, pp. 89–100, Springer, 2010.

[89] F. Ringeval, A. Sonderegger, J. Sauer, and D. Lalanne, "Introducing the recola multimodal corpus of remote collaborative and affective interactions," in *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*, pp. 1–8, IEEE, 2013.

[90] M. K. Abadi, R. Subramanian, S. M. Kia, P. Avesani, I. Patras, and N. Sebe, "Decaf: Meg-based multimodal database for decoding affective physiological responses," *IEEE Transactions on Affective Computing*, vol. 6, pp. 209–222, July 2015.

[91] S. Walter, J. Kim, D. Hrabal, S. C. Crawcour, H. Kessler, and H. C. Traue, "Transsituational individual-specific biopsychological classification of emotions," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 43, no. 4, pp. 988–995, 2013.

[92] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, "International affective picture system (IAPS): Technical manual and affective ratings," *NIMH Center for the Study of Emotion and Attention*, pp. 39–58, 1997.

[93] G. Boccignone, D. Conte, V. Cuculo, and R. Lanzarotti, "AMHUSE: A multimodal dataset for humour sensing," in *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, pp. 438–445, ACM, 2017.

[94] J. Broekens and W.-P. Brinkman, "AffectButton: A method for reliable and valid affective self-report," *International Journal of Human-Computer Studies*, vol. 71, no. 6, pp. 641–667, 2013.

[95] G. N. Yannakakis, H. P. Martínez, and A. Jhala, "Towards affective camera control in games," *User Modeling and User-Adapted Interaction*, vol. 20, no. 4, pp. 313–340, 2010.

[96] N. Shaker, S. Asteriadis, G. N. Yannakakis, and K. Karpouzis, "Fusing visual and behavioral cues for modeling user experience in games," *IEEE transactions on cybernetics*, vol. 43, no. 6, pp. 1519–1531, 2013.

[97] S. Asteriadis, K. Karpouzis, N. Shaker, and G. N. Yannakakis, "Towards detecting clusters of players using visual and gameplay behavioral cues," *Procedia Computer Science*, vol. 15, pp. 140–147, 2012.

[98] E. Hudlicka, "Affective computing for game design," *Proceedings of the 4th International North American Conference on Intelligent Games and Simulation (GAMEON-NA)*, 2008.

[99] N. Fourati and C. Pelachaud, "Perception of emotions and body movement in the emilya database," *IEEE Transactions on Affective Computing*, vol. 9, no. 1, pp. 90–101, 2018.

[100] R. Folgieri, C. Lucchiari, M. Granato, and D. Grechi, "Brain, technology and creativity. brainart: A BCI-based entertainment tool to enact creativity and create drawing from cerebral rhythms," in *Digital Da Vinci*, pp. 65–97, Springer, 2014.

[101] T. Christy and L. I. Kuncheva, "Technological advancements in affective gaming: A historical survey," *GSTF Journal on Computing (JoC)*, vol. 3, no. 4, 2018.

[102] B. Bontchev, "Adaptation in affective video games: A literature review," *Cybernetics and Information Technologies*, vol. 16, no. 3, pp. 3–34, 2016.

[103] G. N. Yannakakis and J. Hallam, "Entertainment modeling through physiology in physical play," *International Journal of Human-Computer Studies*, vol. 66, no. 10, pp. 741–755, 2008.

[104] L. Nacke and C. A. Lindley, "Flow and immersion in first-person shooters: measuring the player's gameplay experience," in *Proceedings of the 2008 Conference on Future Play: Research, Play, Share*, pp. 81–88, ACM, 2008.

[105] F. Levillain, J. O. Orero, M. Rifqi, and B. Bouchon-Meunier, "Characterizing player's experience from physiological signals using fuzzy decision trees," in *Computational Intelligence and Games (CIG), 2010 IEEE Symposium on*, pp. 75–82, IEEE, 2010.

[106] R. L. Hazlett, "Measuring emotional valence during interactive experiences: boys at video game play," in *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 1023–1026, ACM, 2006.

[107] S. Tognetti, M. Garbarino, A. Bonarini, and M. Matteucci, "Modeling enjoyment preference from physiological responses in a car racing game," in *Computational Intelligence and Games (CIG), 2010 IEEE Symposium on*, pp. 321–328, IEEE, 2010.

[108] P. Mirza-Babaei, S. Long, E. Foley, and G. McAllister, "Understanding the contribution of biometrics to games user research.," in *DiGRA conference*, Citeseer, 2011.

[109] V. Vachiratamporn, R. Legaspi, K. Moriyama, K.-i. Fukui, and M. Numao, "An analysis of player affect transitions in survival horror games," *Journal on Multimodal User Interfaces*, vol. 9, no. 1, pp. 43–54, 2015.

[110] M. Marras, D. Maggiorini, M. Granato, D. Gadia, and L. A. Ripamonti, "A touch-based configurable gamepad for gamers with physical disabilities," in *Proceedings of the International Conference on Computer-Human Interaction Research and Applications*, SCITEPRESS - Science and Technology Publications, 2017.

[111] L.-D. Liao, C.-Y. Chen, I.-J. Wang, S.-F. Chen, S.-Y. Li, B.-W. Chen, J.-Y. Chang, and C.-T. Lin, "Gaming control using a wearable and wireless eeg-based brain-computer interface device with novel dry foam-based sensors," *Journal of neuroengineering and rehabilitation*, vol. 9, no. 1, p. 5, 2012.

[112] D. Schwarz, V. Subramanian, K. Zhuang, and C. Adamczyk, "Educational neurogaming: Eeg-controlled videogames as interactive teaching tools for introductory neuroscience," in *Tenth Artificial Intelligence and Interactive Digital Entertainment Conference*, 2014.

[113] L. E. Nacke, M. Kalyn, C. Lough, and R. L. Mandryk, "Biofeedback game design: using direct and indirect physiological control to enhance game interaction," in *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 103–112, ACM, 2011.

[114] D. H. Emmen and G. Lampropoulos, "BioPong: Adaptive gaming using biofeedback," *Creating the Difference*, p. 100, 2014.

[115] G. Chanel, C. Rebetez, M. Bétrancourt, and T. Pun, "Emotion assessment from physiological signals for adaptation of game difficulty," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 41, no. 6, pp. 1052–1063, 2011.

[116] C. T. Tan, T. W. Leong, and S. Shen, "Combining think-aloud and physiological data to understand video game experiences," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 381–390, ACM, 2014.

[117] F. W. Petersen, L. E. Thomsen, P. Mirza-Babaei, and A. Drachen, "Evaluating the onboarding phase of free-toplay mobile games: A mixed-method approach," in *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, pp. 377–388, ACM, 2017.

[118] M. Zafar, B. Ahmed, R. Al-Rihawi, and R. Gutierrez-Osuna, "Gaming away stress: Using biofeedback games to learn paced breathing," *IEEE Transactions on Affective Computing - In Press*, 2018.

[119] J. M. Kivikangas, G. Chanel, B. Cowley, I. Ekman, M. Salminen, S. Järvelä, and N. Ravaja, "A review of the use of psychophysiological methods in game research," *journal of gaming & virtual worlds*, vol. 3, no. 3, pp. 181–199, 2011.

[120] A. Parnandi and R. Gutierrez-Osuna, "A comparative study of game mechanics and control laws for an adaptive physiological game," *Journal on Multimodal User Interfaces*, vol. 9, no. 1, pp. 31–42, 2015.

[121] M. B. Conover, *Understanding electrocardiography*. Elsevier Health Sciences, 2003.

[122] T.-M. Tsai, H.-C. Lin, S.-Y. Lee, and S.-J. Chang, "Heart rate detection through bone-conduction headset," in *Biomedical Circuits and Systems Conference (BioCAS), 2014 IEEE*, pp. 65–68, IEEE, 2014.

[123] F. Adib, H. Mao, Z. Kabelac, D. Katabi, and R. C. Miller, "Smart homes that monitor breathing and heart rate," in *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, pp. 837–846, ACM, 2015.

[124] C. Wang, T. Pun, and G. Chanel, "A comparative survey of methods for remote heart rate detection from frontal face videos," *Frontiers in bioengineering and biotechnology*, vol. 6, 2018.

[125] P. H. Bacchini, T. M. G. De, A. Barbosa, *et al.*, "Emopad: An affective gamepad," *International Journal of Computer Applications*, vol. 102, no. 15, 2014.

[126] M. K. Marks, M. South, and B. G. Carter, "Measurement of respiratory rate and timing using a nasal thermocouple," *Journal of clinical monitoring*, vol. 11, no. 3, pp. 159–164, 1995.

[127] A. D. Angie, S. Connelly, E. P. Waples, and V. Kligyte, "The influence of discrete emotions on judgement and decision-making: A meta-analytic review," *Cognition & Emotion*, vol. 25, no. 8, pp. 1393–1422, 2011.

[128] A. Furnham, "Response bias, social desirability and dissimulation," *Personality and individual differences*, vol. 7, no. 3, pp. 385–400, 1986.

[129] B. Sheffield, "The journey to create journey – the quest for emotion," February 2013.

[130] J. A. Russell, "Culture and the categorization of emotions.," *Psychological bulletin*, vol. 110, no. 3, pp. 426–450, 1991.

[131] J. A. Russell and A. Mehrabian, "Evidence for a three-factor theory of emotions," *Journal of research in Personality*, vol. 11, no. 3, pp. 273–294, 1977.

[132] M. Granato, D. Gadia, D. Maggiorini, and L. Ripamonti, "Software and hardware setup for emotion recognition during video game fruition," in *Proceedings of the International Conference on Smart Objects and Technologies for Social Good, GOODTECHS 2018*, 2018.

[133] Y. Renard, F. Lotte, G. Gibert, M. Congedo, E. Maby, V. Delannoy, O. Bertrand, and A. Lécuyer, "Openvibe: An open-source software platform to design, test, and use brain–computer interfaces in real and virtual environments," *Presence: teleoperators and virtual environments*, vol. 19, no. 1, pp. 35–53, 2010.

[134] C. J. Kalkman, "LabVIEW: a software system for data acquisition, data analysis, and instrument control," *Journal of clinical monitoring*, vol. 11, no. 1, pp. 51–58, 1995.

[135] R. Cowie, E. Douglas-Cowie, S. Savvidou, E. McMahon, M. Sawey, and M. Schröder, "'FEELTRACE': An instrument for recording perceived emotion in real time," in *ISCA tutorial and research workshop (ITRW) on speech and emotion*, 2000.

[136] R. Stojanović, A. Čaplánová, Ž. Kovačević, N. Nemanja, and Z. Bundalo, "Alternative approach to addressing infrastructure needs in biomedical engineering programs (case of emerging economies)," *Folia Medica Facultatis Medicinae Universitatis Saraeviensis*, vol. 50, no. 1, pp. 29–33, 2015.

[137] M. Granato, D. Gadia, D. Maggiorini, and L. Ripamonti, "Emotions detection through the analysis of physiological information during video games fruition," in *Springer Lecture Notes in Computer Science (Proceedings of 6th International Conference of Games and Learning Alliance - GALA 2017)*, vol. 10653, pp. 197–207, 2017.

[138] L. A. Ripamonti, M. Mannalà, D. Gadia, and D. Maggiorini, "Procedural content generation for platformers: designing and testing fun pledge," *Multimedia Tools and Applications*, pp. 1–50, 2016.

[139] C. Mazza, L. A. Ripamonti, D. Maggiorini, and D. Gadia, "Fun pledge 2.0: a funny platformers levels generator (rhythm based)," in *Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter*, p. 22, ACM, 2017.

[140] A. Van Boxtel, "Facial EMG as a tool for inferring affective states," in *Proceedings of measuring behavior*, pp. 104–108, Noldus Information Technology Wageningen, 2010.

[141] R. Rosenthal, "Experimenter effects in behavioral research.," 1966.

[142] S. Grimnes, "Impedance measurement of individual skin surface electrodes," *Medical and Biological Engineering and Computing*, vol. 21, no. 6, pp. 750–755, 1983.

[143] A. Savitzky and M. J. Golay, "Smoothing and differentiation of data by simplified least squares procedures.," *Analytical chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.

[144] M. Lavielle, "Detection of multiple changes in a sequence of dependent variables," *Stochastic Processes and their Applications*, vol. 83, no. 1, pp. 79–102, 1999.

[145] D. Dubin, *Rapid Interpretation of EKG's, Sixth Edition*. Cover Pub Co, 2000.

[146] A. Phinyomark, C. Limsakul, and P. Phukpattaranont, "A novel feature extraction for robust EMG pattern recognition," *Journal of Computing*, vol. 1, no. 1, pp. 71–80, 2009.

[147] P. Hamilton, "Open source ECG analysis," in *Computers in Cardiology, 2002*, pp. 101–104, IEEE, 2002.

[148] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[149] T. K. Ho, "Random decision forests," in *Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on*, vol. 1, pp. 278–282, IEEE, 1995.

[150] W.-Y. Loh, "Regression tress with unbiased variable selection and interaction detection," *Statistica Sinica*, pp. 361–386, 2002.

[151] H. R. Bonab and F. Can, "A theoretical framework on the ideal number of classifiers for online ensembles in data streams," in *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, pp. 2053–2056, ACM, 2016.

[152] D. Gadia, M. Granato, D. Maggiorini, L. Ripamonti, and C. Vismara, "Consumer-oriented head mounted displays: Analysis and evaluation of stereoscopic characteristics and user preferences," *Mobile Networks and Applications*, vol. 23, no. 1, pp. 136–146, 2018.

[153] C. Vismara, M. Granato, L. Ripamonti, D. Maggiorini, and D. Gadia, "Analysis of stereoscopic visualization in a consumer-oriented head mounted display," in *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST*, vol. 195 LNICST, pp. 274–283, 2017.

[154] D. Gadia, G. Garipoli, C. Bonanomi, L. Albani, and A. Rizzi, "Assessing stereo blindness and stereo acuity on digital displays," *Displays*, vol. 35, no. 4, pp. 206–212, 2014.

[155] E. Kreyszig, *Advanced Engineering Mathematics*. Wiley, 2011.

[156] A. Fedotov, "Selection of parameters of bandpass filtering of the ECG signal for heart rhythm monitoring systems," *Biomedical Engineering*, vol. 50, no. 2, pp. 114–118, 2016.

[157] O. Herrmann, "Design of nonrecursive digital filters with linear phase," *Electronics Letters*, vol. 6, no. 11, pp. 328–329, 1970.

[158] G. Boccignone, V. Cuculo, G. Grossi, R. Lanzarotti, and R. Migliaccio, "Virtual EMG via facial video analysis," in *International Conference on Image Analysis and Processing*, pp. 197–207, Springer, 2017.

[159] R. G. Lyons, *Understanding Digital Signal Processing*. Pearson Education India, 2011.

[160] D. T. Lykken and P. H. Venables, "Direct measurement of skin conductance: A proposal for standardization," *Psychophysiology*, vol. 8, no. 5, pp. 656–672, 1971.

[161] A. Oppenheim, *Discrete-time signal processing*. Upper Saddle River, N.J: Prentice Hall, 1999.

[162] E. J. Rechy-Ramirez and H. Hu, "Stages for developing control systems using EMG and EEG signals: A survey," *School of Computer Science and Electronic Engineering, University of Essex*, pp. 1744–8050, 2011.

[163] M. A. Oskoei and H. Hu, "GA-based feature subset selection for myoelectric classification," in *Robotics and Biomimetics, 2006. ROBIO'06. IEEE International Conference on*, pp. 1465–1470, IEEE, 2006.

[164] H.-P. Huang and C.-Y. Chen, "Development of a myoelectric discrimination system for a multi-degree prosthetic hand," in *Robotics and Automation, 1999. Proceedings. 1999 IEEE International Conference on*, vol. 3, pp. 2392–2397, IEEE, 1999.

[165] K. Englehart and B. Hudgins, "A robust, real-time control scheme for multifunction myoelectric control," *IEEE transactions on biomedical engineering*, vol. 50, no. 7, pp. 848–854, 2003.

[166] A. Phinyomark, C. Limsakul, and P. Phukpattaranont, "A novel feature extraction for robust EMG pattern recognition," *Journal of Computing*, vol. 1, no. 1, pp. 71–80, 2009.

[167] J.-S. Han, W.-K. Song, J.-S. Kim, W.-C. Bang, H. Lee, and Z. Bien, "New emg pattern recognition based on soft computing techniques and its application to control of a rehabilitation robotic arm," in *Proc. of 6th International Conference on Soft Computing (IIZUKA2000)*, pp. 890–897, 2000.

[168] E. A. Clancy, E. L. Morin, and R. Merletti, "Sampling, noise-reduction and amplitude estimation issues in surface electromyography," *Journal of electromyography and kinesiology*, vol. 12, no. 1, pp. 1–16, 2002.

[169] E. Clancy and N. Hogan, "Theoretic and experimental comparison of root-mean-square and mean-absolute-value electromyogram amplitude detectors," in *Engineering in Medicine and Biology Society, 1997. Proceedings of the 19th Annual International Conference of the IEEE*, vol. 3, pp. 1267–1270, IEEE, 1997.

[170] J. J. Braithwaite, D. G. Watson, R. Jones, and M. Rowe, "A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments," *Psychophysiology*, vol. 49, no. 1, pp. 1017–1034, 2013.

[171] D. L. Donoho *et al.*, "High-dimensional data analysis: The curses and blessings of dimensionality," *AMS Math Challenges Lecture*, vol. 1, p. 32, 2000.

[172] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Computers & Electrical Engineering*, vol. 40, no. 1, pp. 16–28, 2014.

[173] T. M. Oshiro, P. S. Perez, and J. A. Baranauskas, "How many trees in a random forest?," in *International Workshop on Machine Learning and Data Mining in Pattern Recognition*, pp. 154–168, Springer, 2012.

[174] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*. the MIT Press, 2006.

[175] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp. 1189–1232, 2001.

[176] H. Drucker, C. J. Burges, L. Kaufman, A. J. Smola, and V. Vapnik, "Support vector regression machines," in *Advances in neural information processing systems*, pp. 155–161, 1997.

[177] M. S. Caywood, D. M. Roberts, J. B. Colombe, H. S. Greenwald, and M. Z. Weiland, "Gaussian process regression for predictive but interpretable machine learning models: An example of predicting mental workload across tasks," *Frontiers in human neuroscience*, vol. 10, p. 647, 2017.

[178] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," in *Advances in neural information processing systems*, pp. 2951–2959, 2012.

[179] B. H. Menze, B. M. Kelm, R. Masuch, U. Himmelreich, P. Bachert, W. Petrich, and F. A. Hamprecht, "A comparison of random forest and its gini importance with standard chemometric methods for the feature selection and classification of spectral data," *BMC bioinformatics*, vol. 10, no. 1, p. 213, 2009.

[180] M. Claesen and B. De Moor, "Hyperparameter search in machine learning," *arXiv preprint arXiv:1502.02127*, 2015.

[181] S. Deterding, D. Dixon, R. Khaled, and L. Nacke, "From game design elements to gamefulness: defining gamification," in *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*, pp. 9–15, ACM, 2011.

[182] W. P. Krijnen, "Applied statistics for bioinformatics using R," *Institute for Life Science and Technology, Hanze University*, 2009.