POLITECNICO

MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

A New Virtual Reality Experimental Framework for Targeted Emotion Elicitation

TESI DI LAUREA MAGISTRALE IN
BIOMEDICAL ENGINEERING - INGEGNERIA BIOMEDICA

Authors: **Beatrice Prato, Andrea Signorelli**

Student IDs: 968799, 976126
Advisor: Prof. Riccardo Barbieri
Co-advisors: Edoardo Maria Polo, Maximiliano Mollura
Academic Year: 2021-22

Abstract

Emotions have a pivotal role in many cognitive functions thus it's important to study them in the context of human behaviour.

There is a need to define and characterize human emotion, and it can be performed using discrete and continuous models. The first rely on the universality of facial expressions to define a set of basic emotions; the latter use a multidimensional space to represent fundamental emotional characteristics, like the Circumplex Model of Affect that maps emotions in a space defined by valence, arousal, and dominance.

Affective computing uses these models to create systems that interact with human emotions. To build affective recognition models it's necessary to evoke specific emotions in controlled laboratory environments through rigorous elicitation methods.

This work explores the integration of Virtual Reality and Affective Computing; a novel experimental framework was developed in an immersive virtual environment to test the validity of VR as medium for targeted emotional triggers, against a traditional method that uses pictures from the OASIS database, presenting them through a Flat Screen (FS). The comparison was two-fold: an analysis of subjective data, collected by two post-elicitation surveys; and an analysis of physiological signals. Trends of the extracted features were observed and showed that VR-induced variations were generally more pronounced and showed a higher arousal than the Flat Screen-ones (as also confirmed by a Friedman's test). A square method for feature selection isolated the best three features for each method in differentiating the four emotions and a 3D graphical representation of the emotional separation was presented for both stimulations. A Machine Learning classification approach was carried out for both methods; the models trained on VR data outperformed the one trained on Flat Screen data in all cases.

This work defines a new VR-based protocol for targeted emotion elicitation that induces more specific responses in the subjects than traditional FS picture-based methods and proves the efficacy of VR as an emotional elicitation tool and its potential in enhancing the field of affective computing.

Keywords: Affective Computing, Virtual Reality, Emotion Elicitation

Abstract in lingua italiana

Le emozioni hanno un ruolo centrale in molte funzioni cognitive, è quindi importante studiarle nel contesto del comportamento umano.

È necessario, in primo luogo, definire e caratterizzare le emozioni utilizzando modelli discreti o continui. I primi si basano sull'universalità delle espressioni facciali per definire un insieme di emozioni di base; i secondi rappresentano le principali caratteristiche emotive in un continuo, come il Circumplex Model che mappa le emozioni tramite valenza, arousal e dominanza.

L'Affective Computing usa questi modelli per creare sistemi che interagiscano con le emozioni. Per costruire modelli di riconoscimento affettivo è necessario evocare emozioni specifiche in ambienti di laboratorio controllati con metodi di stimolazione rigorosi.

Questo lavoro esplora l'integrazione della Realtà Virtuale e dell'Affective Computing; è stato sviluppato un nuovo framework sperimentale in un ambiente virtuale immersivo per testare la validità della VR come mezzo per stimolare emozioni mirate, rispetto a un metodo tradizionale utilizzando immagini dal database OASIS. Il confronto è stato duplice: un'analisi dei dati soggettivi, raccolti tramite due sondaggi, e un'analisi dei segnali fisiologici raccolti. Sono state osservate i trend delle features estratte, le variazioni indotte dalla VR erano generalmente più pronunciate e mostravano maggior arousal rispetto a quelle della stimolazione a schermo, come confermato dal test di Friedman. Lo square method è stato usato per isolare le migliori tre features per ciascun metodo nella separazione delle quattro emozioni; è stata poi creata una rappresentazione 3D delle conseguenti separazioni emotiva. Per entrambe le stimolazioni è stato usato un metodo di classificazione Machine learning; i risultati dei modelli trainati sui dati VR hanno sempre superato i corrispettivi dai dati provenienti dalla stimolazione tradizionale. Questo lavoro definisce un nuovo protocollo basato sulla VR per l'elicitazione mirata delle emozioni che induce nei soggetti risposte più specifiche rispetto ai metodi tradizionali basati sulle immagini e dimostra l'efficacia della VR come strumento di elicitazione emotiva e il suo potenziale nel migliorare il campo dell'affective computing.

Parole chiave: Affective Computing, Realtà Virtuale, Machine Learning

Summary

Emotions serve a pivotal role in numerous cognitive functions, including rational decision-making, perception, and learning, among others. The physiological and psychological statuses of humans are profoundly influenced by emotions, thus underscoring the critical importance of recognizing and comprehending them in the context of human behavior research.

Affective Computing has emerged in the last few decades as a significant area of research that aims at the development of systems and devices able to recognize, interpret, process, and simulate human emotions. The interdisciplinary nature of this field encompasses psychophysiology, computer science, biomedical engineering, and artificial intelligence.

To create affective recognition models a lot of implicit data, such a physiological signals, is needed; thus a rigorous and effective way to elicit specific emotions is necessary. It's also important that researchers evoke affective states in controlled laboratory environments through various elicitation methods. To measure affection, it is necessary to define and characterize human emotion; this can be done through discrete and continuous models. Discrete models rely on the universality of facial expressions to define a set of basic emotions, such as the Pick-A-Mood (PAM) model or the Self-Assessment-Mannequin (SAM); Continuous models use instead a multidimensional space to represent fundamental emotional characteristics, like the Circumplex Model of Affect (CMA) in which emotive states are identified through valence, arousal, and dominance.

This work explores the integration of Virtual Reality and Affective Computing using rigorous quantitative methods, such as advanced signal processing, to extract and analyse high-quality features. A novel experimental protocol was developed in a fully immersive virtual environment to test the validity of the VR technology as medium for targeted emotional triggers, against an already validated picture-based method for emotion elicitation, using pictures from the Open Affective Standardized Image Set (OASIS).

The comparison was carried out in two ways: an analysis of self-reported data, collected with two post-elicitation assessment surveys using the previously cited PAM and SAM models to place the reported states on Russell CMA's two dimensions of arousal and valence; and an analysis of acquired physiological signals such as galvanic skin response

(GSR or EDA), respiration, Electrocardiograms (ECG) and blood volume pressure (BVP or PPG) collected using a ProComp device with four channels. From the four acquired physiological signals, a total of 30 features were extracted. Specifically, for the cardiac features a Point Process Model of Heart Rate Variability was created and 9 features were computed, both regarding the time component and spectral component of the signal; other cardiac features computed were the Pulse Arrival Time (PAT) and Pulse Pressure (PP).

Regarding the extracted features, their trends throughout the experiment were first observed, and showed that VR-induced variations in the features were generally more pronounced than the Flat Screen-induced ones, and that the VR stimulation generated a general higher level of arousal in the subjects; this was confirmed by a Friedman's test aimed at identifying statistically significant features in distinguishing between different emotions. The results showed that none of the extracted features from the more traditional elicitation were statistically significant, 17 features from the VR stimulation were instead significant in distinguish at least two of the elicitated emotional states, with GSR being the most relevant signal in this context. A Wilcoxon signed rank test was also carried out to assess differences between corresponding elicitated emotions in the two stimulations, only one feature resulted significantly different for all four emotions.

A square method for feature selection was carried out to isolate the best three features for each stimulating method in differentiating the four emotional states, and a 3D graphical representation of the emotional separation was presented, highlighting the better capability in distinguish the four emotions of the VR extracted features against the Flat Screen ones. This insight was then confirmed by analysing the outcomes of the machine learning classification models that were developed. Three different models were created for each stimulation method: a multiclass classification model for the four emotions and two binary classifiers to distinguish arousal level and valence level. Arousal separation achieved the highest accuracy, followed by valence separation and multiclass separation for both elicitation methods. The model trained on VR data consistently outperformed the one trained on Flat Screen data in all cases.

In conclusion, our work defines a new VR-based protocol for targeted emotion elicitation that induces more specific responses in the subjects than traditional flat screen image-based methods and demonstrates the efficacy of virtual reality as an emotional elicitation tool and its potential in enhancing the field of affective computing.

Contents

Introduction

Emotions play a pivotal role in rational decision-making, the learning process, general perception, and various other physiological and psychological functions in humans. Therefore, the recognition of emotions is a crucial aspect of human behavioural science.

In laboratory settings, affective states can be induced using various stimulation modalities that can both be active (such as Behavioral Manipulation and Active Recalling) and passive such as images, audio, videos, and, more recently, virtual reality (VR) and augmented reality (AR) [1][2].
The scientific interest and use of VR began in the 1990s and has significantly increased due to the advent of a new generation of low-cost headsets, made possible by the considerable investments in the gaming market and the rise of new VR hardware and software companies [3]. These new devices and the developer tools that come with them, led to the start of a change in the direction of the scientific studies in Affective Computing.

VR presents a unique prospect to create simulated environments that can elicit a powerful sense of presence while providing a cost-effective mean of isolating and controlling the variables of an experiment, which is unfeasible in real-world settings. The integration of VR, physiological signal recordings, and advanced signal processing techniques offers novel prospects to extract valuable features that can be utilized for emotion recognition.

Previous studies examining the emotional aspects and elicitation potential of VR have shown a tendency towards moderation. These studies focused on the feasibility of eliciting diverse emotional states using simple stimuli, such as static images or moderately intricate settings, without fully utilizing the potential of virtual reality as a medium for stimulation [4]. In real-world scenarios, it's not possible to isolate emotional stimuli. Therefore, a VR protocol tests the hypothesis that more immersive and realistic scenarios may be better triggers for specific emotional responses compared to traditional visual stimulation using validated pictures on a flat screen surface.

The aim of this work is to develop a new framework for targeted emotion elicitation using VR; to compare it with a traditional flat screen, image-based method, and evaluate the differences in subjects' responses, both through subjective assessments and analysis of

physiological signals. The results will help to identify the strengths and weaknesses of this new approach in respect with the traditional one.

The resulting data will also be used to train a machine learning classifier that can separate the emotional state of a person and two binary classifiers that separate levels of arousal and valence, based on a certain subset of features derived from the acquired physiological signals. This tools could be used to more accurately assess the emotional responses of future subjects, improving the efficacy of the elicitation method.

Thesis Outline

To better understand the structure of the document a brief outline of what will follow in the next chapters is presented below:

In Chapter 1 an introduction to the context of emotion and the field of Affective Computing is proposed, discussing then the Autonomic Nervous System and its role in this field. The focus will then shift to affective modeling and affective elicitation modalities, introducing in more detail the most commonly used and the chosen ones for this thesis work. The chapter ends with an introduction to the physiological signals acquired during this project.

Chapter 2 will present the materials and the methods used in this thesis work. At first the tools and strategies used to build the stimulation protocol are explained in detail, for both the traditional, the VR elicitation and the post experience emotional assessments. Then the focus will shift to signal acquisition and processing, and move to the performed Statistical Analysis and feature selection. To conclude the used Machine Learning approaches are described in the last portion of the chapter.

In Chapter 3 the results obtained from the various analysis described in the previous chapter are presented, so it will follow a similar outline. At firs the results from the two post-elicitation assessment are shown and compared in detail for each part of the stimulation. The results from the physiological response analysis and the statistical analysis are then presented in detail. The chapter will conclude with the feature selection results and machine learning models performances.

The document will end in chapter 4 with a critical discussion of the obtained results and the conclusions, which will summarize the biggest achievements of this work, its limitations and possible future improvements.

1 | Background

Emotions are very complex phenomena that make up a significant part of human life; they affect behaviour, perception, memory retrieval and cognition. They can be described as a sensory condition that leads to instinctive physical and psychological changes [5].

1.1. Affective Computing

To investigate the connections and interactions between emotional states, cognition and decision-making processes, the field of Affective Computing was introduced. Affective Computing analyses human responses using implicit measures, such as psychophysiological responses, with the aim to develop automatic emotion recognition tools using a combination of computer science and physiology [6], aspiring to narrow the communicative gap of human-computer interfaces [7].

Various sources are efficiently used in Affective Computing as an input to classify emotions [7]. We will briefly introduce the most used ones (fig.1.1):

- **Voice**: paralinguistic features of speech; affective information is carried in what is said (explicit message, often analysed through text) and how it is said (implicit paralinguistic message), for example pitch appears to be an index of arousal.

- **Facial Expressions**: online face video capturing, used to track face movements in real time and use them to identify expressions linked to specific emotions.

- **Body language and posture**: usually acquired through pressure sensors and video tracking, posture and body movements contain a great amount of nonverbal affective information.

- **Neuroimaging**: new developments in brain imaging, such as Functional Magnetic Resonance Imaging (fMRI), are used in attempts to map the neural pathways that are involved in emotional experiences and use them to identify emotions.

- **Physiological Activity**: research in this field is known as psychophysiology and it studies how an affective stimulus (independent variable) influences a physiological measure (dependent variable). Most of the measures used are "non-invasive" such as:

  - Electroencephalography (EEG)

  - Respiration (RESP)

  - Galvanic Skin Response (GSR)

  - Skin temperature (SKT)

  - Electromyogram (EMG)

  - Electrooculogram (EOG), measuring eye movement

  - Electrocardiography (ECG or EKG)

  - Heart rate (HR) and Heart Rate Variability (HRV)

  - Photopletismography (PPG), used as a non-invasive measure for blood pressure

All these physiological signals bring information related to the central nervous system activity (the EEG signal in particular) and the autonomic central system (ANS) activity [8].
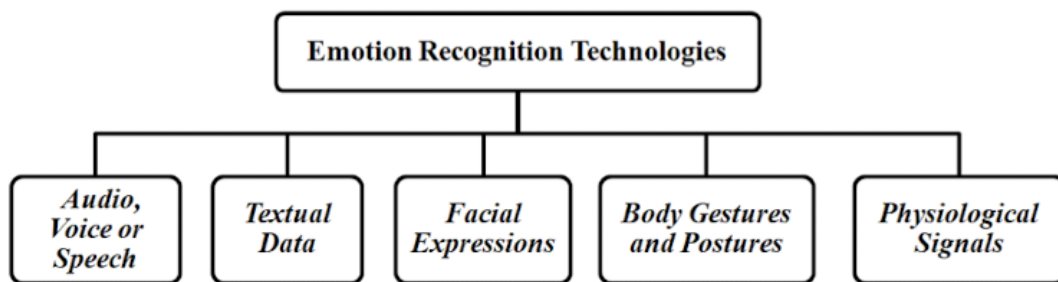
Figure 1.1: Emotion Recognition Technologies

1.1.1.   Autonomic Nervous System

The **Autonomic Nervous System (ANS)** is the portion of the nervous system that innervates various internal organs: stomach, intestine, genitals, lungs, bladder, pupils, heart, blood vessels, kidneys and liver and multiple glands (digestive, sweat, salivary). Also, skin is innervated by the sympathetic system only. A graphical representation of ANS can be seen in figure 1.2.

The ANS regulates body processes such as the respiration pattern, the heart rate, the contraction of the digestive tract's muscles or the blood pressure; all these processes are autonomous, they occur without conscious exertion from the subject, thus the name *"Autonomous Nervous System"* (or Autonomic).

The ANS can be divided in two subsystems:

- **Sympathetic nervous system**

- **Parasympathetic nervous system**
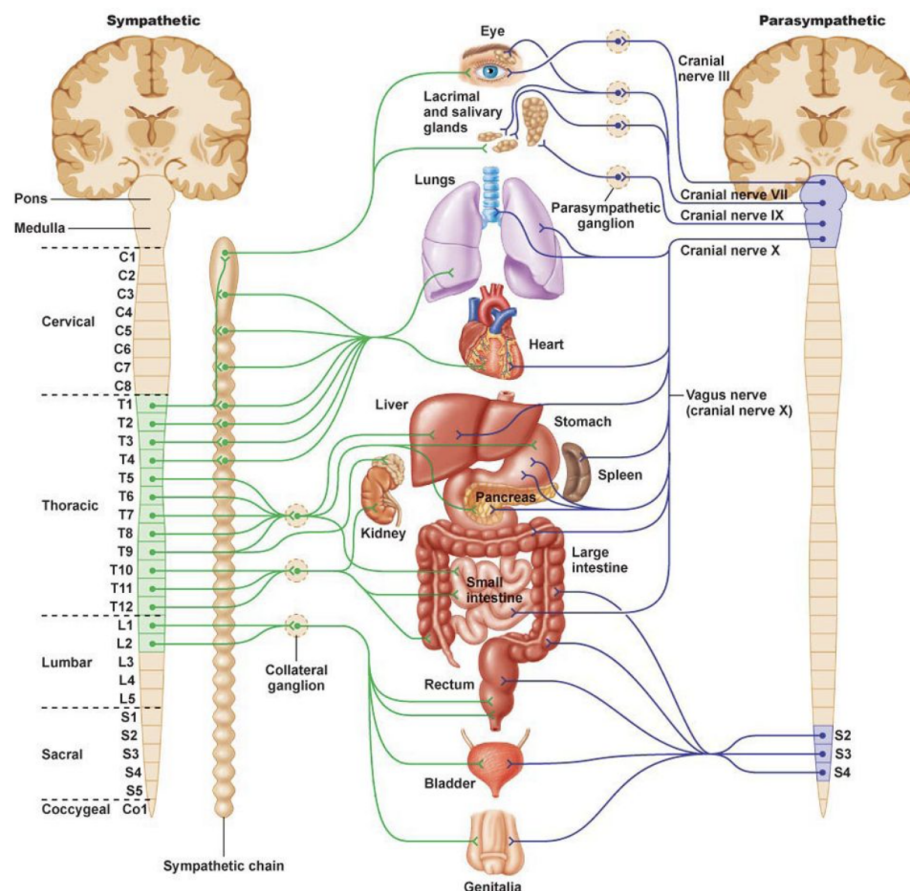
Figure 1.2: The Autonomic Nervous System and its branches.

These two subsystems innervate mostly the same organs and tissues but have diametrically opposed effects: the sympathetic system prepares the body for action by increasing the heart rate, blood pressure and accelerating the respiratory rate, the so called *"fight or flight"* response; in contrast, the parasympathetic (or vagal) system predisposes the body to rest through the *"rest and digest"* response: it slows the heartbeat and respiratory rate and reduces blood pressure, concentrating the blood flow and energy resources to processes like digestion [9].

Different affective states such as relaxation, sadness or fear are defined by different activity of the parasympathetic and sympathetic systems, and these differences can be studied observing physiological signals as was done in numerous studies.

The results of these studies, in which different specific ANS activations are associated with four main emotions (of which we will discuss in more detail in the following section), can be summarized as follows [10]:

- **Sadness**: The feeling of sadness is physiologically characterized by a heterogeneous pattern of parasympathetic-sympathetic co-activation. Regarding sadness there's an important differentiation between a crying and a non-crying sadness status; physiologically a crying sadness is correlated to an uncoupled sympathetic activation and a non-crying status is linked to parasympathetic-sympathetic withdrawal.

- **Relaxation or contentment**: Characterized by a strong deactivation of the sympathetic branch and mild cardiac vagal activation, related to a HR deceleration and a decrease of both respiratory and electrodermal (GSR) activation.

- **Happiness**: The autonomic response pattern for the feeling of happiness is distinguished by vagal withdrawal, that causes increased cardiac activity, increased respiratory activity, vasodilation, and increased electrodermal activity. The physiological response to happiness includes increased HR and increased blood pressure, but a few exceptions can occur in the case of elicitation with visual material (videos, pictures): instead of the typical increased HR some studies reported unchanged or decreased HR; this variance may lead to the possibility that a large range of positive feelings is clustered under the umbrella term *"happiness"*.

- **Fear**: The physiological effect of fear is related to broad sympathetic activation, that causes vasoconstriction, increased myocardial contractility, cardiac acceleration (increased HR) and increased electrodermal activity; this response is accompanied by decreased cardiac vagal influence, that increases respiratory activity, leading to faster breathing resulting in decreased carbon dioxide blood levels.

1.2.   Affective Models

To measure affection, we first need to define and model human emotions.

Although the two terms are often used interchangeably in scientific literature, affect and emotion refer to two different things: in [11], affect is defined as a short-term perceptual feeling of pleasure-displeasure (valence) and of calmness-excitement (arousal), while emotion is a long-term complex mental construction.

Both can be measured through signals in the form of bodily activation and responses to inputs.

In this work, we will use these two terms as in literature, but it is important for the reader to understand their differences; both terms will refer to emotion and not affect.

Proposed models of affection can be divided in two categories: continuous models and discrete models.

1.2.1.   Continuous Models

Circumplex Model of Affect (CMA)

This model was the first to represent affective states as a circle in a two-dimensional bipolar space, taking as variables the ranges of valence and arousal felt by the subject [12]. Such values are then used as coordinates and, depending on the position of the resulting point, the most likely emotion felt is selected. When the model was developed, the ranges of valence and arousal were derived from a subjective assessment through the Mehrabian and Russell's state affect scales [13], and the results were linked with a rating of how accurately 518 adjectives described how the subjects were feeling. In the end, 15 affect states were derived and identified into the two-dimensional space, defining the Circumplex Model (fig. 1.3).

Notice that this graph represents a specter, a continuous space where affective states overlap over one another.
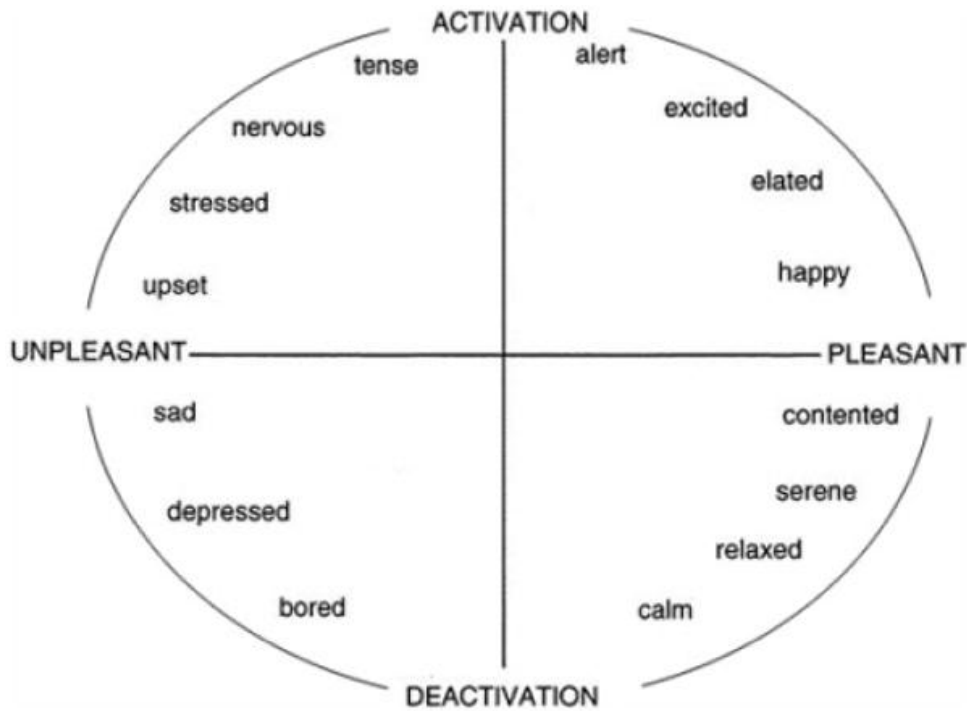
Figure 1.3: 2-dimensional CMA

1.2.2.   Discrete Models

Discrete models define a set of basic emotions such as anger, disgust, tension, happiness, sadness, and surprise; this method is based on the concept of emotional facial expressions being universal, transcending cultural differences, an example of this approach, that also finds application in this thesis work, is the Pick-A-Mood model [14].

Pick-A-Mood Model (PAM)

The Pick-A-Mood (PAM) model is a cartoon-based graphical tool for reporting and expressing subjective moods. The idea is to use cartoon characters instead of pictures of real people in order to enable the subjects to unambiguously and visually express their mood in an intuitive way. PAM requires little time and effort to be compiled, which makes it useful for design research applications. This model focuses on moods, which are not evoked by a single stimulus (unlike emotions) but are rather complex phenomena that are difficult to explain and describe for people. At the same time, mood can be influenced by little changes in the environment and in one's emotional state and can bring to highly different behaviors [15]. Finally, mood can only be measured through self-report.

The PAM model represents eight moods that can be grouped in four main mood cate-

gories:

- excited and cheerful (energized-pleasant),

- irritated and tense (energized-unpleasant),

- relaxed and calm (calm-pleasant),

- bored and sad (calm-unpleasant),

plus a neutral mood.

The final iteration of the PAM model is composed of nine drawings, each for one mood state in the form of a neutral-aged woman, as can be seen in figure 1.4.

Figure 1.4: Pick-A-Mood Model graphical representation.

SAM

The Self-Assessment Manikin (SAM) is a pictorial assessment technique that directly measures the pleasure, arousal, and dominance associated with a person's affective reaction to stimuli [16]. A different set of five images was developed to subjectively assess each variable, and it was validated against the Semantic Differential Scale (SDS), which is a psychometric technique used to measure the attitude of an individual towards an object or concept. It is an adaptation of the Likert Scale [17], that is a 7-point scale made up of bipolar adjectives (for example, good-bad, strong-weak, etc.) that are used to measure the connotation of a stimulus.

Other then the already cited valence and arousal, the SAM model also measures dominance, which is the degree of control perceived by the subject in response to a stimulus, be it an image, a video, an event etc.

Below, the graphical representations of the SAM Arousal (fig. 1.5), Valence (fig. 1.6) and Dominance (fig. 1.7) scales are reported.
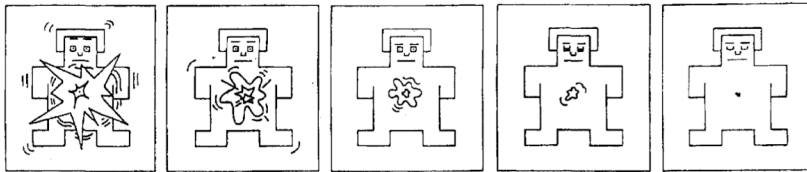
Figure 1.5: SAM Arousal graphical representation
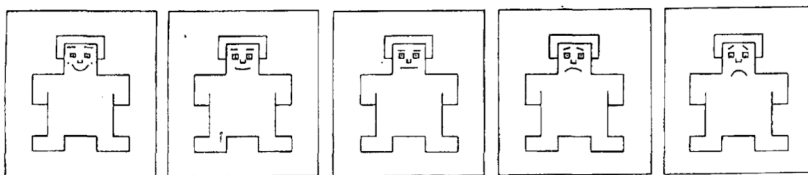
Figure 1.6: SAM Valence graphical representation
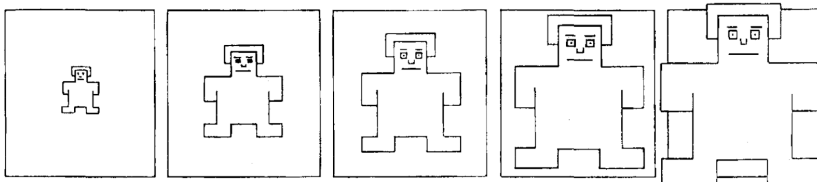
Figure 1.7: SAM Dominance graphical representation

1.3.   Affective Elicitation Methods

Emotion elicitation refers to the capacity to consistently and ethically elicit affective states. This process is essential in the development of a system capable of detecting, interpreting, and adapting to human emotions. Elicitation may be either active or passive, with the former utilizing means such as behavioral manipulation, social interaction, and dyadic engagement to directly impact individuals. Conversely, passive methods typically present stimuli through visual, auditory, or audiovisual media [1].

1.3.1.   IAPS Database

Behavioral and brain sciences often rely on the use of images that can elicit a diverse range of emotions associated with social and non-social phenomena. To facilitate such research, the Center for the Study of Emotion and Attention started developing the International Affective Picture System (IAPS) [18], which is a standardized and internationally available set of visual emotional stimuli. The latest version of the IAPS consists of 1,195 color photographs, each of which has been assigned normative ratings on three dimensions: valence, arousal, and dominance. The IAPS has been extensively utilized in psychological, psychophysiological, and neuroscience research, covering a broad spectrum of topics. However, it should be noted that the IAPS was developed in a pre-Internet era, and the images are therefore dated and subject to copyright restrictions.

Figure 1.8: Sample of pictures from the IAPS.

1.3.2. OASIS Database

The Open Affective Standardized Image Set (OASIS) [19] is a freely accessible online database that includes 900 color images depicting a wide range of themes such as humans, animals, objects, and scenes. The images are accompanied by normative ratings on two affective dimensions, valence and arousal, which measure the positive or negative affective response and the intensity of the affective response, respectively. The images in the OASIS database were sourced from online platforms, and the related ratings were obtained through an online study involving more than 800 participants, with diverse backgrounds and ages. The valence and arousal ratings covered a significant proportion of the CMA's space and exhibited high reliability and consistency across gender groups.

There are four key advantages of the OASIS dataset. First, it contains a substantial number of images across four distinct categories. Second, the data collection process was conducted in 2015, allowing for a more contemporary selection of images and up-to-date ratings of valence and arousal than pre-existing datasets [19]. Third, the OASIS database allows users to interactively explore images by category and ratings. Finally, and most importantly, OASIS gives free use of its images for both online and offline research studies, as they are not subject to copyright restrictions that apply to other similar databases, such as the IAPS.
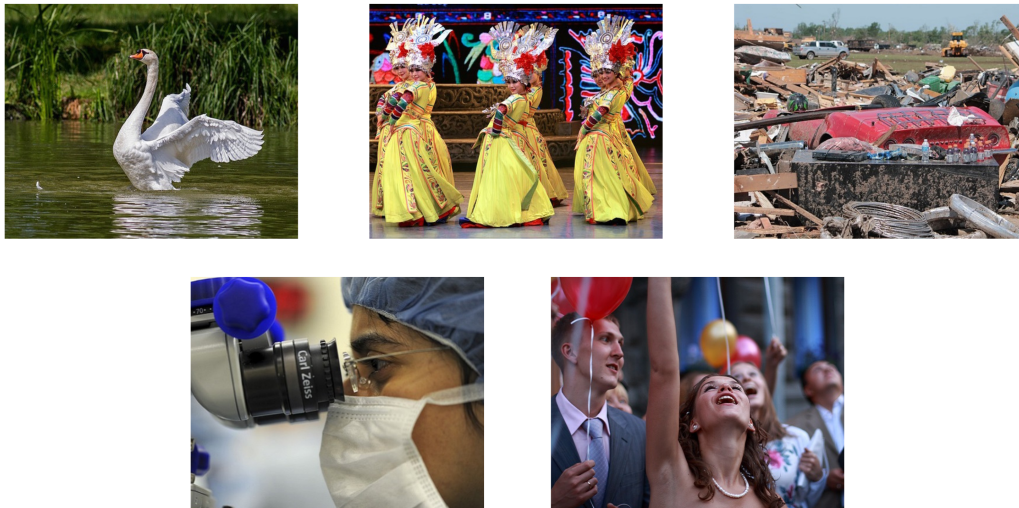
Figure 1.9: Sample of pictures from the OASIS.

1.3.3. Virtual Reality

VR can be defined as a computer-simulated environment that replicates physical presence in a three-dimensional space. It is achieved through the use of specialized hardware and software (usually VR headsets and hand controllers) to generate a simulated environment that allows users to interact with objects, people, and entities, although VR can also be used in a passive way, like through the vision of 360° movies. The experience of being inside the virtual world can be enhanced by the use of immersive audio, visual, and haptic (tactile) feedback.

VR has been used as a medium for games, entertainment, education, physical and cognitive training, rehabilitation and therapy. In recent years, studies have registered an increase in interest in human emotional behavioural studies using VR to evoke emotions in the subjects [5].

The previously cited methods of emotional elicitation use their stimulating content as if it possessed intrinsic affective qualities, without taking context into consideration. The response of a subject will be different if they see the image of a spider on a computer's screen, and if they see it in real life. As in [20], *«there is a lack of ecological validity to presenting a stimulus that has an assumed intrinsic emotional quality but is otherwise disconnected from its meaning in other components of emotion»*.

Where talking about "ecological validity", the VR-related concepts of immersion and presence are brought up. Immersion, that *«relates to the feeling of being physically present in a nonphysical world»* [21] is achievable through the removal of real-world sensations, substituting them with virtual ones that absorb the user, limiting a sense of separation from what is experienced in the simulation. As in [22], *«Immersion is intended to instill a sense of belief that one has left the real world and is now "present" in the virtual environment»*. Presence is the psychological, subjective, cognitive consequence of the more objective and technologically related immersion and can be thought of as the perception of *"being in"* and *"existing in"* the virtual environment.

Through enhanced immersion and, thus, presence, the connection between the subject and the stimuli is reinforced, bringing to heightened responses than those acquired through more traditional protocols.

VR can be integrated with numerous elicitation modes as bidimensional images, videos, games, 360-degree panoramas or a combination of more modes, with or without other sensorial information than visual ones.

As a result, with proper development, VR can be used to represent complex, realistic and three-dimensional scenarios that are fully controlled by the developer, without taking into consideration the complexities of building or finding the same scenario in physical life.

To create such environments, free and easy-to-use software such as Unity, or similar game engines, can be used.

Figure 1.10: Examples of commercial VR headsets.

1.4. Physiological Signals

Various sources can be efficiently used in Affective Computing. This work will focus on physiological activity, by acquiring signals such as the Electrocardiogram, Blood Volume Pulse, Galvanic Skin Response and Respiration. In the following section an introduction to these signals is proposed.

1.4.1. ECG

Electrocardiograms (ECGs) have been a subject of scientific investigation for over a century. ECGs are essential biomedical signals, typically depicted as a graph of voltage versus time that represents the electrical activity of the heart. Electrodes placed on the skin detect the electrical changes that occur during each cardiac cycle as a result of cardiac muscle depolarization and re-polarization.

The typical morphology of the ECG signal is shown in figure 1.11 and consists of three segments within a single cycle. The first segment, known as the P wave, corresponds to atrial depolarization. The second segment is the QRS complex, which is the highest amplitude waveform caused by ventricular depolarization. The distance between successive R peaks is used to obtain the inter-beat interval (IBI), also known as RR interval from the name of the peaks, for heart rate detection. QRS detection is essential for extracting heart rate variability (HRV) features from ECG signals. The T wave appears after a few milliseconds of plateau, as a consequence of ventricular re-polarization, and the cycle repeats. ECG signals are considered high-sensitivity physiological signals, with a low recording voltage ranging from 0.5 to 5 mV. The typical frequency for high-quality ECG acquisition is 256 Hz or higher, while the minimum frequency required to appreciate the QRS morphology of the ECG signal is 100 Hz [23]. The brain and heart are connected through the ANS, whereby both organs influence each other's behavior in a closed loop. As such, emotional experiences can cause changes in heart rhythm, which can be detected through ECG readings.
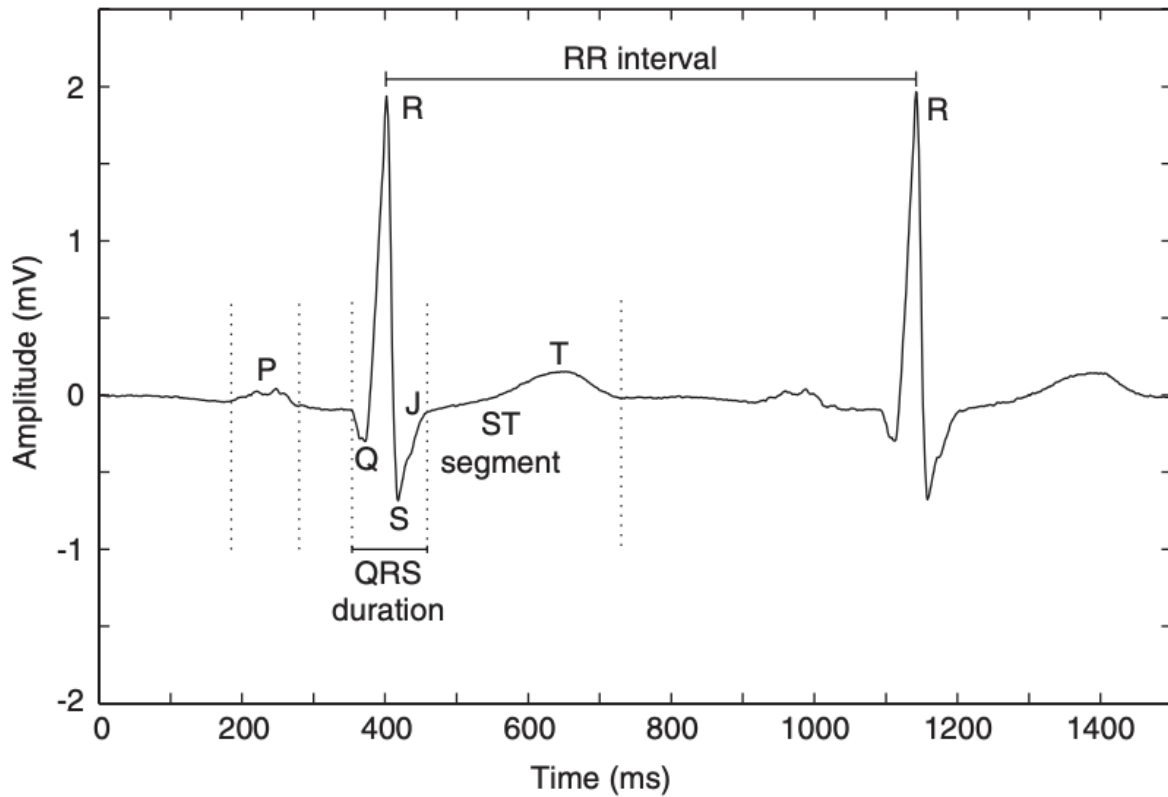
Figure 1.11: ECG wave

1.4.2.   Blood Volume Pulse

The Blood Volume Pulse (BVP) signal is a valuable tool for extracting blood pressure information, when acquired using specific sensors it can also be called Photoplethysmography (PPG) (fig.1.12). BVP is composed of a pulsatile (AC) component and a superimposed (DC) component [24], where the AC component is contributed by variations in blood volume that occur synchronously with cardiac activity; meanwhile, the DC component is shaped by respiration, sympathetic nervous system activity, and thermoregulation. The AC component of BVP reflects changes in blood volume, which are dependent on the systolic and diastolic phases. The systolic phase, referred to as the "rise time," begins with a valley and concludes with the pulse wave systolic peak, while the end of the diastolic phase is marked by another valley. The PPG signal is divided into two blocks: the rising edge of the pulse (anacrotic) corresponding to systole and the falling edge (catacrotic) corresponding to diastole. In addition, a dicrotic notch is typically observable at the catacrotic phase.

Features such as rise time, amplitude, and shape can predict vascular changes in the

bloodstream. Moreover, PPG can be used in addition or alternative to the ECG to assess HRV or variations in the intervals between heartbeats.
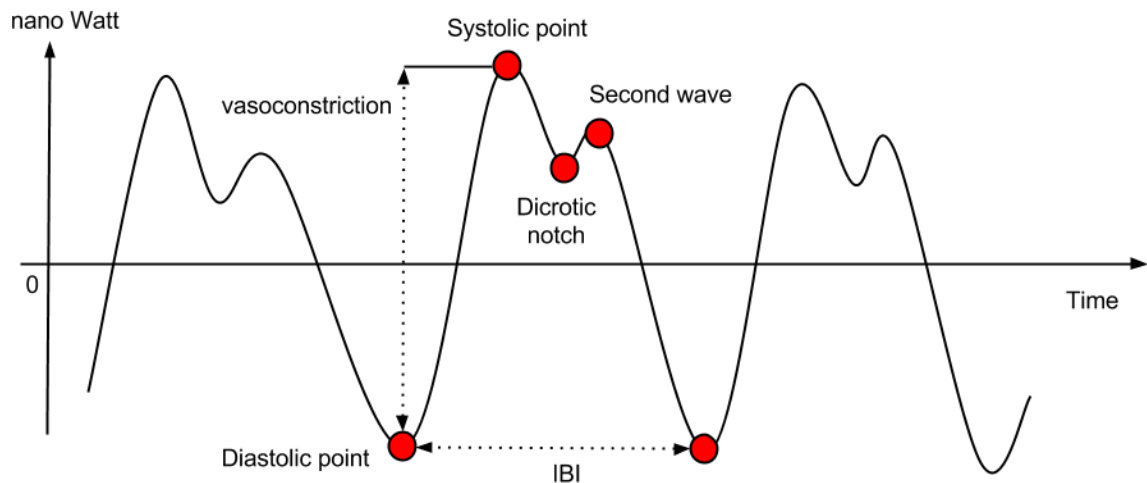
Figure 1.12: Photoplethysmographic Signal

1.4.3. Galvanic Skin Response

GSR, which is the signal shown in figure 1.13, refers to the variations in electrical conductance of the skin due to sweat secretion. This data is acquired by applying a low constant voltage to the skin and measuring the resulting variations in skin conductance. GSR is linked to the regulation of internal body temperature and has been found to be strongly associated with emotional arousal. The Skin Conductance Response (SCR), which is often used as an indicator, is modulated autonomously by the sympathetic nervous system, leading to changes in skin conductance that offer direct insights into emotional regulation. The time course of the GSR signal comprises two additive processes: a tonic base level driver that fluctuates slowly (seconds to minutes) and a faster-varying phasic component that fluctuates within seconds [25]. Changes in phasic activity can be identified in the continuous data stream as bursts characterized by a steep incline to a distinctive peak and a slow decline relative to the baseline level. When significant changes in GSR activity occur in response to a stimulus, an Event-Related Skin Conductance Response (ER-SCR) occurs; these responses are called GSR peaks and provide information about emotional arousal to stimuli.
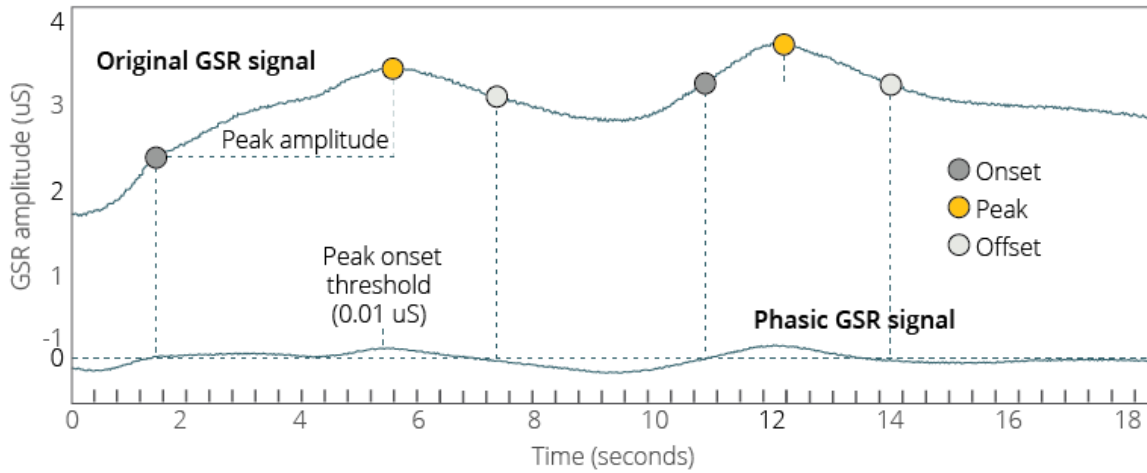
Figure 1.13: GSR Signal.

1.4.4.   Respiration

The respiratory system is primarily responsible for oxygen supply and carbon dioxide removal from the body's tissues. The process of breathing is regulated by muscular contractions that increase and decrease lung volume, facilitating precise and sensitive control of arterial blood carbon dioxide levels. In healthy individuals, this process occurs rhythmically, respiratory rate can be derived from the respiratory signal and as said in 1.1.1 it can be highly correlated to emotion [26, 27]. Respiratory pattern contains rich information about emotional states. Respiration velocity and depth usually varies with human emotion. For example, deep and fast breathing shows excitement that is accompanied by happy, angry or afraid emotion; shallow and fast breathing shows tension; relaxed people often have deep and slow breathing; shallow and slow breathing shows a calm or negative state [28].

This signal can be acquired by a belt with a displacement sensor placed around the subject's chest since the respiratory cycle is accompanied by changes in the thoracic volume. In the figure below a typical morphology of the respiratory signal is shown in figure 1.14. The respiratory activity also reflects on the heart rate variability in a physiological phenomenon known as Respiratory Sinus Arrhythmia, the high-frequency component of HRV is affected by respiration as the heart rate tends to increase during inspiration and decrease during expiration.
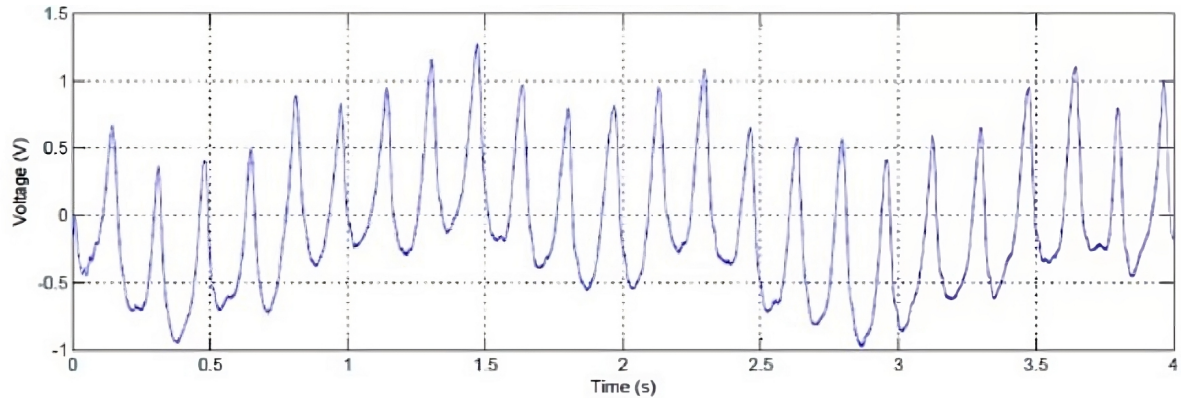
Figure 1.14: Respiratory Signal.

In conclusion, this section reported information about the research field this work is placed in, Affective Computing. After a brief introduction on the ANS and its spontaneous mechanisms when an emotion is felt and the main emotional models that have been presented and used in literature, some elicitation strategies were discussed. It is here that this work's aim finds its roots: such methods use one single type of stimulation, be it pictures, sounds or compound like videos, assuming that these contents have intrinsic emotive qualities that can be considered outside of the context in which they are perceived by the subject (which is, usually, a laboratory setting where these inputs are presented through a screen).

Immersive environments, powered by VR technology, present a new venture for Affective Computing with their ability to remove the subject from the aseptic real world laboratory environment and transporting them to a specifically designed setting. Enhancing the feeling of presence, the hypothesis is that VR as a medium has potential for new and more accurate ways to elicit specific emotional states in humans.

2 | Methods

2.1. Development of the Emotion Elicitation Protocol

As said before in 1.3, emotion elicitation is essential in the development of a system capable of detecting, interpreting, and adapting to human emotions. In this work an emotion elicitation protocol composed by two different elicitation techniques is proposed. The two methods are a traditional Flat Screen picture-based elicitation method (FS) and a fully immersive VR elicitation protocol. Both are designed to elicit sadness, relaxation, happiness, and fear, and were developed using the Unity graphic engine.

Several tools were used to build the protocol:

- **Data Acquisition Tools**: as said in 1.4 the ECG, BVP, GSR and Respiratory signal were collected for this protocol.

  The data acquisition method utilized in the study involved the use of the ProComp Infiniti (fig. 2.1), which is a portable multimedia system for acquiring biofeedback signals.

Figure 2.1: Procomp Infiniti Device

The system is designed to interface with any computer featuring a USB port and is equipped with 8 channels to enable parallel signal acquisition. The specific sensors

used in the study include a 3-lead ECG (fig. 2.2 a), a respiration band (fig. 2.2 c), a reflectance pulse oximeter (fig. 2.2 b), and an EDA sensor (fig. 2.2 d).

(a)

(b)

(c)

(d)

Figure 2.2: Procomp Infiniti Sensors:
(a) ECG sensors, (b) PPG sensor, (c) Respiratory Band, (d) GSR sensors.

- **Graphic Engine**:
  Unity (formerly known as Unity 3D) is a cross-platform game engine developed by Unity Technologies. It was first released in 2005 and has since been used to create videogames, VR experiences and mobile apps, both in 2D and in 3D [29]. Being free and easy to use, Unity has rapidly gained popularity among developers in a variety of industries, including education, engineering, architecture, medical, film

and advertising.

The engine, whose starting screen is shown in figure 2.3 (b), is written in the C# programming language and includes a comprehensive suite of tools for creating 3D content, including a terrain editor, animation tools and a particle system. A physics engine is also integrated, which can be used to create realistic physics-based simulations.

The engine natively supports the use of VR headsets, facilitating the development of immersive experiences.

(a)
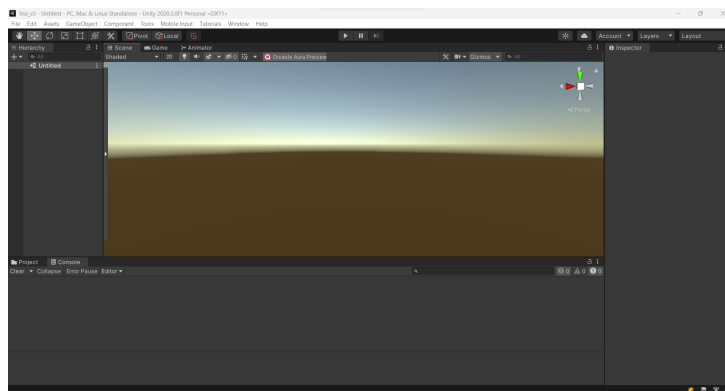
(b)

Figure 2.3: Unity Engine
(a) Unity logo, (b) Unity's environment

- **Questionnaires**: two different surveys were created to qualitatively assess, through self-report, the perceived affective state of the subjects after each section of the protocol. These surveys were created and distributed through the Microsoft Form platform.
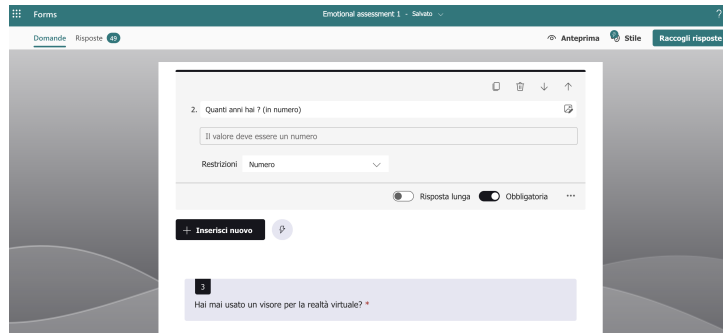
Figure 2.4: Microsoft Forms Environment

- **Data Analysis Tools**: The acquired data have been processed for statistical analysis using the MATLAB environment. For the Machine Learning part, Jupiter Notebook's environment was used.

2.1.1.   Experimental Protocol Structure

The developed emotion elicitation protocol is made of two different parts, each followed by an emotional assessment survey; for each acquisition a rigorous protocol was followed to minimize inter-subject variability in the response.

Each subject was sat on a chair at a standard distance of 80-100cm from a 24' monitor, sensors were applied to standard positionings on the hands (the non-dominant hand was always chosen to position the electrodes; the skin conductance's on the pointer and middle finger and the PPG sensor on the index's fingertip) as can be seen in figure 2.2 b) and d), torso (ECG sensors on the two shoulders and one on the abdomen above the navel) and abdomen (respiratory band).

A schematic representation of the experimental station is presented in figure 2.5.

Prior to each use, the electrodes were cleaned with alcohol, and the virtual reality headset was sterilized and cleaned.

The room temperature was maintained within a range of 20 to 25°C.

The protocol has a total duration of about one hour and 15 minutes, considering the set-up process, the two elicitation protocols and the time needed to answer the two post elicitation assessment forms.

Figure 2.5: Experimental setup.

Each of the two elicitation methods starts with a 90 seconds adaptation period, followed by eight parts: four sections of emotion elicitation, each portion representative of a different quadrant of the Circumplex plane (1.2), interspersed with four baselines. The order was designed to progress in a counterclockwise direction through the valence-arousal plane, commencing with sadness, followed by relaxation, happiness and ultimately, fear.

2.1.2. Flat Screen Elicitation

As stated before, the FS elicitation part is composed by 4 baselines (each lasting 1 and a half minutes), alternated with 4 stimulant sections (each lasting 2 and a half minutes). Each elicitation scene shows 5 pictures to the subjects, each picture is present on the

screen for 30 seconds, and they are always presented in the same order, chosen by rising arousal value.

These images were selected from the OASIS database, for its advantages over the IAPS database reported in 1.3.2, and were carefully picked in the most outlying regions of the CMA plane, so to induce highly distinguished reactions in the various scenes. The baselines were instead as stimulus-less as possible, in order to induce a neutral reaction in the subject and making them ready and unbiased for the next stimulant section.

Baseline

Since the entire FS part of the experiment is based on validated pictures, for the baseline moments, whose importance resides in their supposed ability to "reset" the emotional response of the subject, it was decided to proceed with the screen only reporting a uniform, neutral color. The color that was ultimately picked was the so-called Eigengrau or Eigenlicht (translated from German and Dutch: own light, brain gray or dark light). This particular color was first mentioned in the XIX century [30], and was related to the color perceived by human beings in total darkness with eyes closed.
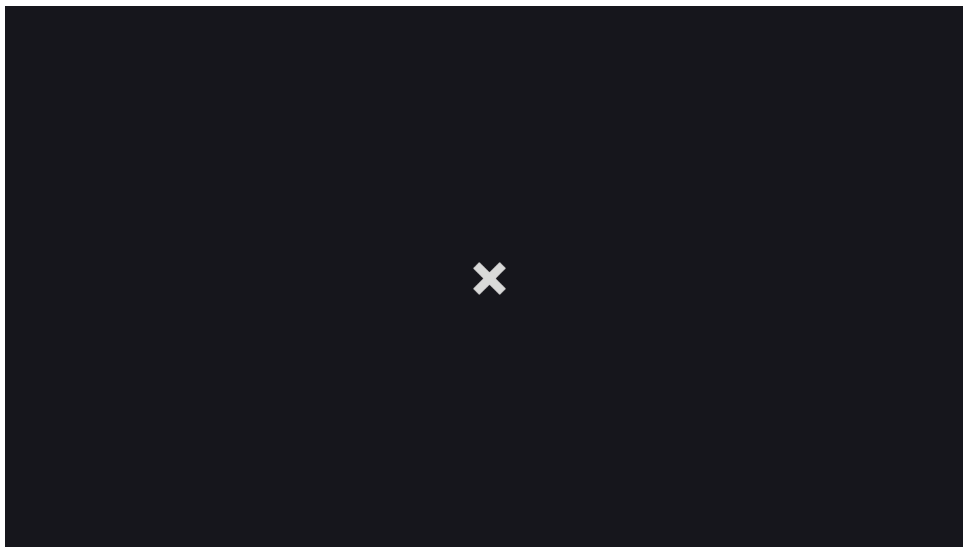
Figure 2.6: The baseline screen for the FS part.

The utility of this choice was twofold: giving a neutral stimulus, and at the same time assuring the subject that the screen was not switched off, that the experiment was still ongoing. Also, a little light-gray cross was added at the center of the screen, to remind the subject not to look around and get distracted. A representation of the baseline screen can be seen in figure 2.6.

Sadness

The first emotion inducing section is Sadness, this emotion is identified by low values of both valence and arousal. Using the OASIS database scores for each picture (for each dimension the scores went from 1 to 7), it was selected the portion of databases with scores of valence under 3,6 and arousal under 3,25 to isolate a significant group of pictures on the periphery of the plane.

From this group a selection of five pictures was chosen, shown in figure 2.7, the corresponding scores of valence and arousal are presented in the following table 2.1 in order of appearance during the experiment, so with rising arousal values; the selected pictures have a mean valence score of 2.509 and a mean arousal score of 3.081.

Figure 2.7: Sadness pictures selected

Sadness Pictures Scores

Picture	Valence	Arousal
Garbage Dump	2,196	2,861
Feces	2,137	3,069
Toilet	2,535	3,155
Fence	2,882	3,158
Miserable Pose	2,796	3,163

Table 2.1: Values of Valence-Arousal for the selected pictures: Sadness

Each picture is presented on the screen for 30 seconds, for a total duration of 2.30 minutes.

Relaxation

The second emotion inducing section is Relax, this emotion represents the fourth quadrant of the Circumplex plane, so it is identified by high values of valence and low arousal. Using the OASIS database scores for each picture (for each dimension the scores went from 1 to 7), the portion of databases with scores of valence over 4.5 and arousal under 2,25 was selected in order to isolate a significant group of pictures on the periphery of the plane.

Figure 2.8: Relax pictures selected

Relaxation Pictures Scores

Picture	Valence	Arousal
Pinecone	4,637	2,069
Bark	4,627	2,089
Grass	5,010	2,109
Rocks	4,520	2,178
Acorns	4,520	2,228

Table 2.2: Values of Valence-Arousal for the selected pictures: Relaxation

From this group a selection of five pictures was chosen, shown in figure 2.8, the corre-

sponding scores of valence and arousal are presented in the following table 2.2 in order of appearance during the experiment; the selected pictures have a mean valence score of 4,663 and a mean arousal score of 2,135.

Each picture is presented on the screen for 30 seconds, for a total duration of 2.30 minutes.

Happiness

The third emotion inducing section is Happiness, this emotion represents the first quadrant of the CMA plane, so it is identified by high values of valence and arousal. Using the OASIS database scores for each picture (for each dimension the scores went from 1 to 7), the portion of databases with scores of valence over 5 and arousal over 4.5 was selected in order to isolate a significant group of pictures on the periphery of the plane.
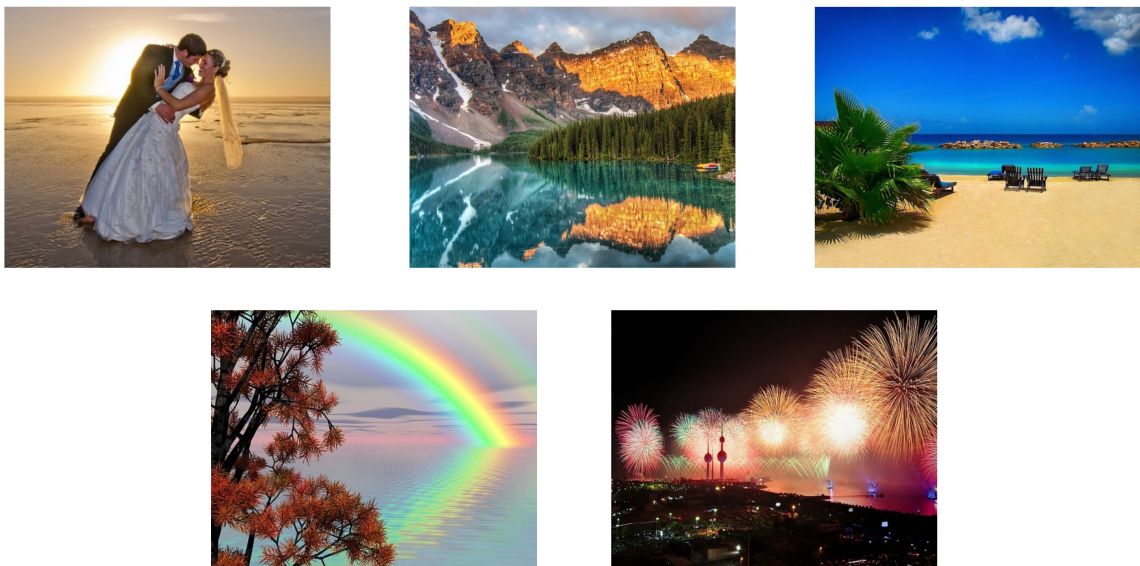
Figure 2.9: Happiness pictures selected

From this group a selection of five pictures was chosen, shown in figure 2.9, the corresponding scores of valence and arousal are presented in the following table 2.3 in order of appearance during the experiment, so for rising arousal; the selected pictures have a mean valence score of 6,249 and a mean arousal score of 4,777.

Each picture is presented on the screen for 30 seconds, for a total duration of 2.30 minutes.

Happiness Pictures Scores

Picture	Valence	Arousal
Couple	5,971	4,535
Lake	6,376	4,728
Beach	6,366	4,738
Rainbow	6,257	4,903
Fireworks	6,275	4,980

Table 2.3: Values of Valence-Arousal for the selected pictures: Happiness

Fear

The fourth emotion inducing section is Fear, this emotion represents the second quadrant of the CMA plane, so it is identified by high values of arousal and low values of valence. Using the OASIS database scores for each picture (for each dimension the scores went from 1 to 7), the portion of databases with scores of valence under 2.5 and arousal over 4.5 was selected in order to isolate a significant group of pictures on the periphery of the plane.

Figure 2.10: Fear pictures selected

From this group a selection of five pictures was chosen, shown in figure 2.10, the corresponding scores of valence and arousal are presented in the following table 2.4 in order of

appearance in the experiment, so with rising arousal values; the selected pictures have a mean valence score of 1,902 and a mean arousal score of 4,851.

Fear Pictures Scores

Picture	Valence	Arousal
Injured Foot	1,971	4,535
Fire	1,743	4,602
Car Crash	2,029	4,663
Explosion	2,376	4,990
Injury	1,392	5,465

Table 2.4: Values of Valence-Arousal for the selected pictures: Fear

Each picture is presented on the screen for 30 seconds, for a total duration of 2.30 minutes.

To better visualize the placement of each section on the Valence-Arousal plane a plot with the mean values of the two dimensions is shown in figure 2.11.
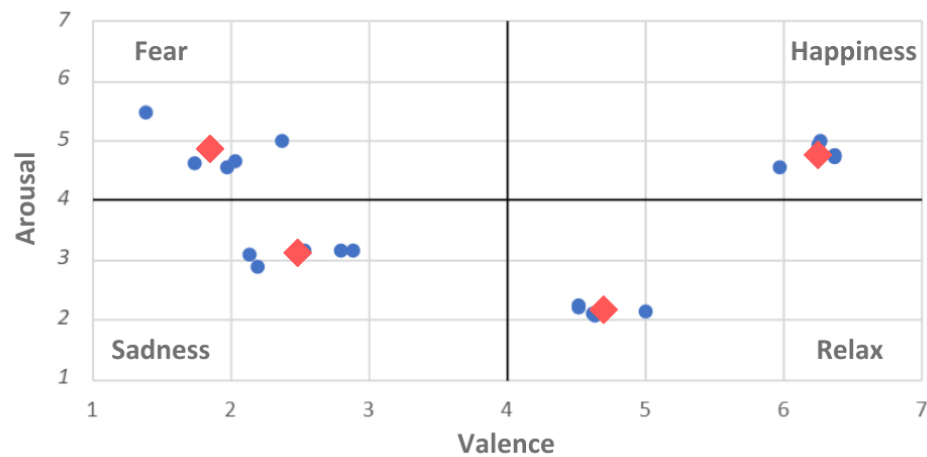
Figure 2.11: Section Placement on Valence-Arousal Plane.

2.1.3. Virtual Reality Elicitation

Similarly of the FS Elicitation part, also the Virtual Reality one is composed of 4 baselines and 4 stimulant parts. Each of these scenes is a VR 360° environment in which the subject is immersed.

		High Arousal	**Low Arousal**	**High valence**	**Low valence**
Static Visual Cues	Shape	Angular	Rounded	Rounded	Angular
	Lines			Rounded	Straight
	Hue	Green-Yellow, blue-green and green	Purple-blue and yellow-red	Blue,green and purple	Yellow
	Saturation	N/A	N/A	Saturated	Unsaturated
	Brightness	N/A	N/A	Bright	Dark
	Visual Complexity	N/A	N/A	Extremely complex or extremely simple	Neither
Non-static Visual cues	Speed	Fast	Slow		
	Motion shape	Spherical	linear	N/A	N/A
	Direction	N/A	N/A	Downwards-right, inward	Upwards-left, outward
	Path curvature	Jerky	Straight	Jerky	Straight
Sound Cues	Source location	Outside field view	Inside field view	N/A	N/A
	Distance from sound source	Near	Far	N/A	N/A
	Sound Source Movements	Approaching	receding	N/A	N/A
	Heartbeat Sound	Fast(over 100bpm)	Slow(under 60bpm)	N/A	N/A
Music	Tempo	Fast	Slow	N/A	N/A
	Harmony	N/A	N/A	Major scale	Minor Scale

Table 2.5: Summary of audio-visual cues associated with affective states.

At all times, the subject is free to look around in any direction, but is unable to physically move and the only device tracked is the headset, as no hand controller was given to the participants. Instructions were given to the participants so that, although they moved their head during the acquisition, such movements always had to be slow and possibly not sudden, in order to reduce motion artifacts in the signals.

The elicitation scenes are based on the OASIS images that were used in the Flat Screen part of the protocol, while the baselines were emotionally neutral environments. In the design process, careful attention was put into using the suggestions derived from the review [31], a summary of them is reported in table 2.5.

Finally, in each stimulating scene certain events were inserted, and they were always activated at the same time and in the same order for all subjects. Such events and their timeline is reported in each of the following sections. Also, while some events were repeated more than once across the scenes, some others were one-shot only.

In all scenes, both assets created internally using the Blender environment and taken from internet free sources were used.

Baseline

The baseline scenes are two neutral environments, previously used and created for [32]: the first is the interior of a modern apartment, while the second is a professional office in a high building. Although fully furnished, the two places are devoid of stimulating elements and only present background and constant sound stimulation. The modern apartment scene is used for the first and third baseline, while the office is used for the second and the fourth (so that they are alternated). Every time, the position of the player inside the environment is different: the perceived novelty was in charge of avoiding the insurgence of boredom in the subject.

(a)

(b)

Figure 2.12: The apartment environment; (a) represents the point of view of the subject in the first baseline of the VR elicitation part, while (b) represents the point of view in the third baseline

(a)

(b)

Figure 2.13: The office environment; (a) represents the point of view of the subject in the second baseline, while (b) represents the point of view in the fourth baseline

Sadness

Getting inspiration from the 5 images chosen from the OASIS dataset for the FS sadness part, it was decided to ambient this scene in a garbage dump (2.14) placed in the periphery of a big city. Fenced with high barriers covered in barber wire, the dump is full of trash clusters and barrels with fires light up in them. Dogs roam the scene, looking for food, and there is an abandoned and rusty toilet lying in a corner of the environment. To complete the sad/depressing atmosphere, a homeless man sits, remaining completely still, against a container, and it is raining but with low intensity. Textures and models are mostly in low resolution, colors have low saturation, movement is rare and the scene is overall pretty dark, all in accordance with [31].

Figure 2.14: Subject's point of view in the VR Sadness scene

(a)

(b)

(c)

(d)

Figure 2.15: Some details from the VR Sadness scene: (a) A dog licking from an abandoned shoe; (b) particular of the barbed wired fence that surrounds the damp; (c) the toilet left in a container; (d) the homeless man.

The events that were selected for this scene are reported in table 2.16 and comprehend light-poles' flickering, dogs barking, trash rolling and a glass bottle falling nearby the

location of the homeless man.

Sadness Scene Events							
0s	60s	120s	150s	180s		240s	300s
Background ambient noise							
Light Flickering							
Dogs Barking							
Trash Falling							
Bottle Falling							
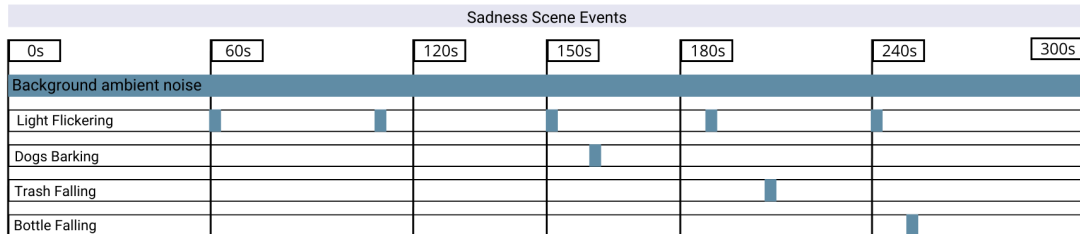
Figure 2.16: Sadness Scene Events

Relax

As all the selected OASIS pictures for the Flat Screen Relax part were related to a natural environment, the scene for the VR Relax stimulation was a pristine wood environment in a sunny day (2.17. A little creek flows right in front of the subject, the surroundings are made of trees and flowers and a soft wind makes the leaves and the grass move. Stones and acorns are also spread all around. It is to be noted that the resulting environment is strikingly similar to those used for video and/or VR meditation apps [21], as natural scenes (be them physical or virtual) show enhanced ability in inducing or facilitating relaxation [33].

Figure 2.17: Subject's point of view in the VR Relax scene

The movements of the leaves and branches are rhythmic with low frequency, the colors are bright and the sun rays are clearly visible through the branches. All sounds are natural, to be expected, and surround the subject enhancing immersion.

Figure 2.18: Some details from the VR Relax scene: (a) a little water stream; (b) flowers spread all around; (c) the sun shining from above; (d) one of the bunnies.

The events that were selected for this scene are reported in table 2.19 and comprehend birds chirping and two bunnies that appear in different positions of the scene and then start wandering around.

Figure 2.19: Relaxation Scene Events

Happiness

For this elicitation section, it was decided to proceed by designing a party setting, at night time, nearby a lake. The subject is placed on a bench right in front of a bonfire (2.20). Five young people can be seen all around and they are either eating, talking or standing in front of the lake. Music, played inside the house, can be heard from the player. The atmosphere is one of inclusion but also freedom, as everyone shows to be enjoying the situation on their own terms. Later on, highly saturated fireworks are fired in the distance, with distinct loud but easily recognizable sounds that also help in their localization. Colors and lights are bright, dynamic, and highly saturated, and they are even more noticeable in contrast with the dark night sky, as in [34] it is stated that these features, and not hue in itself, are mainly related to a positive valence of the stimulus.

Figure 2.20: Subject's point of view in the VR Happiness scene

As for the events of this elicitation part (refer to table 2.22), cicadas sing at fixed intervals. Cheering from inside the house, where people are implicitly dancing and singing, can be heard from outside. Near the end of the scene, fireworks are shot on the other shore of the lake, at the right side of the subject.

(a)

(b)

(c)

(d)
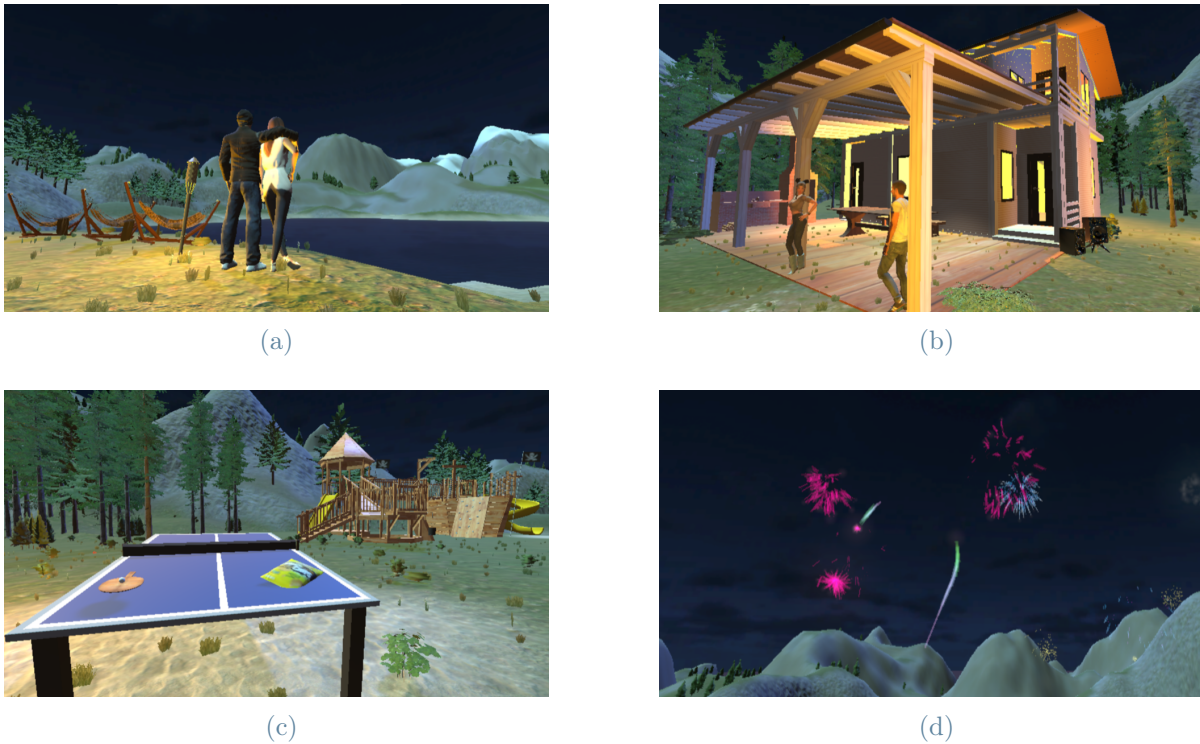
Figure 2.21: Some details from the VR Happiness scene: (a) A couple looking at the lake; (b) the house where the party is going on; (c) ping pong table and a play castle, near the house; (d) the fireworks that appear near the end of the stimulation, shining over the lake.

Figure 2.22: Happiness Scene Events

Fear

The Fear elicitation scene was the most difficult to create from the OASIS pictures previously chosen. It was in the end decided to recreate a car crash, happening in real time nearby the subject (fig. 2.23), in an indifferent city, devoid of people and with only cars and trucks wandering. During the 5 minutes of stimulation a car crash happens in real time in the field of view of the subject.

Figure 2.23: Subject's point of view in the VR Fear scene

(a)

(b)

(c)

(d)

Figure 2.24: Some details from the VR Fear scene: (a) traffic goes on in the streets; (b) a snapshot of the impact; (c) hazard lights blinking during the alarm, right after the impact; (d) the flames coming from the hood of the pick-up.

This scene's elicitating content is mostly made of the events, that induce increasing concern with time. A white pick-up truck appears from a corner, moving at high speed; it takes a turn but it is going too fast, and this leads it to crash into a parked car, at ~10 perceived meters from the subject. As the alarm of the car that was hit goes on, flames appear from the hood of the pick-up truck, and nobody comes out of it. After a while, the flames' intensity increases, until (refer to fig. 2.25 for the timing of the events), the pick-up explodes, leaving only the background city sounds and silence otherwise. In the last 30 seconds of stimulation, an ambulance can be heard getting closer and closer, but it will never actually show up in the subject's visual field. This last design choice was based on [35], where it is stated that when the sound source is not visible to the subject its arousing effect is increased, and also that the sound of something getting near is more arousing than one going away.
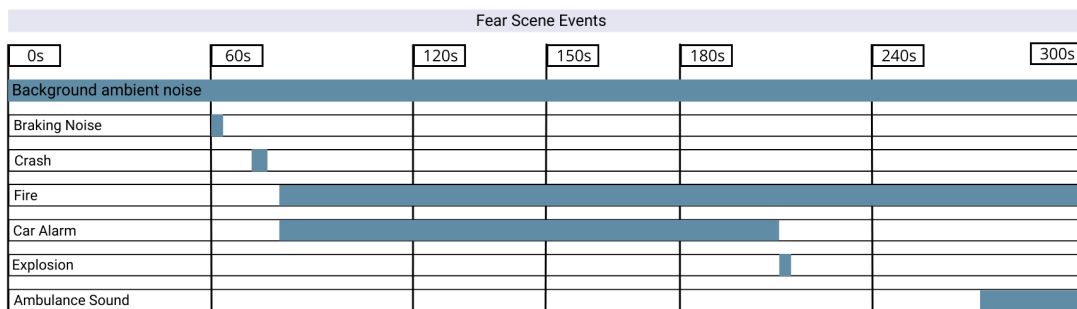
Figure 2.25: Fear Scene Events
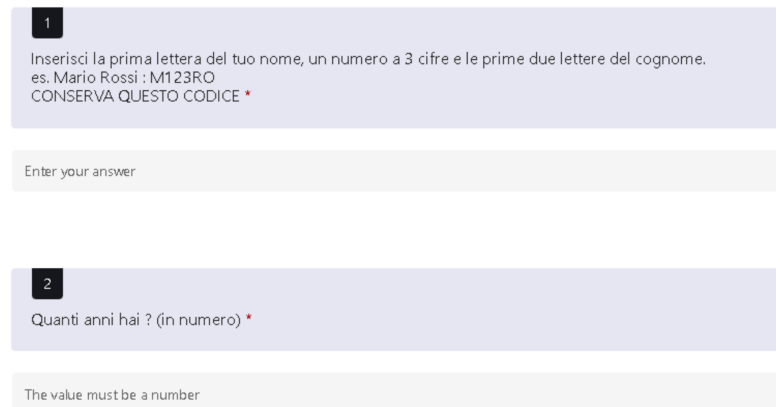
2.1.4. Subjective Emotional Assessment

In this study a post elicitation emotional assessment survey was administered through the Microsoft Forms platform after each portion of affective elicitation. The two forms were designed to assess qualitatively the coherence of the resulting stimuli with the intended elicitation, and also to collect additional information about the test subjects. To achieve these results, discrete models cited in 1.2 such as SAM and PAM were used to perform a self report on the emotional state.

To assess values of valence and arousal, the graphical representations of SAM that are shown in figure 1.5 and 1.6 were used: the subject had to report their state in a scale from 1 to 5 where, for valence, a score of 1 meant unhappy and 5 happy, while for arousal a score of 1 meant excited and 5 calm (this score was then flipped to better be represented on a valence-arousal plane).

Both surveys were divided in different sections:

- In the first section the test subject was asked to create a unique identifier code and type in some useful information, such as their age and if they had previous experience with VR; an estimation on the duration of the elicitation part they just experienced was also surveyed.

Figure 2.26: First section of the form.

- The following sections referred to the 4 elicitating parts (sadness, relaxation, happiness and fear). Each subject had to rate their experience based on the emotion they felt during it, using SAM for self report on valence and arousal and PAM for the main mood they felt (fig. 2.28). For each section a representative picture of the scene was added above the questions that referred to it, an example can be seen in figure 2.27.

Figure 2.27: Emotional assessment introduction to the scene.

Figure 2.28: Emotional assessment example of SAM and PAM.

- An additional part was added in the post-VR survey, to further investigate the effect of the four baselines, using the same tools as the ones used for the scenes.

2.2. Signal Acquisition and Processing

The designed protocol was applied without changes to each participant. Signal acquisition was conducted with strict consistency and stability to ensure optimal quality. In the following sections, the different signals that were recorded and their specific processing methods will be illustrated. Each acquisition session involved the recording of GSR, Respiration, ECG, and PPG, as depicted in the figure 2.29.
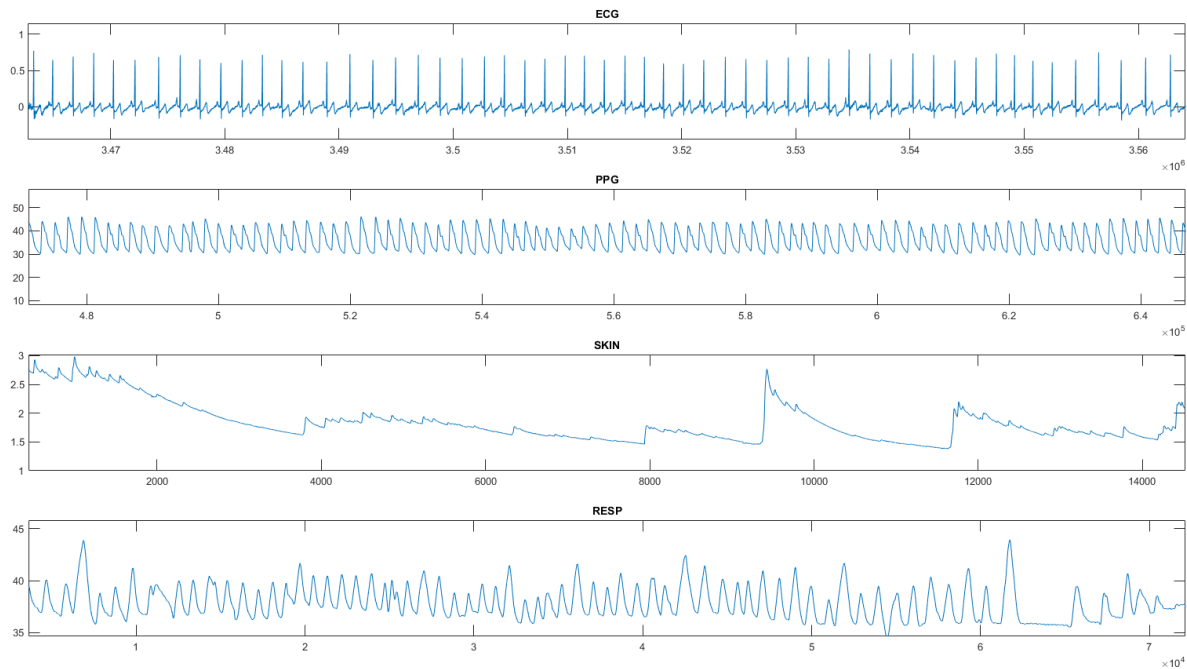
Figure 2.29: The four signals that were sampled for each subject.

As mentioned before in 2.1 the ProComp Infiniti system was used to acquire the signals, the sampling frequency was 2048 Hz for ECG and PPG and 256 Hz for GSR and Respiration. Data and metadata were then stored in a txt file with a time vector.

Signal processing is a crucial step in enhancing the quality of signals by mitigating noise and optimizing spectral components. The acquired signals include GSR, Respiration, ECG, and PPG, as previously mentioned. Starting from the raw signals, additional time series were obtained. HRV was extracted from ECG, and a point process model was created. From ECG and PPG, Pulse Arrival Time (PAT) was extracted. PAT was calculated as the difference between R peaks in ECG and systolic peaks in PPG. The last time series was Pulse Pressure, which was computed using the photoplethysmographic signal as the difference between Systolic pressure and Diastolic pressure each beat.

2.2.1. Galvanic Skin Response Signal

In order to extract the features of GSR (Galvanic Skin Response), the signal was used in both complete (Tonic + Phasic) form and Phasic form as in [36]. A lowpass filter with a cutoff frequency of 2 Hz was used to filter the signal, which was then downsampled to 8 Hz. The phasic component was extracted from this time series using a median filter. A second version of the signal was obtained by applying a bandpass filter from 0.5 Hz to 1 Hz in addition to the lowpass filtering.
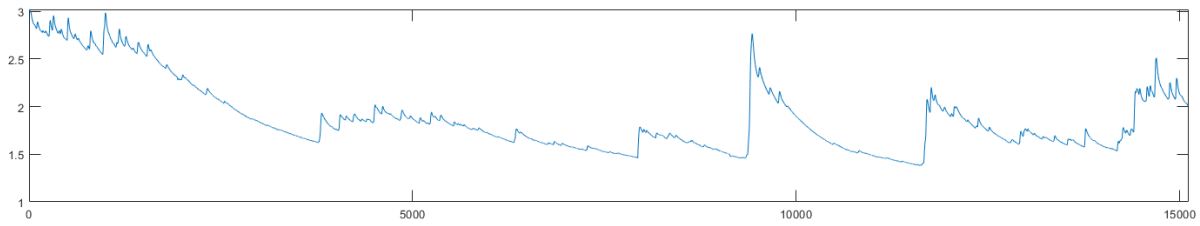
(a)
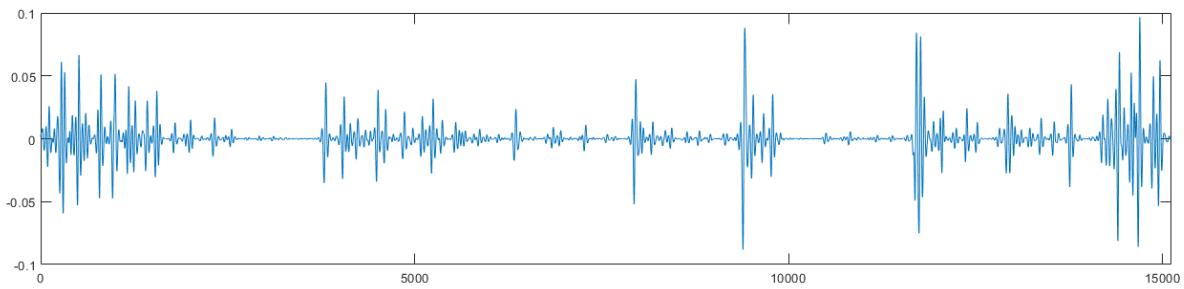
(b)
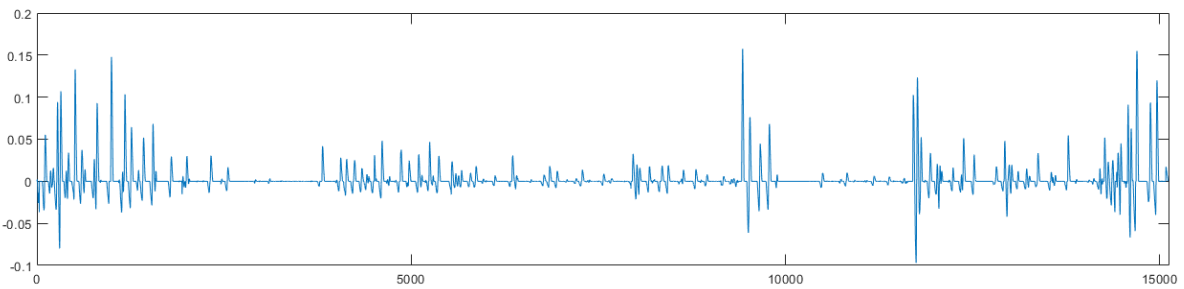
(c)

Figure 2.30: Time series obtained from GSR : (a) GSR processed with low pass filter; (b) GSR processed with a bandpass filter; (c) phasic component of GSR signal

By combining the information extracted from the phasic component, it was possible to identify the onset and offset in the low pass filtered signal. The peak was identified as the maximum value in the two-point range. This allowed the computation of the amplitude of the peaks as the difference between the onset and the actual peak value. Peaks with a difference between onset and offset less than 1 sample were excluded from the analysis. The features extracted from the GSR signal are shown in the table 2.6.

Feature	Description
Avg amplitude peaks	The average amplitude of the peaks is calculated by determining the height of the individual peaks in the GSR signal and then taking the average of those values.
Avg rise time	The "average rise time" refers to the average time expressed in seconds between the peak and the onset and provides information about the speed at which the GSR signal changes in response to a stimulus.
Avg recovery time	The Average recovery time is computed as the average time in seconds between the found peaks and the offset.
N peaks	Computed as the number of peaks found in the considered time window.
Sd amp peaks	Computed as the standard deviation of the amplitudes of the found peaks.
Slope gsrbp	It is the slope of the band-pass (0.5 -1 Hz) filtered GSR signal.
Max sign amp gsrbp	It is the maximum amplitude value of the band-pass filtered GSR signal.
Avg der gsrbp	It is the average of the first derivative of the band pass filtered GSR signal.
Sd der gsrbp	It is the standard deviation of the first derivative of the band pass filtered GSR signal.
Max der gsrbp	It is the maximum value of the first derivative of the band pass filtered GSR signal.
Average	It is the mean of the low pass filtered signal taking into account both tonic and phasic components.
Sd	It is the standard deviation of the low pass filtered signal taking into account both tonic and phasic components.
Avg abs1	It is the mean of low pass filtered GSR first derivative under absolute value taking into account both tonic and phasic components.
Avg abs1 norm	It is the mean of low pass filtered GSR first derivative under absolute value taking into account both tonic and phasic components normalized by the standard deviation of the signal.
Avg abs2	It is the mean of low pass filtered GSR second derivative under absolute value taking into account both tonic and phasic components.
Avg abs2 norm	It is the mean of low pass filtered GSR second derivative under absolute value taking into account both tonic and phasic components normalized by the standard deviation of the signal.
Env	It is the mean of the envelope of the low pass filtered GSR signal.

Table 2.6: GSR extracted features.

2.2.2.   Respiration

With the aim of generating time series containing the respiration peaks and inter-respiration intervals in the form of frequencies, the raw respiration signal is passed through a band-pass filter with a pass frequency of 1 Hz and a stop frequency of 1.5 Hz. Next, peak detection is carried out using a minimum peak prominence of 0.3 times the standard deviation of the signal, and a minimum peak distance of 2 times the sampling frequency. The identified peak times are stored and differentiated to obtain the second time series. This series is then converted to frequencies by taking the inverse of the time intervals, and can be seen in figure 2.31.
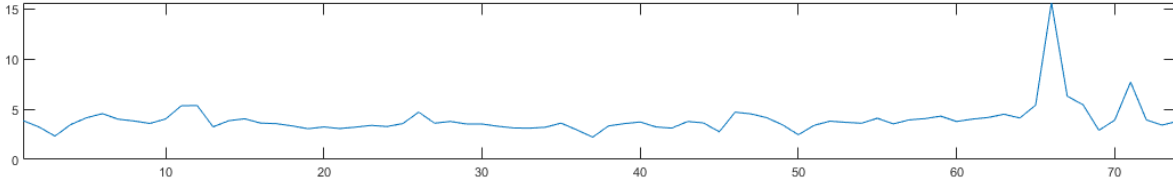
Figure 2.31: Respiratory time series.

The features extracted from the Respiratory signal are shown in table 2.7.

Feature	Description
Respiratory frequency	It is the respiratory frequency expressed in breaths per minute.
Resp amplitude	it is calculated as the mean of the amplitudes of the breathing cycles in the considered time interval.

Table 2.7: Respiration extracted features.

2.2.3.   Cardiac

The raw ECG signal is processed using a Matlab-based tool called SyncTool, which was previously developed at the Spinlabs of Politecnico di Milano and can be seen in figure 2.32. This tool is a semi-automated annotator that identifies the R peaks in the ECG and the systolic, diastolic, and onset of the PPG waves. It has been used in numerous research studies [37–39].
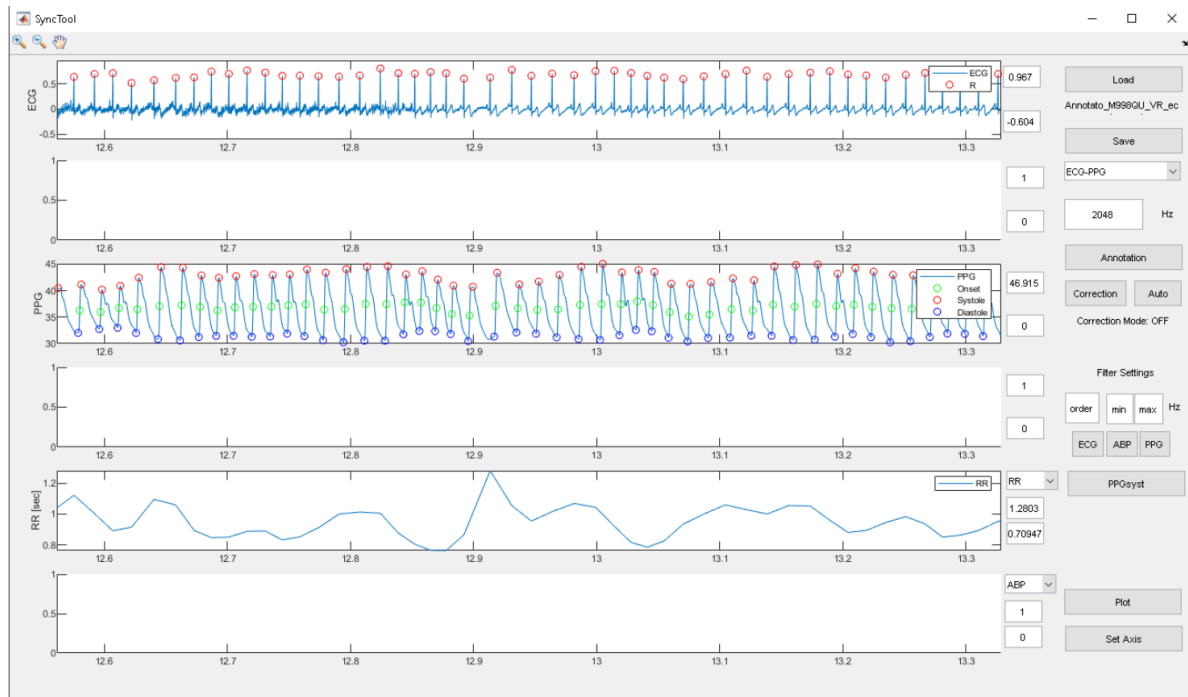
Figure 2.32: Example of signal annotation via SyncTool.

The annotations obtained from SyncTool (that uses the Pan-Tompkins method to recognize R peaks) are essential in defining the HRV time series and instantiating the point process model. However, the annotations obtained using SyncTool were not always suitable for the feature extraction phase due to the presence of anomalous drifts in the signals. Specifically, the PPG and the ECG tend to lose their typical morphology when for example the subject inattentively moves their hand, so the automated annotation with the tool does not perform well. The tool is also equipped with a correction mode to manually erase and replace the incorrect annotations. If the morphology of the signal does not allow any form of annotation, as it can be seen in figure 2.33, a window of at least 3 seconds was left without annotations, to allow for recognition and interpolation in further processing of the signal. Signals that needed to be interpolated for more than 10% of their length were discarded from the dataset.
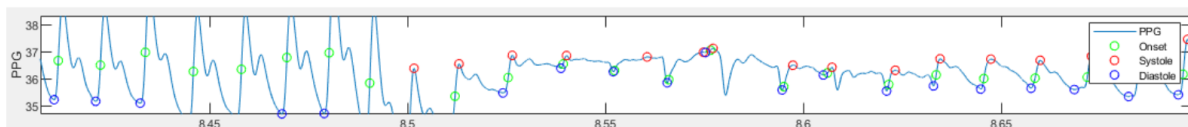
Figure 2.33: Example of incorrect annotation.

Point Process Model

A Point Process is defined as a stochastic process composed of a time series of binary events that occur in continuous time [40]. They are used to describe data that are localized at a finite set of time points. As opposed to continuous-valued processes, which can take on any of countless values at each point in time, a point process can take on only one of two possible values, indicating whether or not an event occurs at that time. In a sense, this makes the probability models used to describe point process data relatively easy to be expressed mathematically, as shown in [41]. In figure 2.34 it is shown an example of an estimated tacogram, obtained through the use of Point Process.
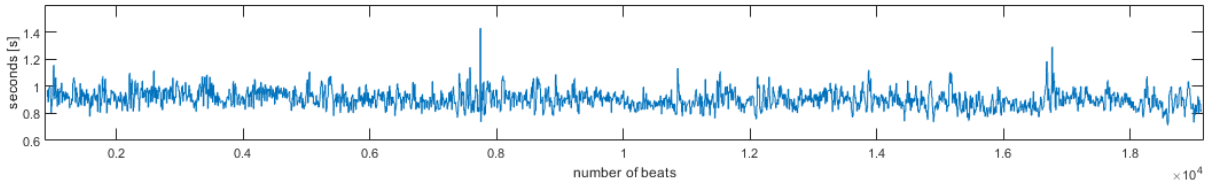
Figure 2.34: Point Process Time series.

The features extracted from the Point Process are shown in table 2.8.

Feature	Description
Mu	It is the average of the estimate RR distance.
S2	Computed as the variance of the modeled RR distances.
lf	It is the estimate of the low frequency power.
hf	It is the estimate high frequency range power.
bal	It is the ratio between lf and hf.
vlf	It is the estimate of very-low frequency range power.
tot	It is the estimate of the total PSD.
lfn	It is the estimate of the low frequency range power normalized.
hfn	It is the estimate of the high frequency range power normalized.

Table 2.8: Point Process extracted features.

PAT

As previously stated in 2.2, PAT is defined as the temporal difference between the occurrence of the R peak on the ECG and the corresponding systolic event in the PPG. More specifically, PAT can be calculated as the difference between the R peak and one of three

fiducial points (systolic, diastolic, or onset) in the PPG pressure wave; the onset-derived PAT was ultimately selected as it proved to be the most consistent and reliable metric.

An example of PAT time series is presented in figure 2.35.
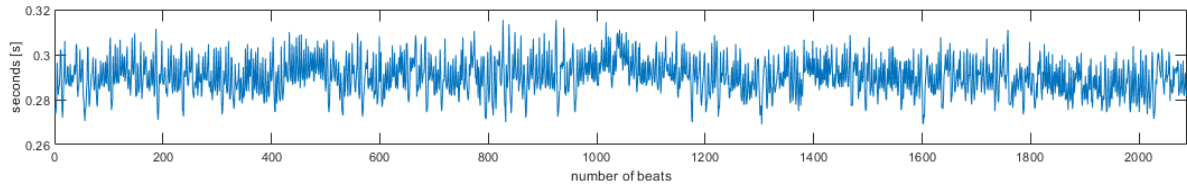
Figure 2.35: Pulse Arrival Time time series.

The features extracted from the PAT are shown in table 2.9.

Feature	Description
PAT	It is the normalized onset PAT computed as the difference between each PPG onset and its corresponding R peak.

Table 2.9: PAT extracted features.

Pulse Pressure
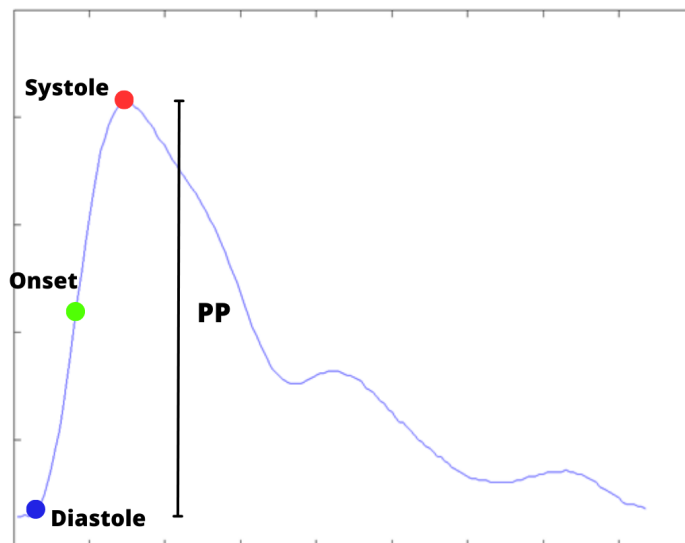
Figure 2.36: PPG annotations.

PP is a time series derived from the blood volume information obtained from PPG. Specifically, PP was computed as the difference between systolic and diastolic blood pressure, as illustrated in figure 2.36.
The obtained signal is reported in figure 2.37.
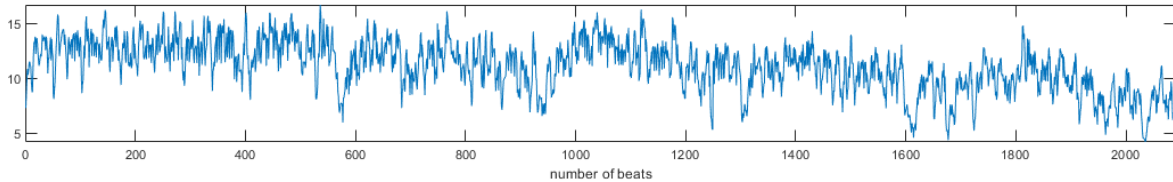
Figure 2.37: PP time series.

The features extracted from the PP are shown in table 2.10.

Feature	Description
Average PP	It is the average of the Pulse Pressure.

Table 2.10: PP extracted features.

2.3.   Statistical Analysis

2.3.1.   Questionnaire Assessment

Once the questionnaire collection is complete in the experimental phase, some generic statistical information about the subjects (such as medium age and previous experiences with VR) were observed and analysed.
A more in-depth study was then carried out for the SAM and PAM results in the various scenes for both parts of the protocol.

SAM Analysis

In both questionnaires, as reported in 2.1.4, the subjects were asked to report their perception of valence and arousal of the various elicitating parts, using the SAM maniquin (as reported in figures 1.5 and 1.6). These values were collected and their mean values and standard deviation in the various parts were computed and then used to plot the global subject's perception on the Circumplex plane.
As for the FS part, the resulting plot was compared with the mean of the validated values of the OASIS pictures used in the elicitation. This analysis is useful to understand

whether the emotional stimulation was not only effective but also accurate.

This kind of comparison was not performed for the VR PAM results: although the scenes were inspired by the same OASIS pictures used in the FS part, more stimuli and an overall higher variability were introduced. This lead to the use of such data mainly to understand the efficacy of the VR scenes in the sense of centrality of the valence and arousal values with respect to the targeted emotional quadrant.

PAM Analysis

The moods of the subjects referring to each elicitation part was inspected through their answers to the PAM graphical assessment model (2.1.4).

The answers were collected and visualized through a histogram for each elicitation section, to have additional qualitative information on the elicitation results. Also, as moods are not directly related to the OASIS values, but are a way more complex phenomenon, a comparison with the scores of the OASIS pictured used was not possible. On the other hand, a comparison between the FS and the VR PAM answers has been be performed in order to check the coherence of the elicitated emotions with the intended ones.

2.3.2.   Extracted Features Analysis

Once the selected 30 features were extracted from all signals, they were organized in a dataset to make the statistical analysis easier.

The dataset was composed by two tables, one for the feature extracted from the signals during the Flat Screen elicitation and a structurally identical one for the VR elicitation's features.

Each table has 30 columns for the extracted features, and one row for each subject, in each cell was contained a vector of 8 elements, one for each elicitating section, in order baseline 1, Sadness, baseline 2, Relaxation, baseline 3, Happiness, baseline 4 and Fear.

At first the median values of each feature were plotted for all 8 parts, to better visualize the trends within each elicitation method and to qualitatively compare the two. An example of the obtained plot is shown in figure 2.38.
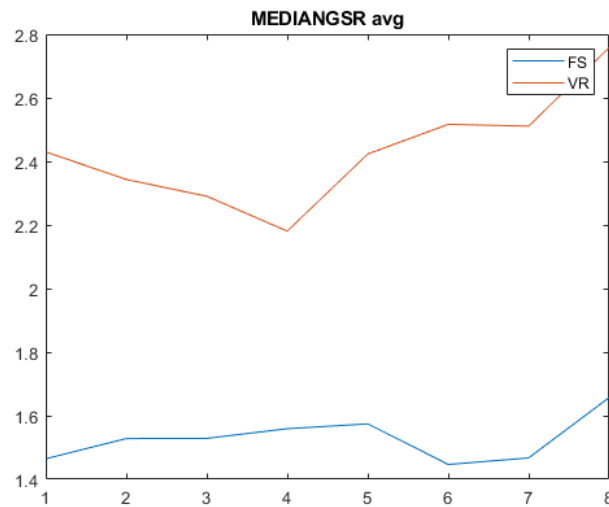
Figure 2.38: Median trends for Flat Screen and VR elicitation of the feature avg from GSR.

With these plots the neutralizing effect of the baselines is also surveyed. To move forward and perform statistical testing a dynamic depolarization based on the baseline before the scene was performed, with the aim to identify the actual variation from a neutral state to the response to the emotional stimuli of each portion of the protocol and better compare the two elicitation methods.

Once the depolarized features were computed some statistical test were performed.
The first chosen approach was a Friedman's test, on each elicitating method, to assess the differences between the 4 elicitation parts. The Friedman's test is a non-parametric statistical test used to determine whether there are differences among two or more related groups. It is similar in nature to the one-way ANOVA with repeated measures, but it is more suitable for analyzing data that are not normally distributed.

The null hypothesis of the Friedman's test is that there is no difference among the groups. If the test statistic is significant at a chosen level of significance (e.g., $\alpha = 0.05$), the null hypothesis is rejected in favor of the alternative hypothesis that at least one group is different from the others.
One of the advantages of the Friedman's test is that it does not assume that the data are normally distributed, and it is more robust to outliers than parametric tests. However, it does not provide information about which group is different from the others, so a post-hoc test with the Bonferroni correction was performed to determine which groups significantly differ from each other.
The Bonferroni correction is a statistical adjustment method that is often used to con-

trol for multiple comparisons in hypothesis testing; it involves dividing the desired level of alpha by the number of comparisons being made, this adjusted significance level is then used to determine whether the null hypothesis can be rejected for each individual comparison.

After the Friedman's test a second statistical test was performed to assess the differences between the two elicitation methods. A Wilcoxon signed-rank test was performed between each pair of corresponding section of the two elicitations modalities.

The Wilcoxon signed-rank test, also known as the Wilcoxon paired test, is a non-parametric statistical hypothesis test used to determine whether two related samples have equal medians. It is often used as an alternative to the paired t-test when the normality assumption is violated.

The Wilcoxon signed-rank test works by comparing the differences between the paired observations, ranking these differences from smallest to largest in absolute value, and then summing the ranks of the positive and negative differences separately.

The null hypothesis of the Wilcoxon signed-rank test is that there is no difference between the medians of the two related samples. If the test statistic is significant at a chosen level of significance (e.g., $\alpha = 0.05$), the null hypothesis is rejected in favor of the alternative hypothesis that the medians are different.

2.3.3. Feature Selection

Further analysis was performed to assess what were the best features capable of distinguish the four emotional states for each elicitation method.

A Square Method feature selection approach was selected. This feature selection procedure is based on successive binary comparisons. For each possible pair of features the mean values, the standard errors (calculated as the standard deviation divided by the square root of the number of samples) and 95% confidence intervals (calculated as 1.96 multiplied by the standard error) were computed and then plotted on the plane defined by the two compared features as squares, where the average was highlighted at the center of the square and the sides were computed as the average $\pm 1.96 * standard\ error$ [42].
To create a ranking of importance of the available features a metric was assigned to each feature, the metric is determined by finding the area of the intersection between two distributions and normalize it dividing it by the total area of the distribution being evaluated.

Afterwards, the amount of times each feature appears in valid couples is counted and a weight proportional to the magnitude of the occurrences (with higher values indicating

better performance) is assigned to each feature. Subsequently, the correlation matrix is calculated and a second weight is assigned to the analyzed feature by counting how many times the feature has a correlation higher than 80% with another (with lower values indicating better performance). In cases where the features have the same value, the second weight is taken into account and the one with the highest value is placed second on the ranking. The outcome of this process is a list of uncorrelated features combined with the most pertinent correlated features.

To isolate only the three best features per method a distributions' inter- section threshold has been set for each elicitation method.

This representation was also implemented to visualize how the selected features could distinguish high valence states from low valence ones and high arousal states from low arousal ones.

The best 3 features for both elicitation methods are then used for a three-dimensional visualization of the emotion separation they induce. Such graphical representation will be useful in determining, at a glance, the efficacy of the stimulation in generating different emotional responses.

2.4. Machine Learning Classification Model

Although the subject count, as initially designed, was going to be too low to have a high accuracy in classification through a machine learning approach, it was still attempted to train various models to try and discriminate between: all 4 emotional stimulations (multiclass classification); high/low arousal; high/low valence (binary classification).

A new dataset were created, containing the features cited in 2.2 but only for the VR data. Such dataset was obtained starting from the one described in 2.3.2: each row (that refer to a single subject) was unpacked into 4 rows, one for each elicitating part, bringing the total row count to 176. The dataset has a column for each extracted feature and a final one with the target value, a number from 1 to 4, were 1 identifies Sadness, 2 identifies Relaxation, 3 identifies Happiness and 4 identifies Fear, used to train a multiclass classification model. This target column can be easily modified to accommodate a classification model that aims to distinguish only a different arousal or valence state, so to create a binary classification model. Before experimenting with the classification model, it was necessary to do more tests for feature selection, using different methods other than the one in 2.3.3, including:

- **Variance**: removing the features with low variance, as they could be unable to represent a high enough degree of variability between emotional states;

- **Correlation**: removing redundant features;

- **K-Best Method**: uses an ANOVA F-value based function to give a score to each feature, selecting the k best ones, where k is an hyperparameter.

Once different sets of features were generated, it was possible to proceed with the model testing, after having opportunely separated the dataset into training (80% and test (20%), and having scaled the data of both through the generation of a scaler on the train test and then by its application on both on train and test sets. KNeighbors was used to compare in an efficient way the various sets, to then proceed and train other models.

The score that was used to evaluate the model's results is accuracy, defined as the number of correct predictions over the total number of predictions [43]. To assess the performance of the trained binary classification models the ROC curve was computed. A ROC (Receiver Operating Characteristic) curve is a graphical representation of the performance of a binary classification model. It shows the trade-off between the true positive rate (sensitivity) and the false positive rate (1-specificity) of the model for various threshold values. To have a quantitative metric to assess, the Area Under the Curve (AUC) was also computed.

The tested models are reported in the next bulleted list with a brief description adapted from [44]:

- **KNN**: K-Nearest Neighbors is a supervised machine learning algorithm used for classification and regression tasks. It is a non-parametric, instance-based learning algorithm that assigns a label to a data point by considering its closest "neighbors". In KNN, each data point is compared to its nearest neighbors in the training set, based on a similarity metric, and the majority label of the neighbors is assigned to the data point. KNN is considered to be an effective and simple classification method, although not computationally optimized.

- **Decision Tree**: Decision Tree is a supervised machine learning algorithm used for both classification and regression tasks. It is a tree-like model of decisions, where each branch of the tree represents a possible outcome of a given input. The algorithm builds a model by making predictions based on the available data, and the predictions are made by traversing down the tree. Decision Tree is a powerful and easy to interpret algorithm, however, it can be prone to overfitting if the tree becomes too complex.

- **Random Forest**: Random Forest is an ensemble machine learning algorithm used for both classification and regression tasks. It is an ensemble of decision trees (group of models that work in parallel), where each tree is trained on a different

subset of the data and their predictions are averaged, resulting in the final output. Random Forest is considered to be an accurate and robust algorithm, but it is not recommended for small datasets.

- **MLP Classifier**: Multi-Layer Perceptron Classifier is a supervised machine learning algorithm used for classification. It is a neural network-based algorithm that consists of multiple layers of neurons and is trained through backpropagation. It is considered a powerful model for classification, but can result prone to overfitting if the number of layers and neurons is too large.

- **Linear SVC**: Support Vector Classifier is a supervised machine learning algorithm used for classification. It is a linear classifier that uses a Support Vector Machine (SVM) to build a model that can separate the data points into two classes. Linear SVC is considered to be an accurate and robust algorithm, however, it can be computationally expensive and prone to overfitting if the number of features is too high. This model couldn't be used for multiclass classification.

- **Logistic Regression**: Logistic Regression is a supervised machine learning algorithm used for classification tasks. It is a linear classifier that uses a logistic function to build a model that can separate the data points into two classes. Logistic Regression is considered to be an effective and simple algorithm, however, it can be prone to overfitting if the number of features is too high. This model can only be used for binary classification.

- **Gaussian Bayesian Classifier**: Gaussian Bayesian Classifier is a supervised machine learning algorithm used for classification tasks. It is a probabilistic classifier that uses a Gaussian distribution to build a model that can estimate the probability of an input belonging to a certain class. Tends to overfitting if the number of features is too high.

- **Discriminant Analysis**: Discriminant Analysis is a powerful supervised statistical technique used to classify data into distinct categories. Discriminant Analysis is based on finding a linear combination of features that best separate the different classes.

2.5.   Summary of Methods

In this section the methodology of the signal acquisition, of the protocol's design and of all subsequent analyses (both statistical and machine learning) were explained in detail. To summarize the main points:

– **Protocol Structure**: the protocol was divided into two elicitating methods, that were presented to all the subjects of the experiment. The first part was a traditional image-based flat screen method (FS), in which five selected OASIS pictures were presented for each target emotion. Each elicitating part was separated by a neutral baseline. The second part of the protocol, the VR stimulation, was structured in the same way as the FS and consisted of a series of 3D immersive scenarios that tried to elicit the target emotions using the full range of VR's capabilities. After each part of the protocol, a subjective evaluation questionnaire was submitted to every subject, where their perception of the emotitional elicitation was surveyed using graphical methods such as SAM and PAM.

– **Signal Acquisition and Processing**: for each partecipant, signals such as ECG, BVP, GSR and Respiratory were acquired, filtered and analysed. 30 features were extracted from these signals and, as for the cardiac features, they were obtained through the use of a Point Process Model.

– **Statistical Analysis**: the questionnaire's SAM results of the FS section were used to qualitatively identify the subjective emotional response against the expected arousal and valence taken of the used OASIS pictures, in order to assess the precision of the FS stimulation. A similar work was done for VR responses, as they were plotted against the FS ones to assess the differences within the two protocols. Statistical analysis of the extracted features followed a number of steps, starting from a visual assessment of their medians' trends, following with a depolarization of the features acquired in the elicitating scenes in respect with the previous baseline, and then proceeding with statistical tests such as Friedman's (to assess which features responded significantly to certain emotive states) and Wilcoxon (to view differences in response related more to the method of elicitation than the emotive state).

– **Feature Selection and Machine Learning**: using the Square Method, three features that best responded to the various target emotions were found for both elicitating methods. Then, after having selected a number of possible subsets of features, their values were used to train a number of machine learning-based classification algorithms to assess their separating power in three tasks: multi-class classification (trying to separate all 4 emotive states), arousal level-based classification (binary separation among low and high arousal states) and valence level-based classification (binary separation among low and high arousal states).

3 | Results

3.1. Questionnaires Assessment

In the following section the results collected from both surveys are shown, the results will be divided in subsections for each elicitation section of both experimental modalities.

3.1.1. General Subjects Information

The subject were 23 ± 2.56 years old, in a range from 18 to 29 years.
Half of the subjects reported previous experience with VR headsets. The average perceived protocol duration is of 17.6 minutes for the FS stimulation section of the protocol (while the actual duration was of 17.5 minutes) and 27.4 min for the VR elicitation part (against the actual 32 minutes).
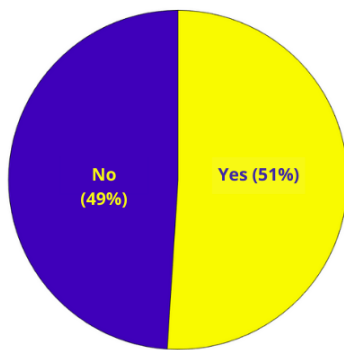
(a) Previous VR experience
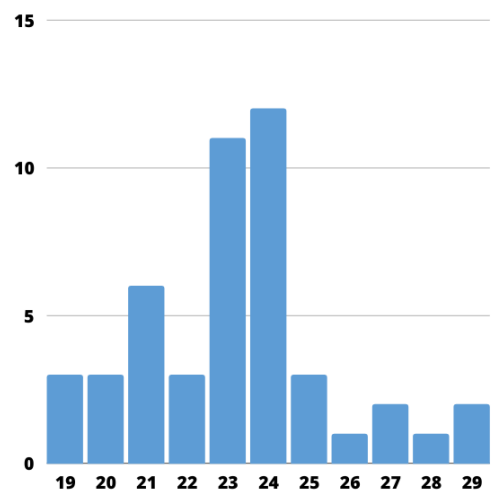
(b) Age distribution

Figure 3.1: General subjects' information

3.1.2. Post-Flat Screen Questionnaire

Sadness

The qualitative assessment of the self-report post elicitation surveys shows a significant displacement towards higher valence than expected from the validated values of the OASIS pictures used (2.08 VS 1.79); arousal is also higher (2.88 vs 2.20). This could be due to a culturally based bias and, as will be shown later in this section, a perceived prevalent neutral mood (that has a more positive valence than the expected one). To better visualize the results figure 3.2 is reported; it is possible to observe the mean value of arousal-valence obtained from the survey (in red) against the mean values from the used OASIS pictures (black dot), the square represents the confidence intervals (calculated as previously stated in 2.3.2) among the votes.
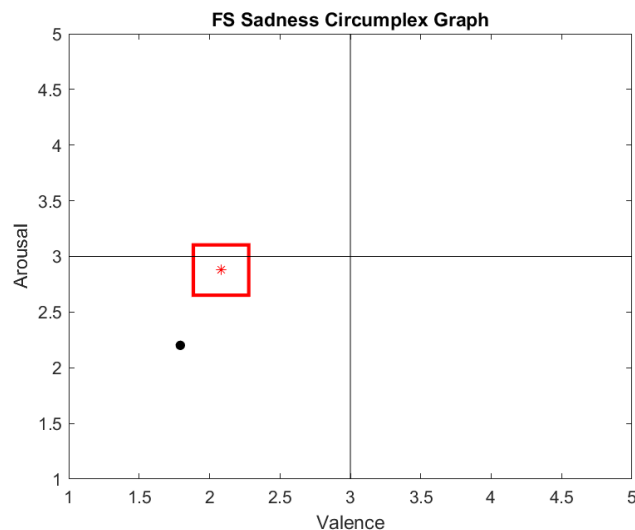
Figure 3.2: Placement of the perceived Sadness valence/arousal values in respect with the OASIS.

As for the PAM results, the prevalent mood was Neutral, followed by Sad and finally Irritated.
The results can be seen in figure 3.3.
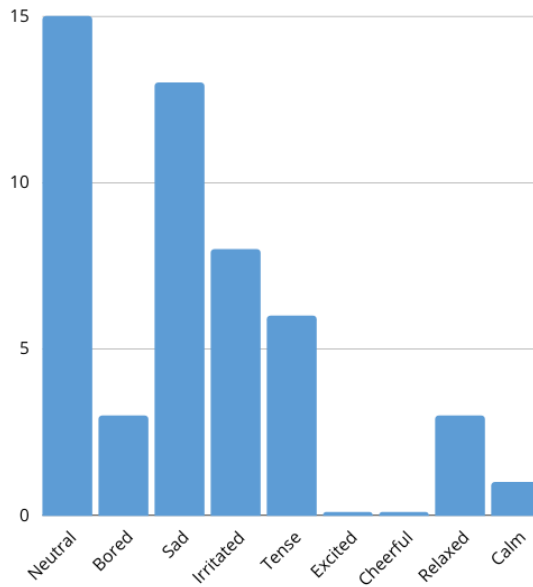
Figure 3.3: PAM results for Flat Screen Sadness

Relax

As for the Relax Flat Screen stimulation, the subjective perception of valence was significantly higher (3.88 vs 3.33), while arousal perception was similar to the expected (1.80 vs 1.52). To better visualize the results figure 3.4 is reported;
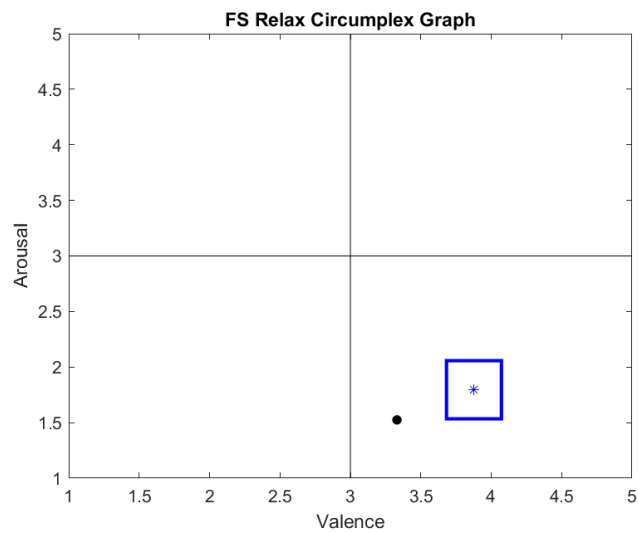
Figure 3.4: Placement of the perceived Relaxation valence/arousal values in respect with the OASIS.

it is possible to observe the mean value of arousal-valence obtained from the survey (in

blue) against the mean values from the used OASIS pictures (black dot), the square represents the square represents the confidence intervals (calculated as previously stated in 2.3.2) among the votes.

The reported PAM moods were highly on point, with the majority of subjects reporting Relaxed and Calm.

The results can be seen in figure 3.5.
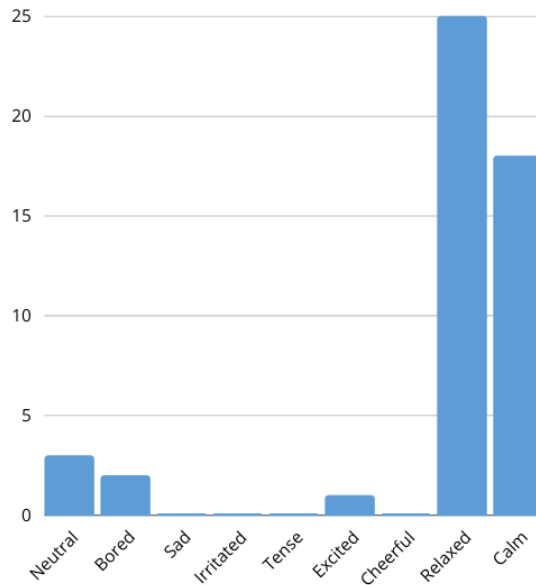
Figure 3.5: PAM results for Flat Screen Relaxation

Happiness

The qualitative assessment of the self-report post elicitation surveys shows a slightly more neutral response with respect to the tabulated values from the OASIS database, for both the valence (4.2 in the self-report vs 4.46 medium value from OASIS) and the Arousal (2.76 in the self-report vs 3.41 medium value from OASIS) scores, it is notable that with a mean score of 2.76 the resulting self-reported Arousal value it is under the neutral value of 3, missing the intended quadrant of the Circumplex plane.

To better visualize the results figure 3.6 is reported; it is possible to observe the mean value of arousal-valence obtained from the survey (in green) against the mean values from the used OASIS pictures (black dot), the square represents the square represents the confidence intervals (calculated as previously stated in 2.3.2) among the votes.
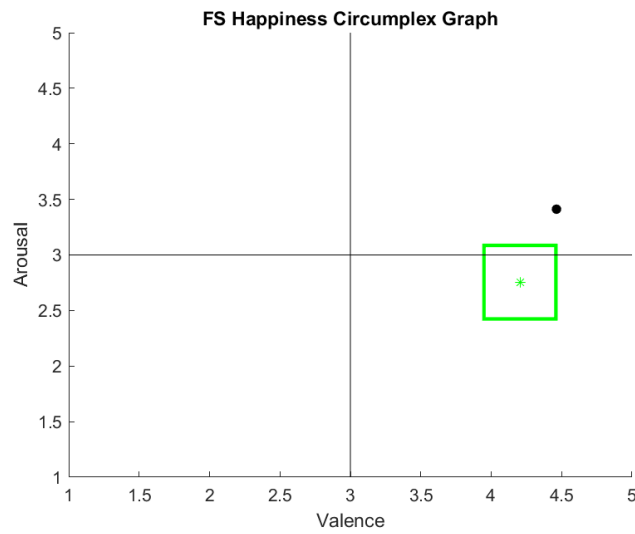
Figure 3.6: Placement of the perceived Happiness valence/arousal values in respect with the OASIS.

As for the PAM results, the prevalent mood was Cheerful, followed by Relaxed and Calm. The results can be seen in figure 3.7.
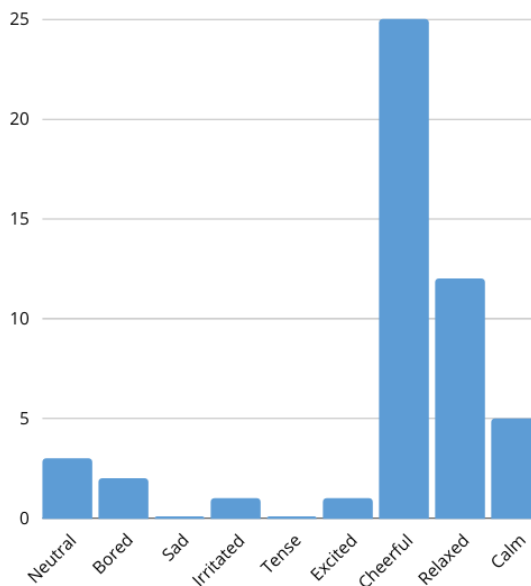
Figure 3.7: PAM results for Flat Screen Happiness

Fear

As for the Fear FS Fear elicitation, the subjective perception of valence was slightly higher than the OASIS tabulated mean value (1.55 vs 1.36), and the arousal perception

was similar to the expected one(3.51 vs 3.47).

To better visualize the results figure 3.8 is reported; it is possible to observe the mean value of arousal-valence obtained from the survey (in light blue) against the mean values from the used OASIS pictures (black dot), the squares represents the square represents the confidence intervals (calculated as previously stated in 2.3.2) among the votes.
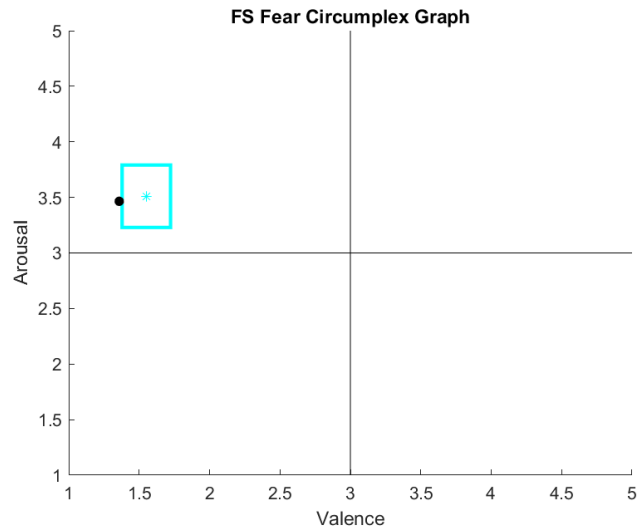
Figure 3.8: Placement of the perceived Fear valence/arousal values in respect with the OASIS.
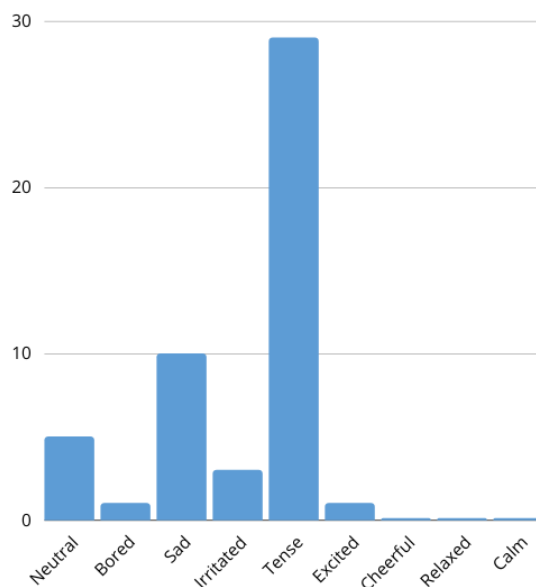
Figure 3.9: PAM results for Flat Screen Fear

The reported PAM moods were on point, with the majority of subjects reporting Tense

and Sad.

The results can be seen in figure 3.9.

3.1.3. Post-Virtual Reality stimulation Questionnaire

Baselines

In the second self-report assessment an additional section to qualitatively assess the effect of the used baselines was introduced.

The subjective perception of valence was slightly positive with a mean value of 3.37; the self-reported arousal was instead located significantly in the low arousal part of the axis, with a mean value of 1.76.

A graphical representation of the baseline self reported values is shown in figure 3.10.
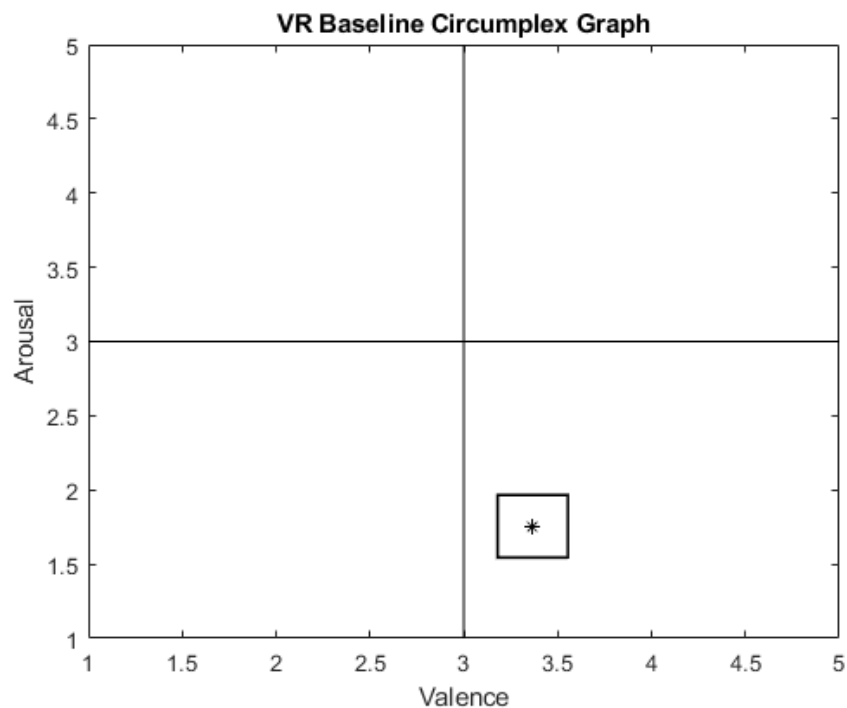
Figure 3.10: Placement of the perceived VR Baseline valence/arousal values

As for the PAM results, the prevalent mood reported was Bored, followed by Calm and Relaxed, coherently with the intended stimulation.

The results can be seen in figure 3.11.
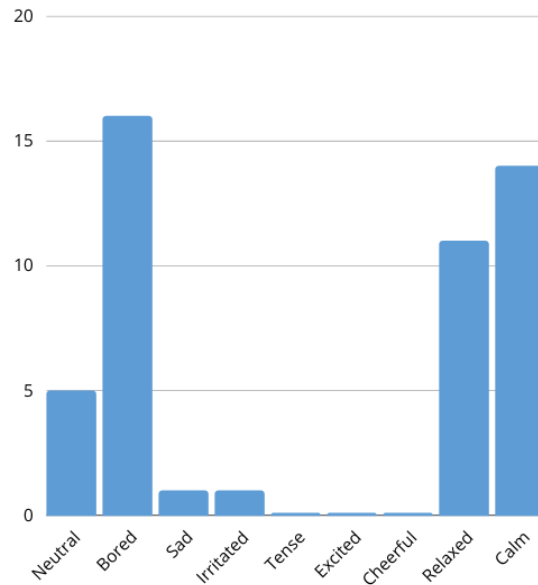
Figure 3.11: PAM results for the baselines.

Sadness

The qualitative assessment of the self-report post elicitation surveys shows a mean valence score of 2.04, slightly less that the Flat Screen elicitation results for the corresponding scene; the mean arousal score is 2.35 (versus the 2.88 from the FS stimulation). The elicitated emotion seems more polarized toward the extreme regions of the quadrant with respect to the one elicitated with the FS stimulation.
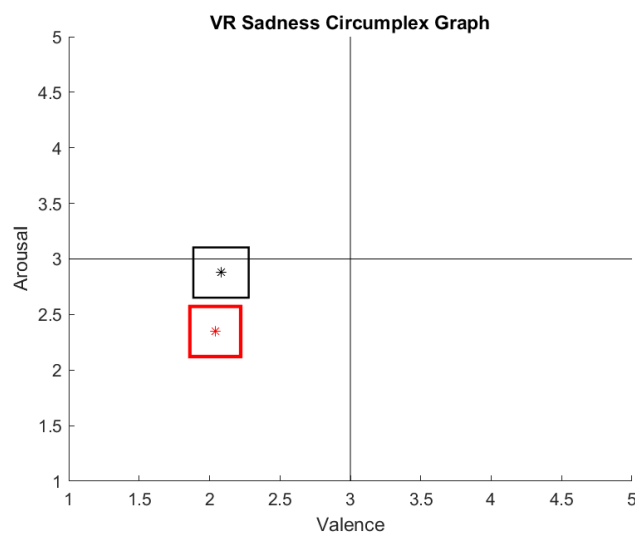
Figure 3.12: Placement of the perceived VR Sadness valence-arousal values in respect with the FS values.

To better visualize the comparison figure 3.12 is reported; it's possible to observe the mean value of arousal-valence obtained from the second survey (in red) against the mean values from the first survey (in black), the square represents the confidence intervals (calculated as previously stated in 2.3.2) among the votes.

As for the PAM results, the prevalent mood was Sad, coherently with what was expected followed by Tense and Neutral (probably due to the low value of arousal).
The results can be seen in figure 3.13.
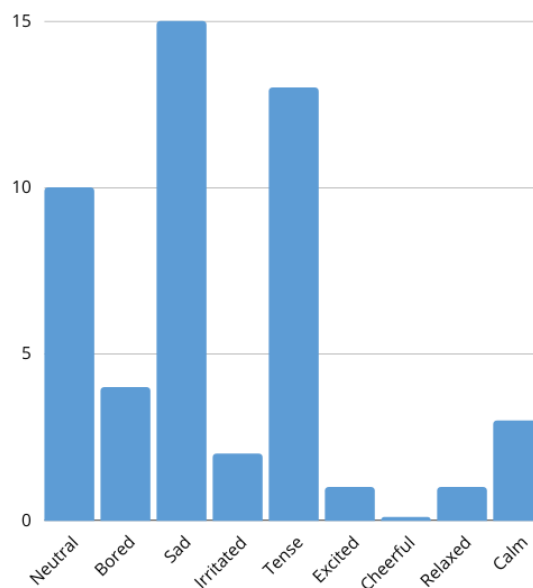
Figure 3.13: PAM results for VR Sadness

Relaxation

The qualitative assessment of the second self-report post elicitation survey shows a mean valence score of 4.31, higher than the Flat Screen elicitation results for the corresponding scene; the mean arousal score is 1.55 (versus the 1.80 from the FS stimulation). The elicitated emotion seems more centered toward the extreme regions of the quadrant with respect to the one elicitated with the FS, for both arousal and valence values.
To better visualize the comparison figure 3.14 is reported; it's possible to observe the mean value of arousal-valence obtained from the second survey (in blue) against the mean values from the first survey (in black), the square represents the confidence intervals (calculated as previously stated in 2.3.2) among the votes.
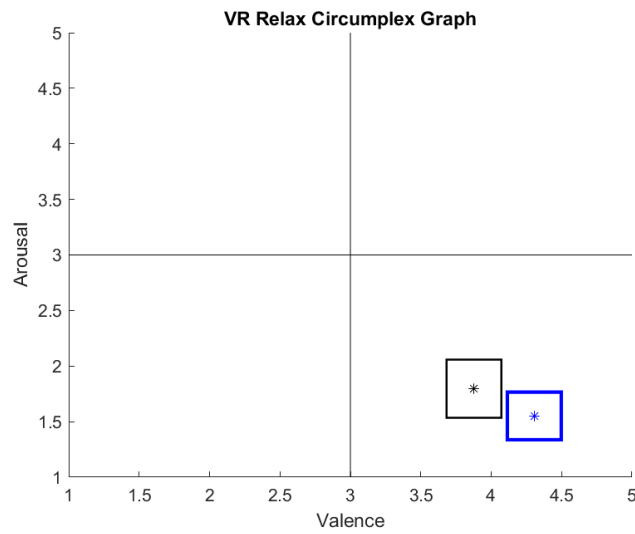
Figure 3.14: Placement of the perceived VR Relaxation valence-arousal values in respect with the FS values.

As for the PAM results, the majority of the collected answers was Relaxed and Calm coherently with what was expected, with a notable number of "Cheerful" (probably due to the positive valence of the elicitated emotion).

The results can be seen in figure 3.15.
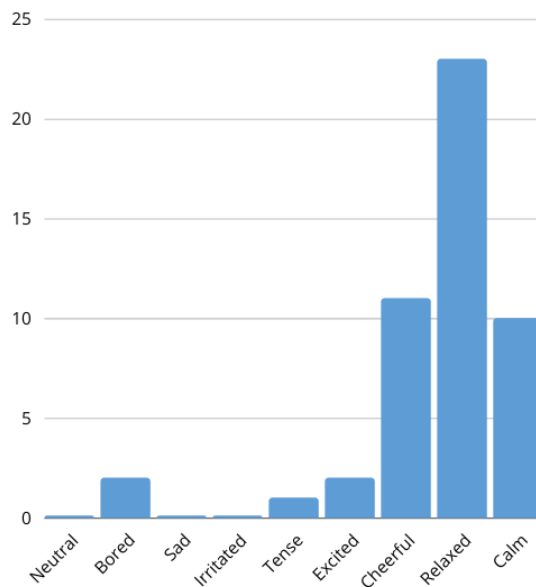
Figure 3.15: PAM results for VR Relaxation

Happiness

As for the Happiness VR stimulation, the subjective perception of valence was similar to the Valence score from the FS elicitation (4.25 vs 4.20), while arousal perception was significantly higher, with a score of 3.81 against the 2.76 from the FS stimulation. It is notable that in the VR stimulation the self reported score for arousal is coherent with the intended elicitated emotion in terms of Circumplex plane's quadrant position.

To better visualize the comparison figure 3.16 is reported; it's possible to observe the mean value of arousal-valence obtained from the second survey (in green) against the mean values from the first survey (in black), the square represents the confidence intervals (calculated as previously stated in 2.3.2) among the votes.
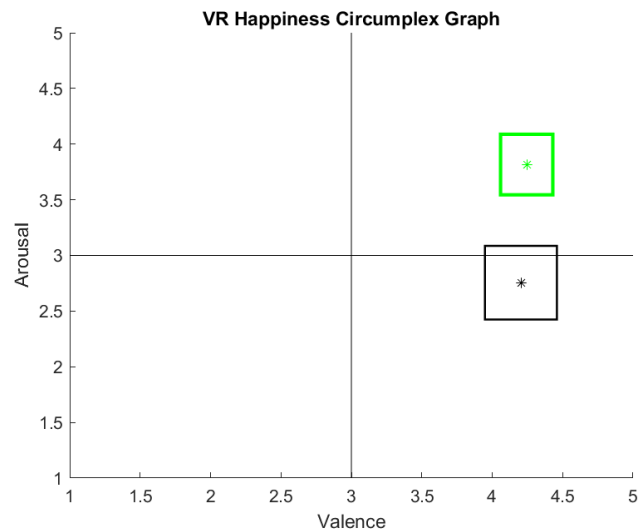
Figure 3.16: Placement of the perceived VR Happiness valence-arousal values in respect with the FS values.

As for the PAM results, the majority of the collected answers was Cheerful and Excited coherently with what was expected, followed by Relaxed and Calm, that refer to the correct value for valence but lower arousal states.

The results can be seen in figure 3.17.
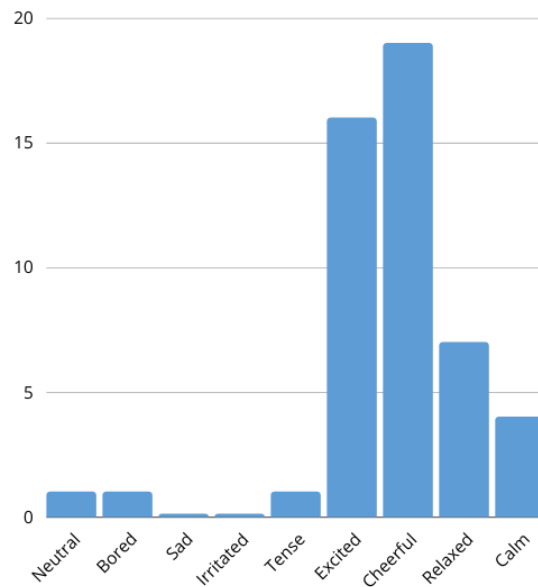
Figure 3.17: PAM results for VR Happiness

Fear

As for the Fear VR stimulation, the subjective perception of valence was slightly higher than the valence score from the FS elicitation (2.06 vs 1.55), probably because of the less visually explicit nature of the stimuli, also the arousal perception was higher, with a score of 4.06 against the 3.51 from the FS stimulation.

To better visualize the comparison figure 3.18 is reported; it's possible to observe the mean value of arousal-valence obtained from the second survey (in Cyan) against the mean values from the first survey (in black), the square represents the confidence intervals (calculated as previously stated in 2.3.2) among the votes.

As for the PAM results, the majority of the collected answers was Tense and Irritated coherently with what was expected.

The results can be seen in figure 3.19.
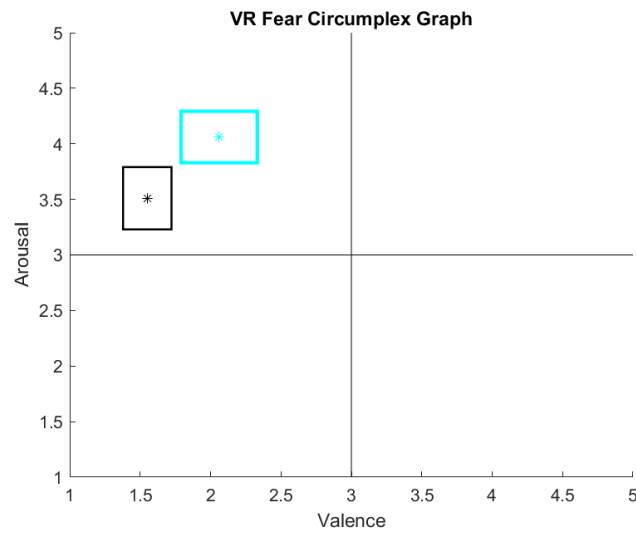
Figure 3.18: Placement of the perceived VR Fear valence/arousal values in respect with the FS values.
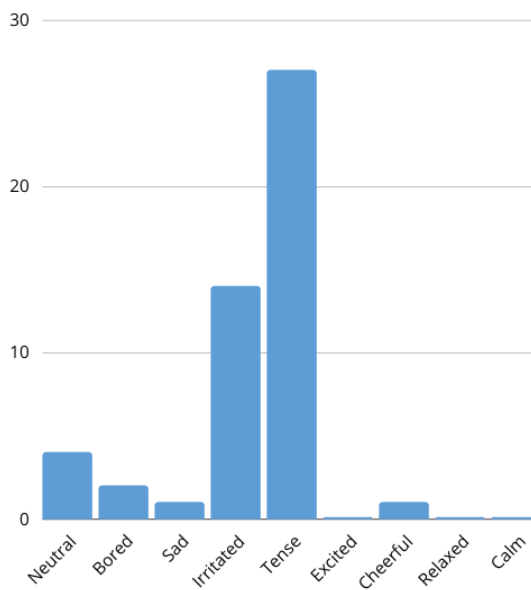
Figure 3.19: PAM results for VR Fear

Summary of Comparisons

To summarize the above results, a plot referring to all 8 parts is reported in figure 3.20; the four elicitated emotional states are differentiated by color, where red corresponds to Sadness, blue corresponds to Relaxation, green corresponds to Happiness and cyan refers to Fear. To distinguish the two different stimulation modalities from one another a darker

version of the four colors was used for the FS elicitation. The mean values of the used OASIS pictures are reported as black dots on the plot.
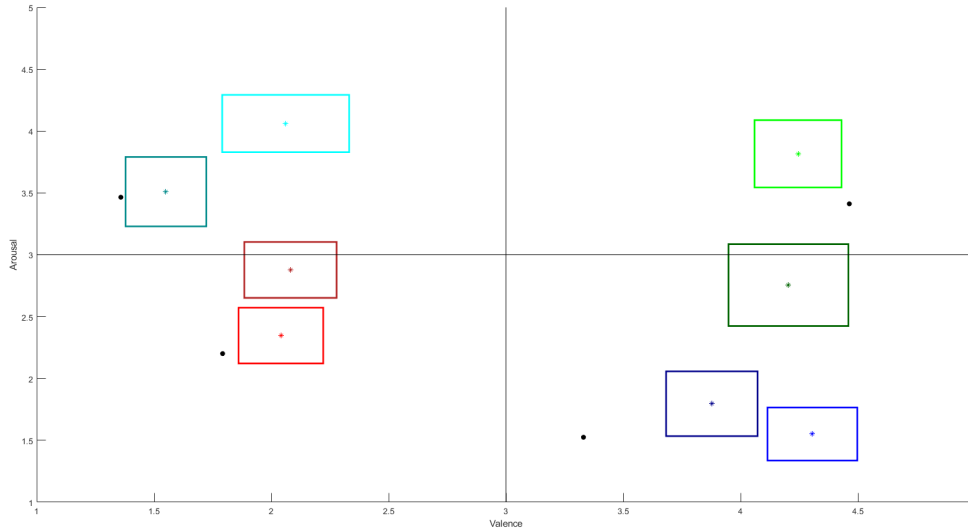
Figure 3.20: Placement of the perceived VR valence-arousal values with respect to the FS and OASIS.

To easily read figure 3.20, the following tables 3.1 have been reported summarizing the comparisons between the scores of the selected OASIS pictures and the self-reported scores of FS stimulation(a), and the aforementioned with the VR stimulation scores (b).

Scene	ΔArousal	ΔValence
Sad	0.48($\uparrow$)	0.29($\uparrow$)
Rlx	0.28($\uparrow$)	0.55($\uparrow$)
Hap	-0.65($\downarrow$)	-0.26($\downarrow$)
Fear	0.04($\simeq$)	0.19($\uparrow$)

(a) Variations of valence-arousal scores between FS stimulation and used OASIS.

Scene	ΔArousal	ΔValence
Sad	-0.53 ($\downarrow$)	-0.04($\simeq$)
Rlx	0.25($\uparrow$)	0.43($\uparrow$)
Hap	1.05($\uparrow\uparrow$)	0.05($\simeq$)
Fear	0.55($\uparrow$)	0.51($\uparrow$)

(b) Variations of valence-arousal scores between FS and VR elicitation.

Table 3.1: Variations of valence-arousal scores

3.2. Physiological Response Analysis

The plots of median values of some of the acquired features for the Flat Screen and Virtual Reality stimulations are shown in the following sections.

3.2.1. GSR Features

The GSR signal is a measure of sweat gland activity in the skin, which reflects changes in the activity of the sympathetic nervous system. It provides information about the strength of the response of the sympathetic nervous system, so in an emotional stimulation scenario it is an indicator of arousal; in general an higher value of the features is indicator of a high arousal emotional state.

Taking as an example the trends of the features shown in the table 3.2, it is shown that the VR stimulation elicits a stronger response of the sympathetic nervous system, so it can be linked with higher states of arousal. From the oscillatory trend between baselines and stimulating scenes in each elicitation modality seen in the feature N Peaks, the neutralizing effect of each baseline is observable.

Feature Name	Median trends
Avg rise time	

N peaks	
Average	
Env	
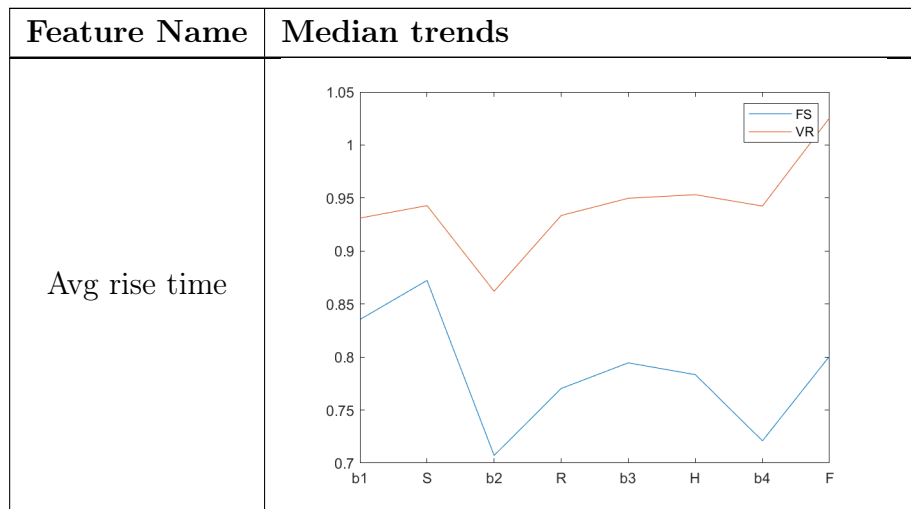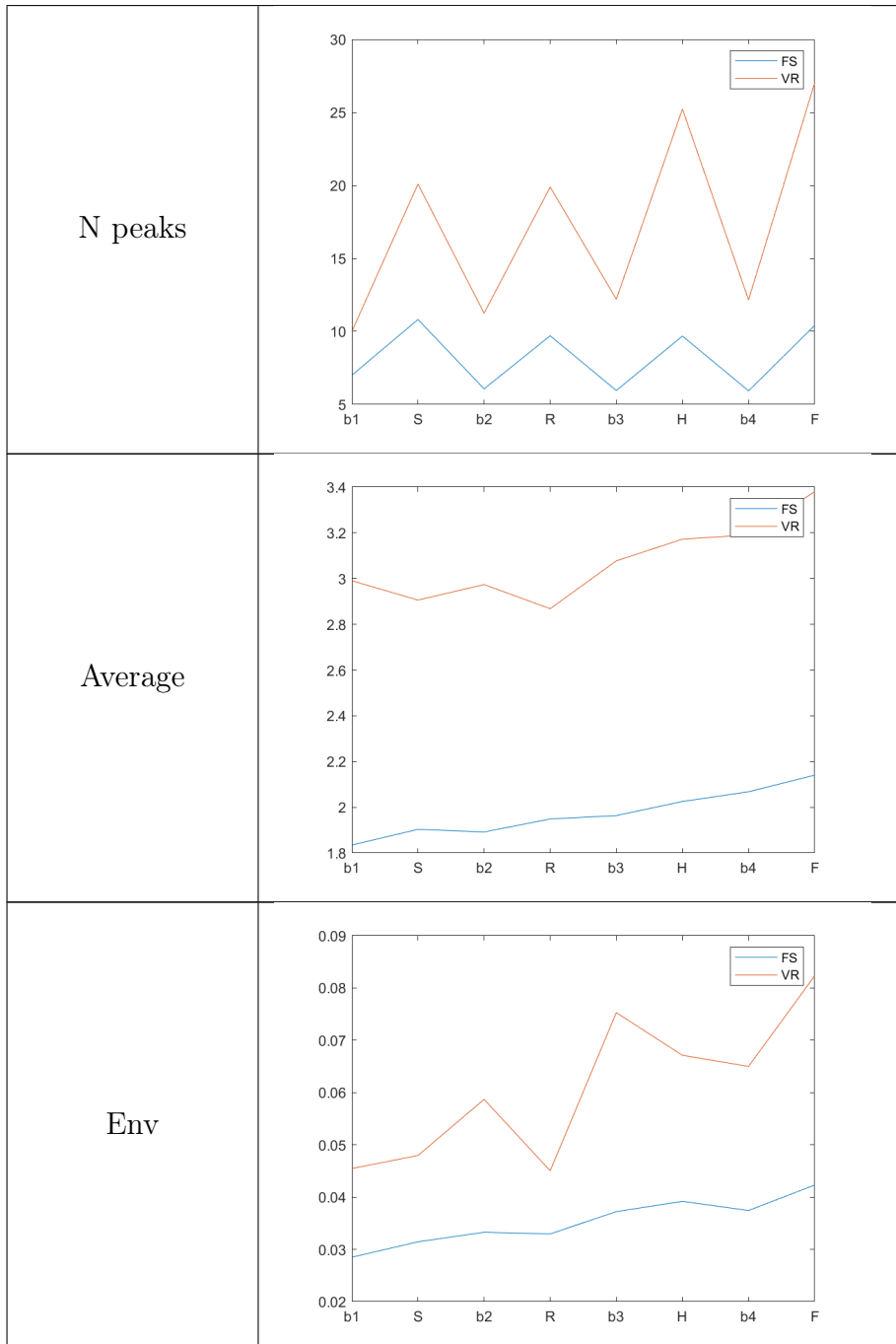
Table 3.2: Example of Median Trend of Extracted GSR Features

3.2.2. Point Process Features

From the features extracted from the Point Process Model time series information about the activity of the Autonomic Nervous System can be deduced.

From the reported trends we can see that the ANS response was coherent with the literature. For example the high frequency (HF) power of the modeled heart rate variability signal is associated with parasympathetic nervous system activity, responsible for slowing down the heart rate and promoting relaxation. Therefore, a higher HRV HF power indicates greater parasympathetic activity, which is associated with a relaxed state. In contrast, a lower HRV HF power indicates a lower level of parasympathetic activity and may indicate a more stressed or aroused state.

Instead, the low frequency power (or the normalized LF) is associated with sympathetic activation, and is expected to be lower in relaxed states and higher in more aroused ones, as can be observed in Table 3.3 the two trends are both rising during the protocol and the VR stimulation results more arousing in general.

Feature Name	Median trends
HF	
lfn	

Table 3.3: Example of Point Process features median trends.

3.2.3. Respiratory

From the extracted features from the respiratory signal information about the activity of
the ANS can be deduced.

From the two extracted features trend it is observable a response that is coherent with
what was expected. Increased respiratory frequency is associated to states of increased
arousal, such as happiness in which it is observable a noticeable peak in the VR stimulation
part, while decreased respiratory frequency is often observed during states of decreased
arousal, like relaxation. From both trends it is observable the presence of higher values
during the first two halves of the elicitation (corresponding to lower arousal states) and
lower values in the remaining halves.

On the opposite higher respiratory amplitude is associated with low states of arousal and
low values with high arousal states, as is confirmed in the shown trends in Table 3.4.
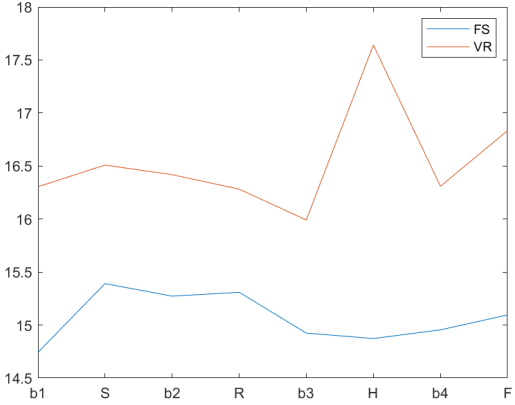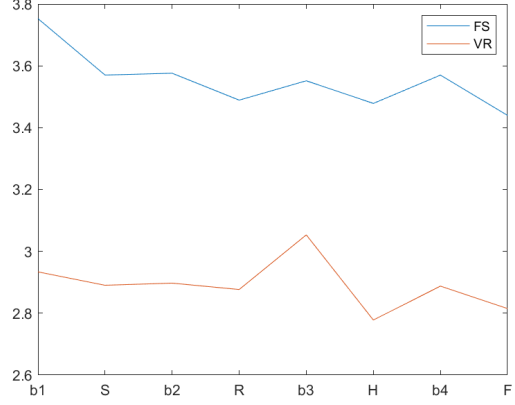
Feature Name	Median trends
Respiratory frequency	
Resp amplitude	

Table 3.4: Respiration features median trends.

3.2.4. PAT

With respect to the ANS, the PAT is affected by changes in sympathetic and parasympathetic activity. Sympathetic activity causes arterial vasoconstriction, increasing arterial stiffness and decreasing the PAT. In contrast, parasympathetic activity causes arterial vasodilation, decreasing arterial stiffness and increasing the PAT, these higher values are associated with lower arousal states. In Table 3.5 is shown that in both trends there is a significant decrease corresponding to the last elicitated emotion, that also corresponds to the most arousing one (Fear).

Feature Name	Median trends
PAT	
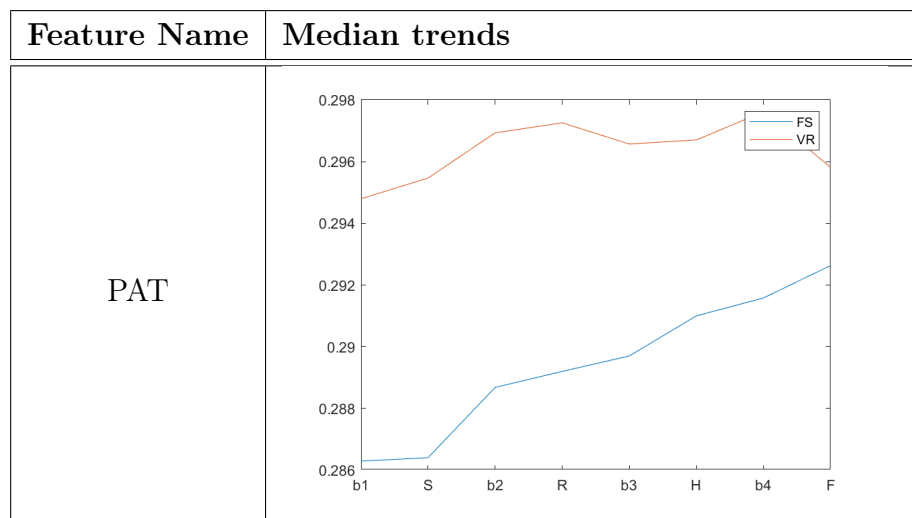
Table 3.5: PAT median trends

3.2.5. PP

A higher Pulse Pressure value indicates a higher parasympathetic activity, in contrast, a lower PP value indicates a lower parasympathetic activity and a high sympathetic activation.

Therefore, PP can be used as an indicator of the ANS's activity, and coherently with what just said the two trends shown in table 3.6 slightly rise from the first baseline to the two low arousal stimulations and then decrease with rising arousal in the elicitation protocol.

Feature Name	Median trends
PP	

Table 3.6: PP Median trends

3.3.   Statistical Analysis Results

As stated in 2.3.2, the statistical testing has been performed over the depolarized dataset, using a dynamic depolarization based on the baseline before the elicitation part of the protocol.

3.3.1.   Friedman's Test Results

Following the method explained in 2.3.2, a Friedman's test was performed between the four elicitation sections for both the FS and the VR part of the protocol, to assess whether the extracted features are effective in distinguish the four elicitated emotions.

For the FS elicitation none of the extracted features resulted significant in distinguish the four different elicitated emotions.
For the VR elicitation, the performed Friedman's Test resulted significant for 20 over 30 extracted features.
The following results refer only to extracted features from the VR stimulation.

To further investigate which couples of comparisons resulted significant a post-hoc test with Bonferroni correction was performed on the features that resulted significant in the previous test.
These comparison results are shown in the following section through the box-plots of the significant Features. The comparisons between emotions are marked with asterisks that represent the significance of the comparison: no asterisks means that the test resulted in a non significant difference, otherwise it resulted significant (there will be one for a

$p-value < 0.05$, two for $p-value < 1*10^{-2}$ or three asterisks for $p-value < 1*10^{-3}$ to easily visualize the magnitude of the significance). Only three of the previously significant features show no significance after the Bonferroni restriction, these are the average Rise Time from the GSR signal, the Mu from the Point Process model and the Respiratory Amplitude.

To better visualize which comparison had most significant features figure 3.21 is reported.
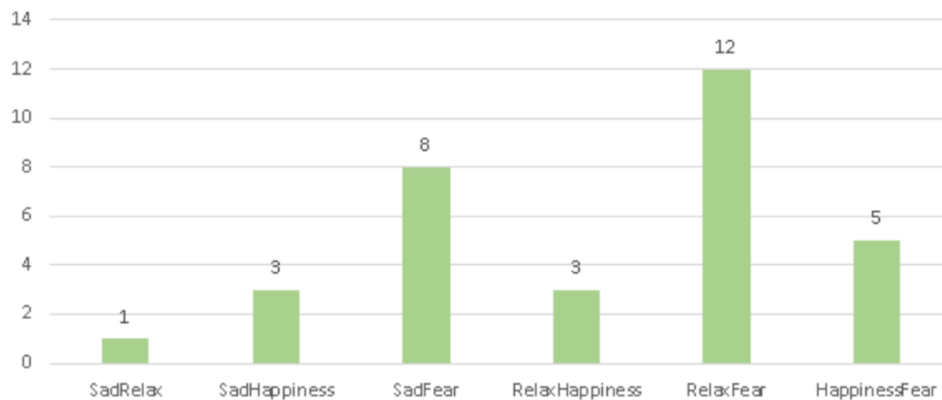
Figure 3.21: Number of significant features per comparison

GSR Features

The skin conductance features proved to be the most significant in term of emotional differences, with 13 significant features over 17 extracted; 12 after the Bonferroni correction. In particular all the significant features can distinguish Fear from the lower arousal state Relax, 6 features are significant for distinguishing Fear from Sadness, 3 for differentiating Happiness from Fear, the comparisons Sadness-Relax, Sadness-Happiness and Relax-Happiness only have one significant feature each.

In the following tables box-plots will have a standard colors for corresponding emotions: Red for Sadness, blue for Relaxation, Green for Happiness and Cyan for Fear, same as the plot colors in sec.3.1.2 and 3.1.3.

Feature Name	Box-Plot
Env	
Average	
N Peaks	

Avg Amplitude of Peaks	
Avg Rise Time	
Sd Amp of Peaks	

Max sign amp gsrbp	
Sd Der gsrbp	
Max Der gsrbp	

Avg Abs 1	
Avg Abs 1 Normalized	
Avg Abs 2	

Avg Abs 2 Normalized	
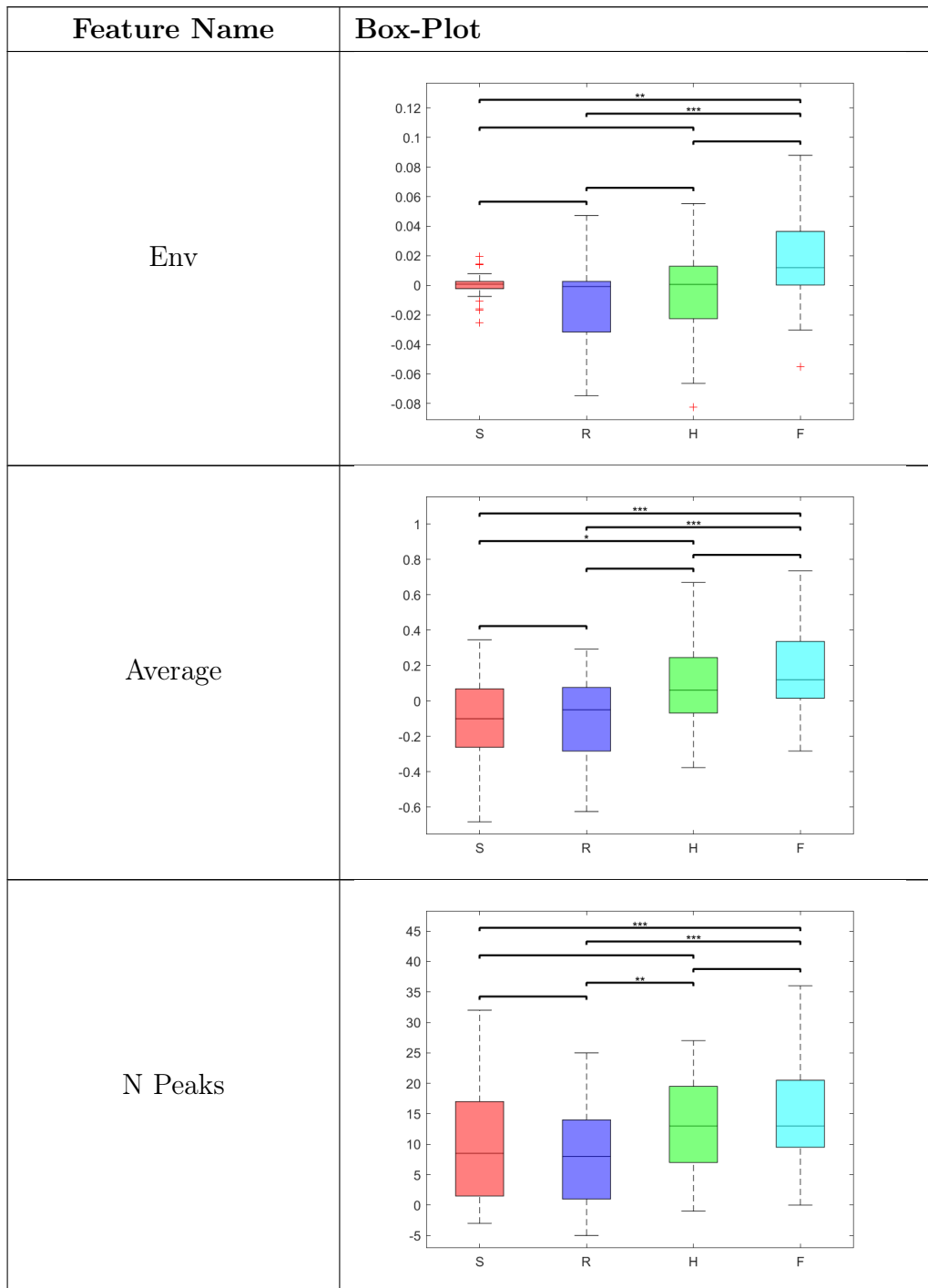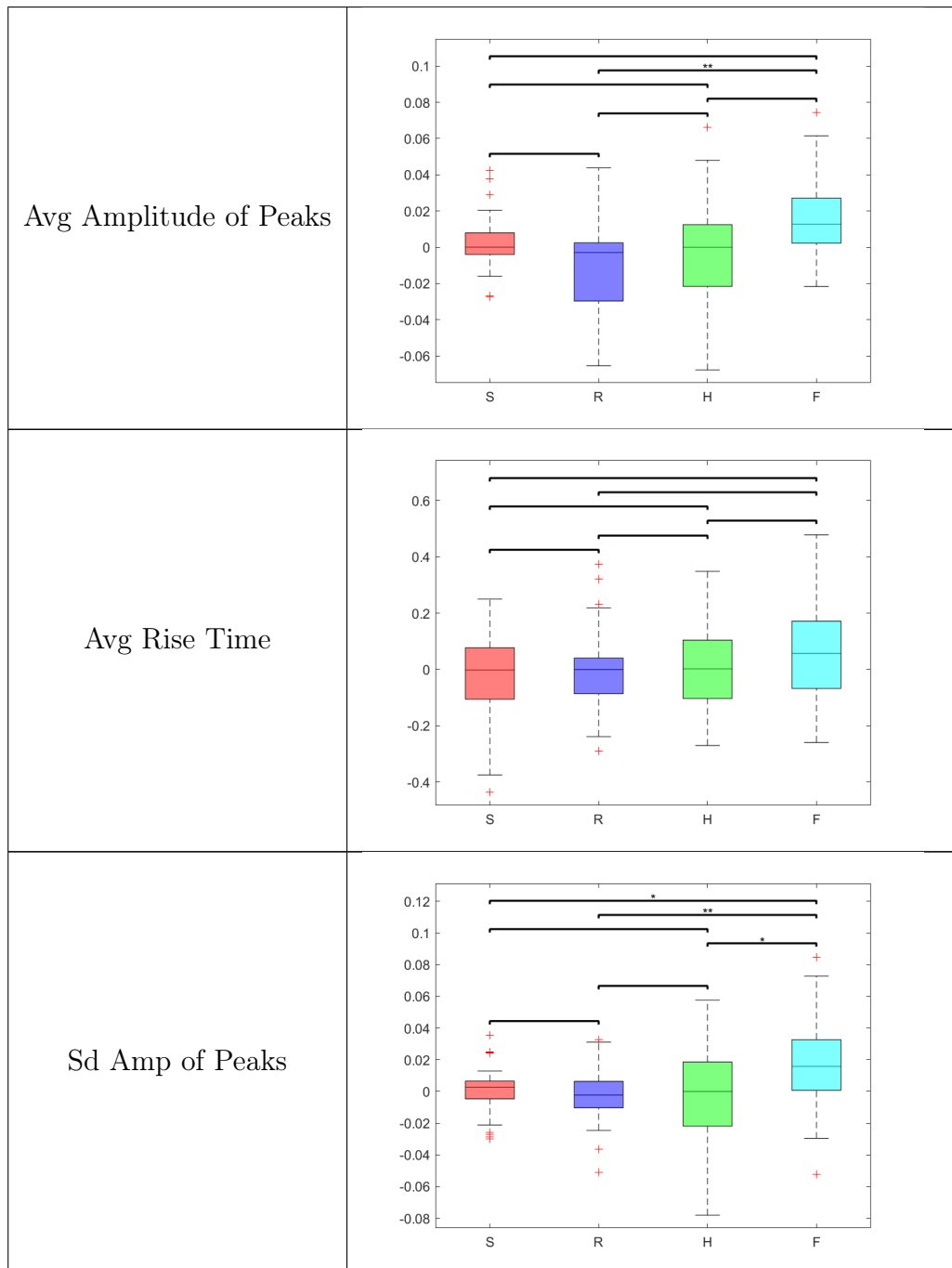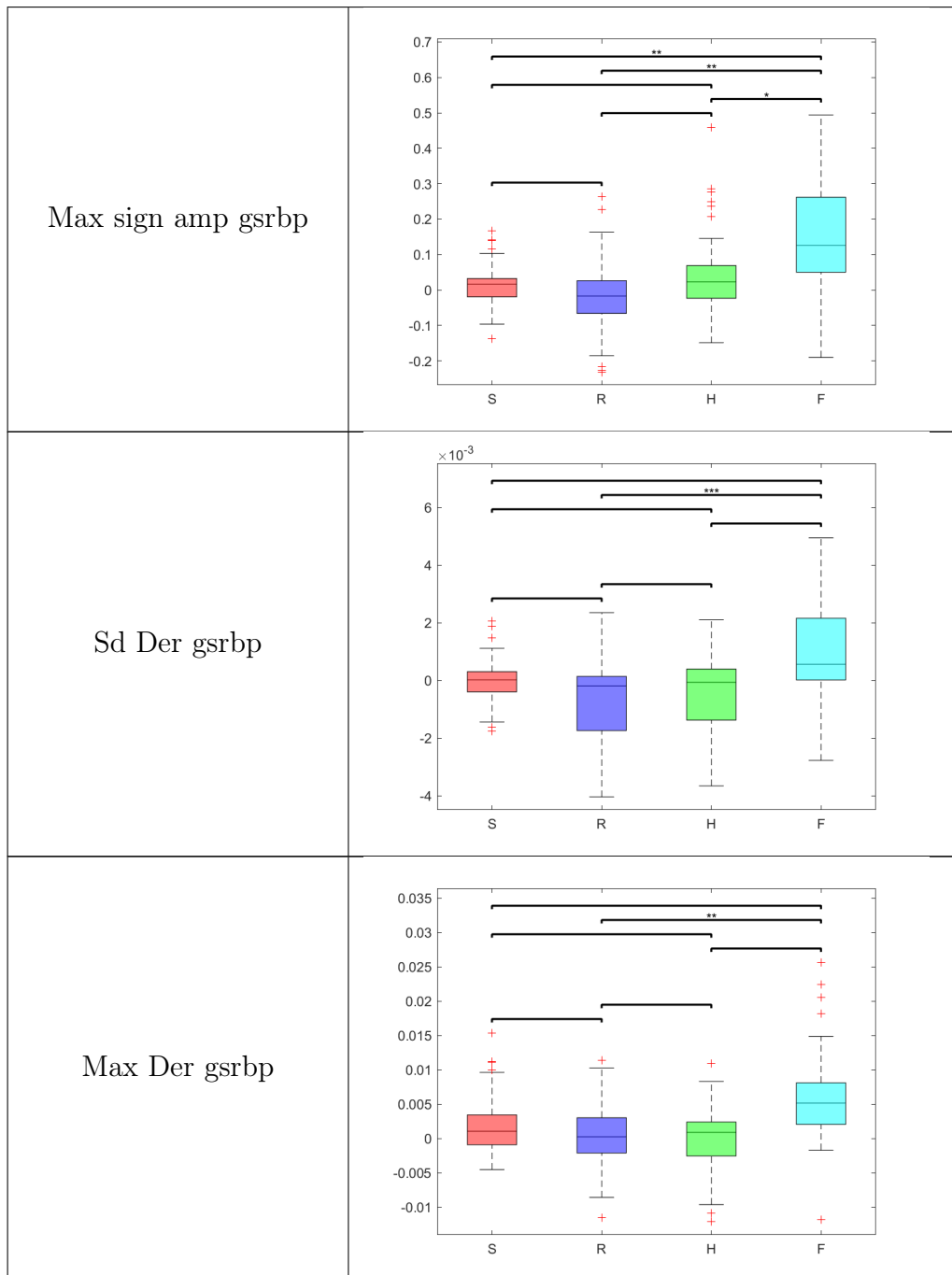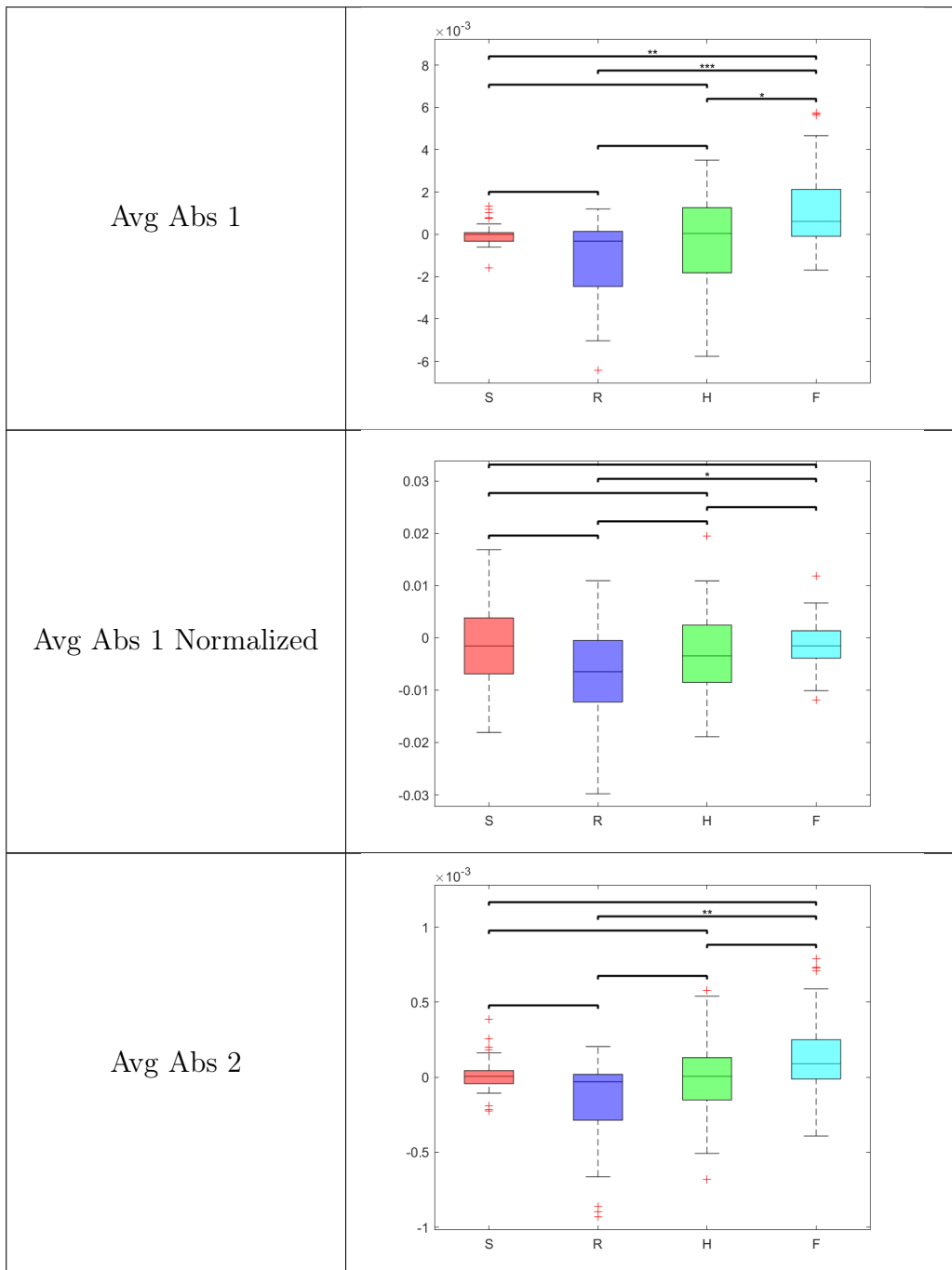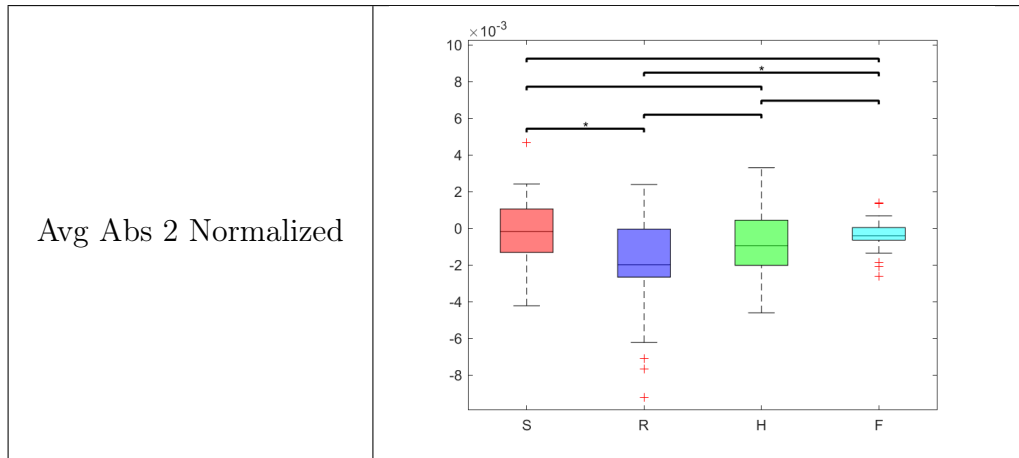
Table 3.7: Statistically significant GSR Features Friedman's Test.

Point Process Features

Three out of the nine Point Process extracted Features from the VR stimulation resulted significant from the Friedman's test, two also resulted significant after the post-hoc test with the Bonferroni restriction.

In particular the HF resulted significant for the comparison Sadness-Fear (so to distinguish a low level of arousal from an high one), and Tot power resulted highly significant for the comparison happiness-fear (so to distinguish a low level of valence from an high one).

In the following table 3.8 box-plots will have a standard colors for corresponding emotions: Red for Sadness, blue for Relaxation, Green for Happiness and Cyan for Fear, same as the plot colors in sec.3.1.2 and 3.1.3.

Feature Name	Box-Plot
Mu	
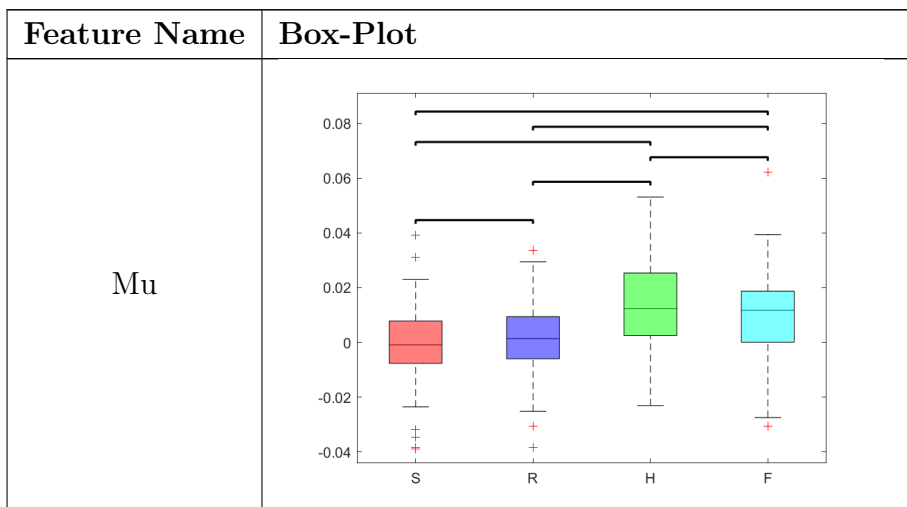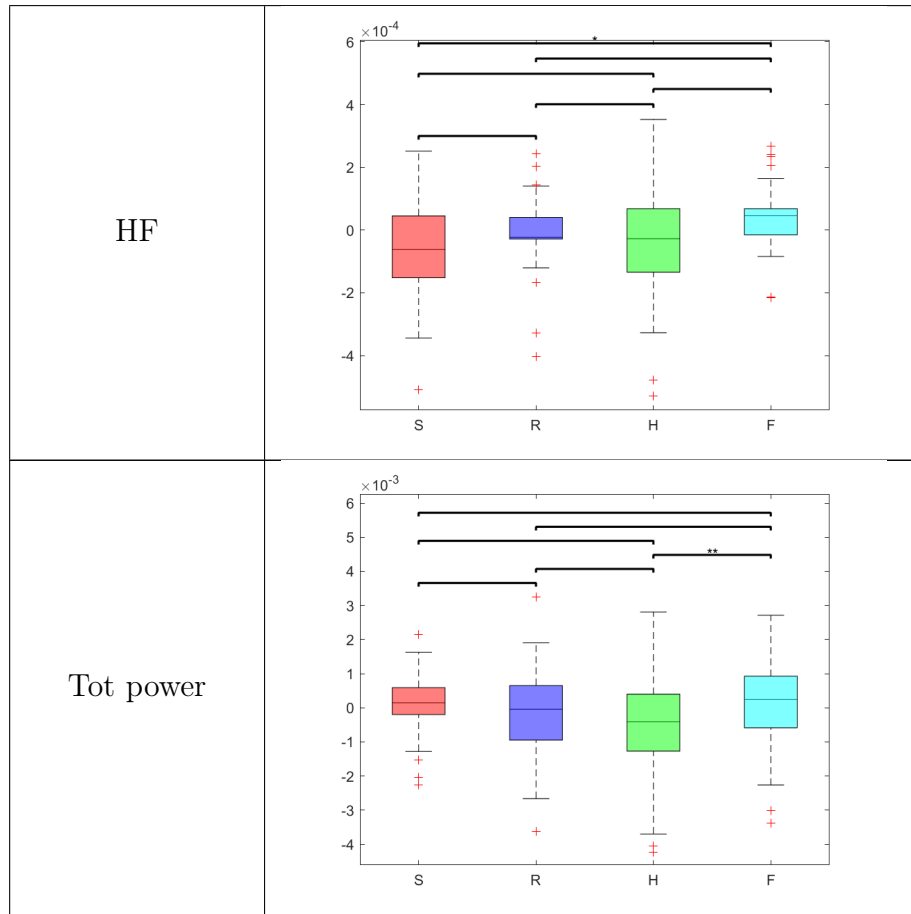
Table 3.8: Statistically significant Point Process Features Friedman's Test.

Respiration Features

Both of the respiration derived features resulted initially significant after the Friedman's test, only the Respiratory frequency remained significant after the post-hoc test with the Bonferroni correction, in particular it resulted significant in distinguishing Happiness from the lower arousal states.

In Table 3.9 box-plots will have a standard colors for corresponding emotions: Red for Sadness, blue for Relaxation, Green for Happiness and Cyan for Fear, same as the plot colors in sec.3.1.2 and 3.1.3.
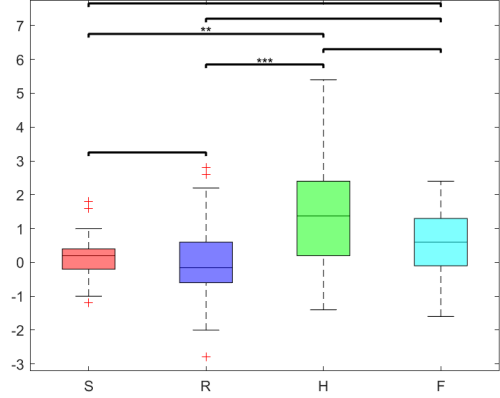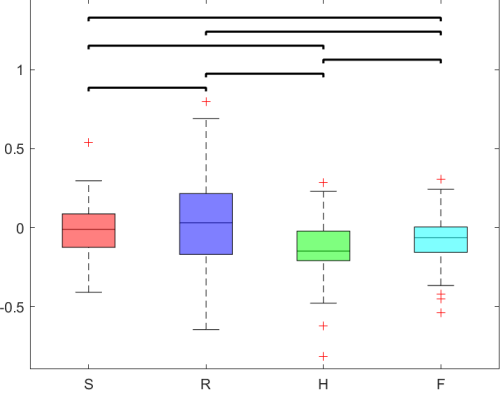
Feature Name	Box-Plot
Respiratory Frequency	
Amp Resp	

Table 3.9: Statistically significant Respiration Features Friedman's test.

PAT

PAT resulted hugely significant for the comparison Sadness-Fear, so in distinguish lower arousal states from, higher ones.

In Table 3.10 box-plots will have a standard colors for corresponding emotions: Red for Sadness, blue for Relaxation, Green for Happiness and Cyan for Fear, same as the plot colors in sections 3.1.2 and 3.1.3.

Feature Name	Box-Plot
PAT	

Table 3.10: Statistically significant PAT Feature Friedman's Test.

PP

The PP time series resulted significant also after the post-hoc test with the Bonferroni correction, it is hugely significant in comparisons that involve Happiness with lower arousal states (Sadness and Relax), and slightly less significant in distinguishing Fear from Happiness.

In table 3.11 box-plots will have a standard colors for corresponding emotions: Red for Sadness, blue for Relaxation, Green for Happiness and Cyan for Fear, same as the plot colors in sec.3.1.2 and 3.1.3.

Feature Name	Box-Plot
PP	

Table 3.11: Statistically significant PP Feature Friedman's Test.

3.3.2.   Wilcoxon Signed-rank Test Results

As stated in 2.3.2 a Wilcoxon Signed-rank test was performed between corresponding stimulation parts of the Flat Screen and the VR part of the protocol, to assess whether there are significant differences between the two elicitation techniques. This test can give information about whether features depend on the stimulation modality or just the elicitated emotion itself.

To easily display these comparison results, the box-plots of the most significant Features are shown in the following section divided by respective signal. In each plot eight box-plots are shown side by side, four for each elicitation modality with standard colors for corresponding emotions: Red for Sadness, blue for Relaxation, Green for Happiness and Cyan for Fear, same as the plot colors in sections 3.1.2, 3.1.3 and 3.3.1.

The comparisons between corresponding emotions are shown with asterisks, those symbols represent the significance of the comparison, no asterisks means that the test resulted in a non significant difference, otherwise the comparison resulted significant (there will be one for a $p-value < 0.05$, two for $p-value < 1*10^{-2}$ or three asterisks for $p-value < 1*10^{-3}$ to easily visualize the magnitude of the significance).

GSR Features

From the GSR extracted features some resulted significantly different in the two elicitation modalities.

In particular Average resulted very significant in the two lower arousal scenes (Fear and Relax); Avg Abs2 Norm resulted instead significantly different for the two low valence sections (Sadness and Fear); the Slope of the band pass filtered signal resulted significantly different for all the elicitated emotions so it seems to distinguish well the two stimulation modes, but as seen in section 3.3.1 it does not result significant in distinguish one emotion from the other.

Feature Name	Box-Plot
Average	
Slope GRSbp	
Avg Abs2 Norm	

Table 3.12: Statistically significant GSR Features Wilcoxon Test.

Point Process Features

From the extracted Point Process features, Balance and HF resulted slightly significantly different for Fear, and Mu for Happiness.

Feature Name	Box-Plot
Mu	
HF	
Bal	

Table 3.13: Statistically significant RESP Features Wilcoxon Test.

Respiration Features

The respiratory frequency resulted highly significantly different only in the Happiness part of the stimulation.

Feature Name	Box-Plot
Respiratory Frequency	
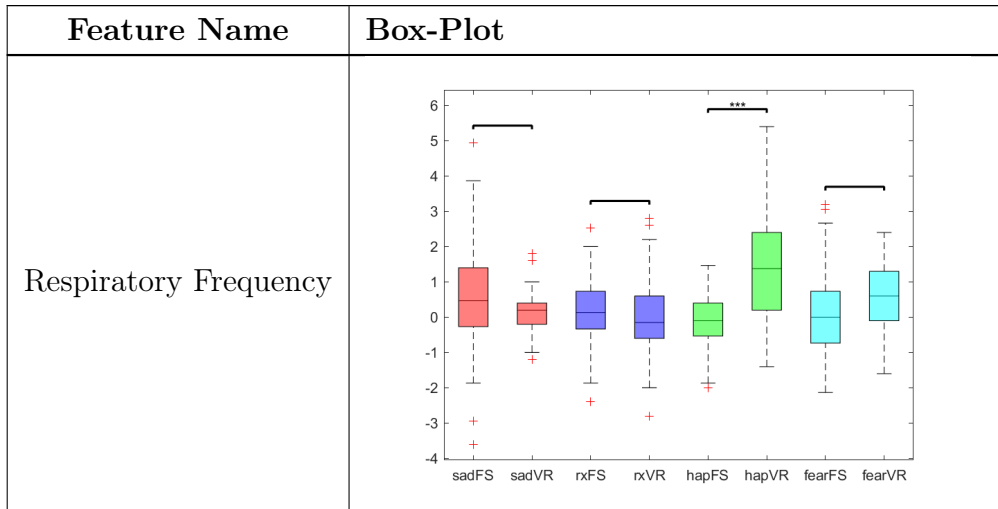
Table 3.14: Statistically significant Point Process Features Wilcoxon Test.

PAT

The Pulse Arrival Time resulted highly significantly different for the Fear section of the two stimulation.

Feature Name	Box-Plot
PAT	
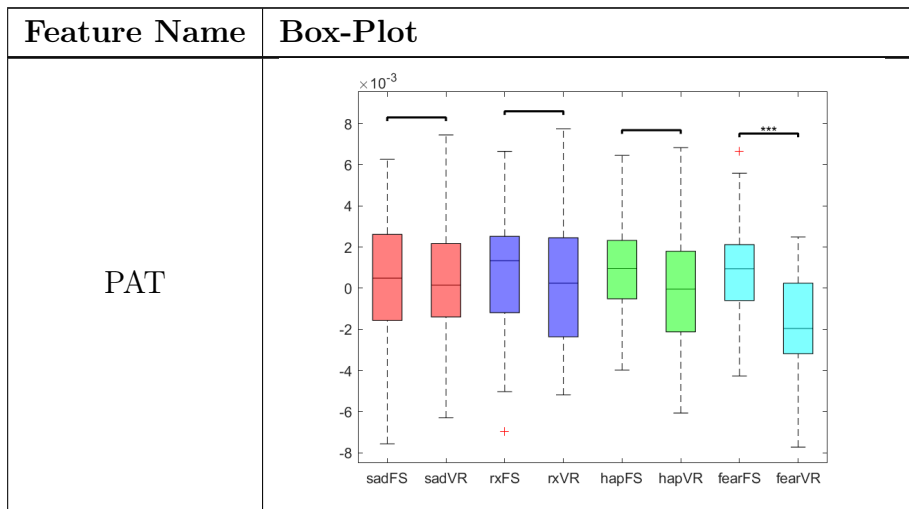
Table 3.15: Statistically significant PAT Feature Wilcoxon Test.

PP

The Pulse Pressure resulted highly significantly different only in the Happiness part of the stimulation.

Feature Name	Box-Plot
PP	
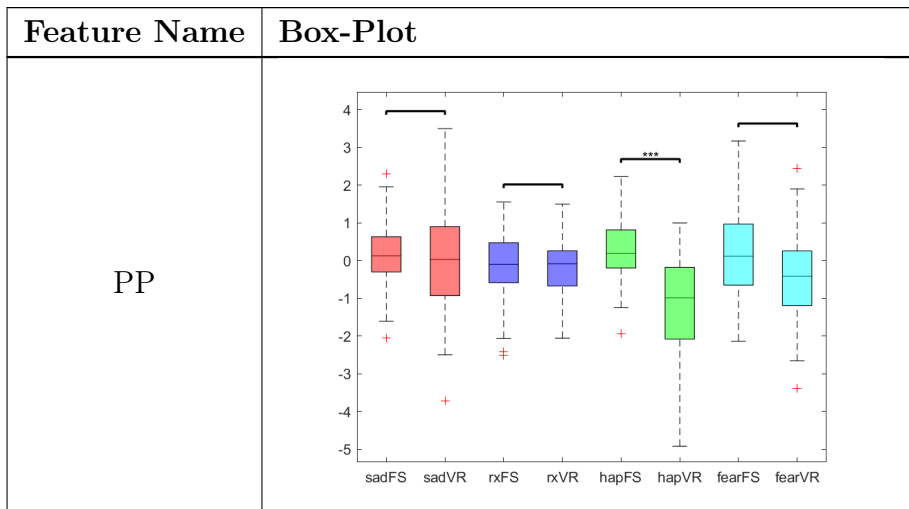
Table 3.16: Statistically significant PP Feature Wilcoxon Test.

3.4. Feature Selection Results

To assess what were the best features capable of distinguishing the four emotional states for each elicitation method the Square Method feature selection approach was chosen. As previously stated in section 2.3.3 the square method gives as output a list of features ranked by their ability to separate the observations based on their emotional nature.

To isolate only the three best features per method a distributions' intersection threshold has been set for each elicitation method. The resulting limit value of the threshold for the FS method was 39% of overlap, placing at the first places in the ranking three Galvanic Skin Response derived features: Avg abs2 normalized, Avg der GSRbp and Slope GSRbp. Using the three selected features a 3D representation of the separation between emotion is represented in the following figure (fig. 3.22), and the three possible two-dimensional views are reported in fig. 3.23, in the following plots each emotion is visible in a different colors, yellow for Sadness, green for Relaxation, blue for Happiness and Red for Fear. Coherently with the results from the Friedman's Test the four emotions are not clearly separated in the 3-feature space.
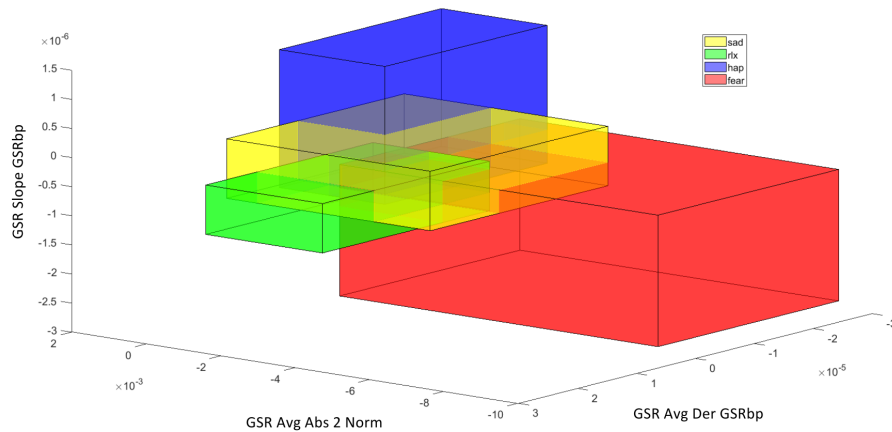
Figure 3.22: 3-dimensional emotion separation Flat Screen stimulation

(a)

(b)

(c)
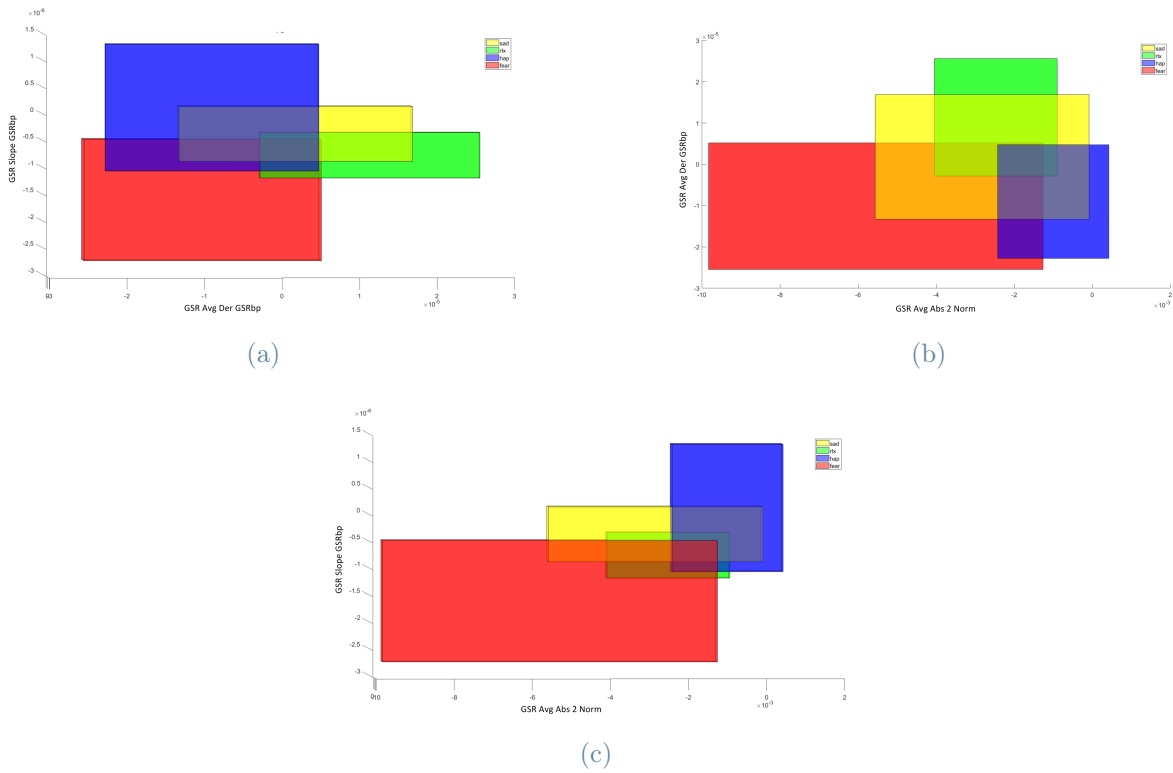
Figure 3.23: 2-dimensional views of emotion separation of Flat Screen stimulation

This representation was also implemented to visualize how the selected features could distinguish high (in red) valence states from low (in blue) valence ones (fig. 3.24) and high arousal states (in red) from low arousal (in blue) ones (fig. 3.25).
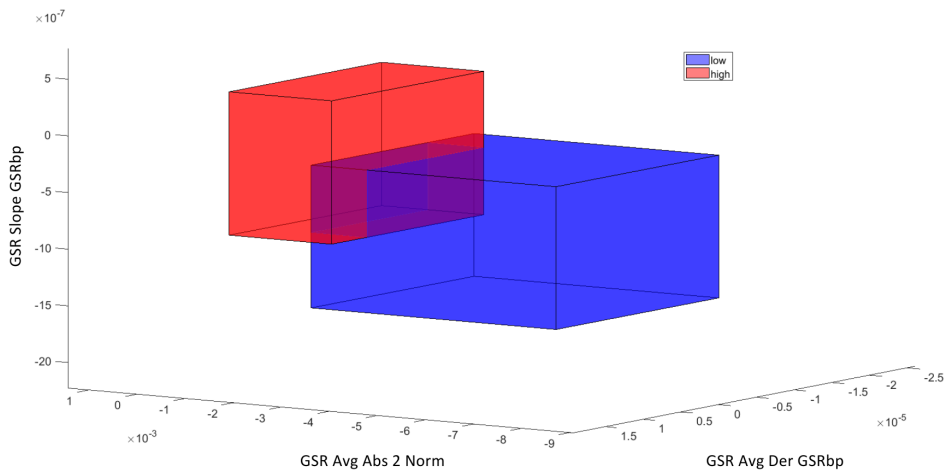
Figure 3.24: 3-dimensional Valence separation Flat Screen stimulation
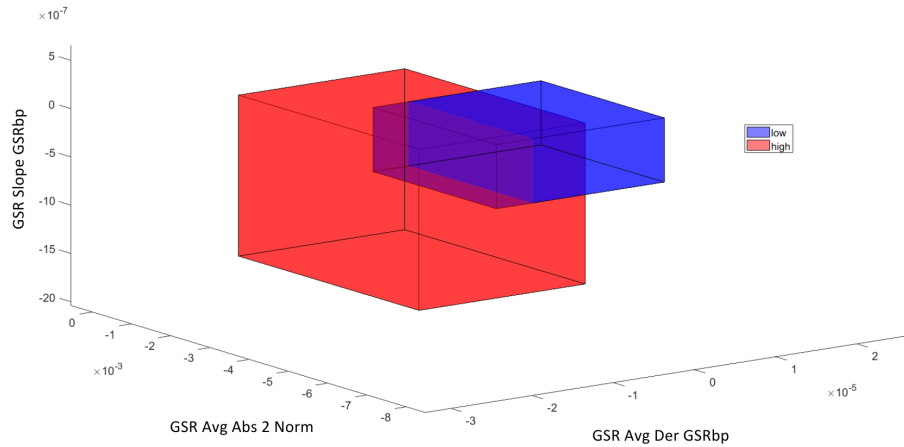
Figure 3.25: 3-dimensional Arousal separation Flat Screen stimulation

As for the VR stimulation, the resulting limit threshold was set at 18.54% of overlap and it isolated four features (third and fourth matching in score, but placed in the ranking following the rules cited in 2.3.3), in rating order: Respiratory frequency, GSR Avg abs2, GSR Average and GSR N peaks.

Using the first three selected features a 3D representation of the separation between emotion is represented in the following figure (fig. 3.26), another 3D space was plotted using the fourth feature (N Peaks) to visually assess the difference between the two matched ranking features (fig. 3.27). In the following plots each emotion is visible in a different colors, yellow for Sadness, green for Relaxation, blue for Happiness and red for Fear.
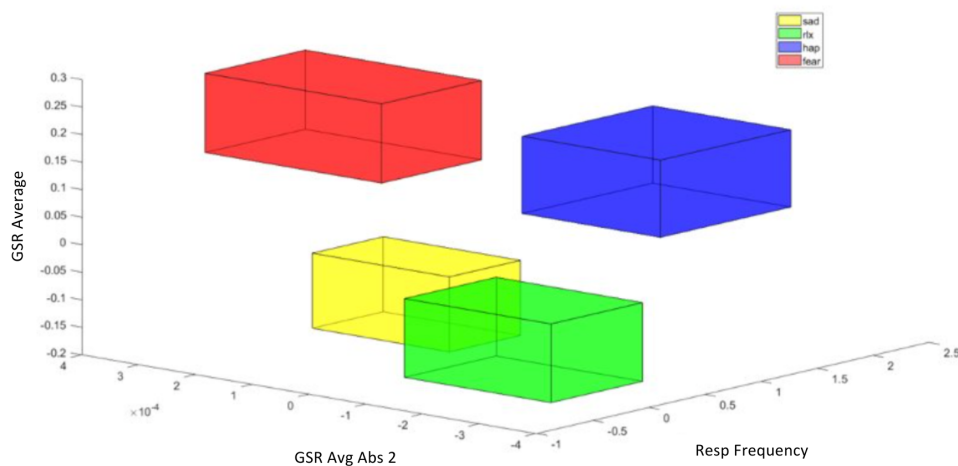
Figure 3.26: 3-dimensional emotion separation of VR stimulation, best three features

The three possible two-dimensional views derived from the best three features are reported

in figure 3.28. As is shown the four elicitated emotional space are separated in the 3-dimensional space defined by the three selected features.
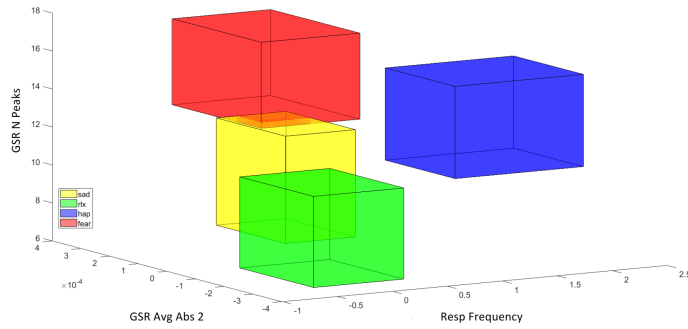
Figure 3.27: 3-dimensional emotion separation of VR stimulation, with N Peaks feature.
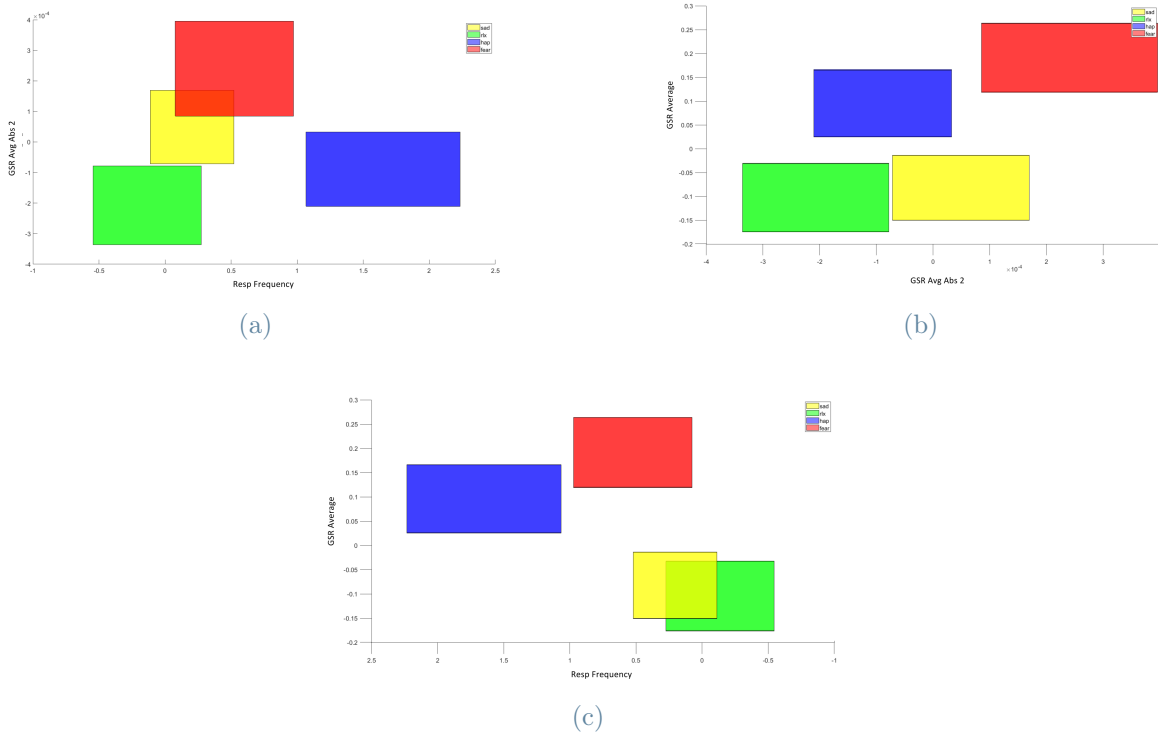
Figure 3.28: 2-dimensional views of emotion separation of VR stimulation

This representation was also implemented to visualize how the selected best three features could distinguish high (in red) valence states from low (in blue) valence ones (fig. 3.29) and high(in red) arousal states from low (in blue) arousal ones (fig. 3.30).
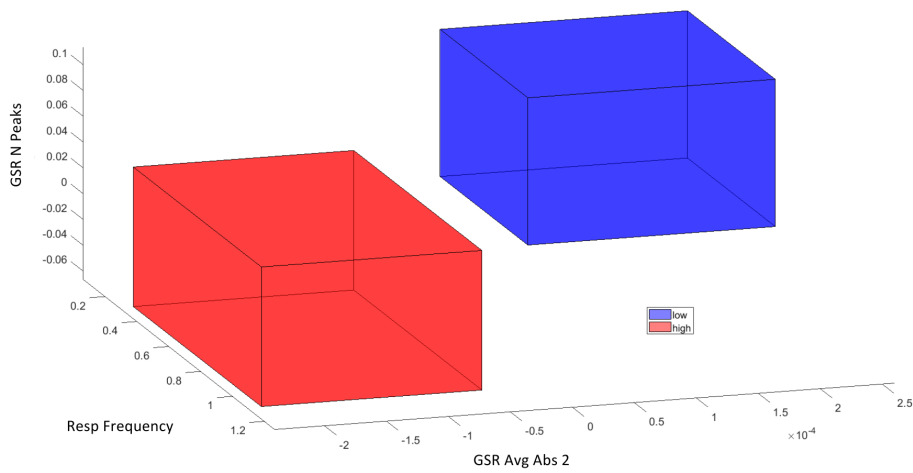
Figure 3.29: 3-dimensional Valence separation VR stimulation
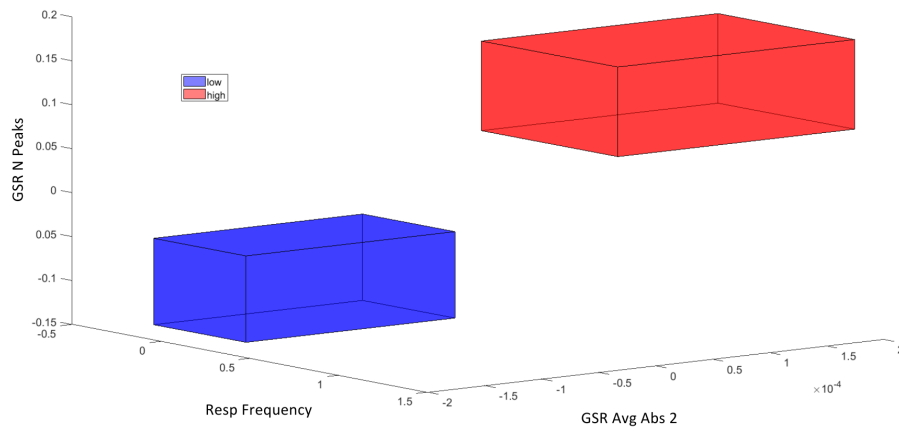
Figure 3.30: 3-dimensional Arousal separation VR stimulation

To better compare the selected features in both elicitation methods the following two 3D plots were created: in each of them it is possible to see the four emotion distribution obtained by using the best three feature selected for the other elicitation modality. In figure 3.31(a) the VR elicitated emotion distribution using the best three feature selected for the FS stimulation can be seen, meanwhile in figure 3.31(b) the FS elicitated emotion distribution using the best three feature selected for the VR stimulation is shown.
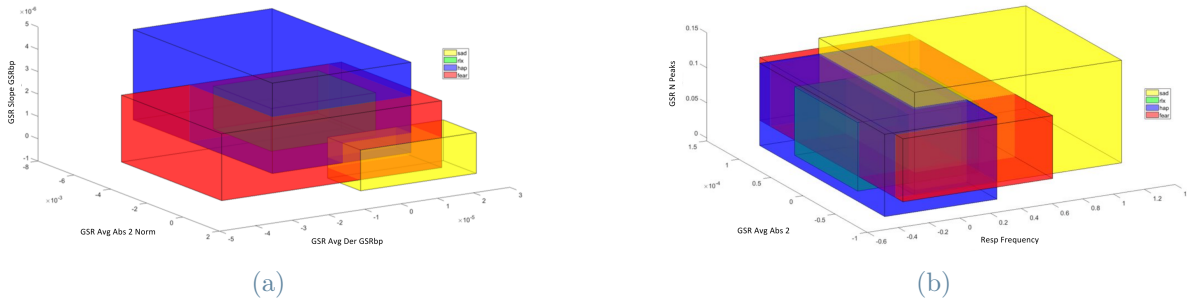
Figure 3.31: 3D emotion separation space, selected features comparison.

3.5.   Machine Learning Classification Results

After conducting a comprehensive evaluation of multiple feature sets derived from the Flat Screen and Virtual Reality results datasets, it was determined that the Square Method was the most effective technique to identify the optimal features.

The selection threshold were raised as necessary to find more than just the three more relevant features listed in 3.4. From the FS part, the top ten features were selected, and the top seven features were retained from the VR part. These feature sets yielded the most beneficial results when employed with a K-Neighbors model. Such sets are:

- **Flat Screen**: GSR: GSR sd amp peaks, GSR max der GSRbp, GSR slope GSRbp, GSR max sign amp, GSR avg der GSRbp, GSR avg abs2 norm; Cardiac: vlf PP, tot PP, PAT; Respiratory: f resp.

- **Virtual Reality**: GSR: GSR avg, GSR n peaks, GSR max der gsrbp, GSR avg abs2; Cardiac: tot (Point process), PP; Respiratory: Respiratory Frequency.

Using these feature sets, the training database was used to train all the models listed in 2.4; their hyperparameters were adjusted in order to obtain the highest accuracy level possible and this was done for each one of the three classification tasks.

Although all models were trained and tested with various combinations of values for their hyperparameters, the best model resulted being K-Neighbors for all classification tasks, for both datasets. It is important to note that these results are supposed to be qualitative and non-final, as the datasets were way too small than what is usually needed to optimally train and test machine learning models. The accuracy level with the selected number of neighbors are reported in the next paragraphs, along with the confusion matrices obtained for all tasks and the ROC curves obtained in the arousal and valence separation tasks (as such graphical method can't be applied in multiclass tasks).

3.5.1.   Results of Classification of Flat Screen data

Starting from the first task, which was the multiclass classification in which the models tried to separate all 4 emotional states, the best accuracy was of 25% (using 9 neighbors), identical to the dataset proportion for each class. The resulting classification matrix can be seen in figure 3.32.
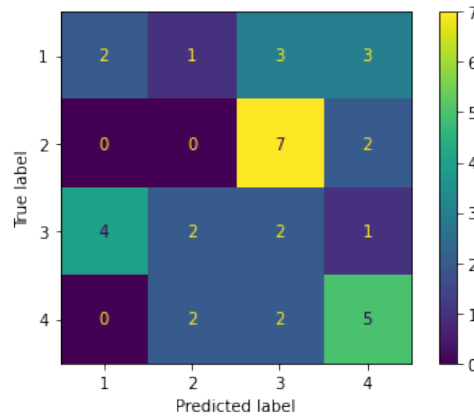
Figure 3.32: Classification matrix of multiclass classification for FS data.

(a)   (b)
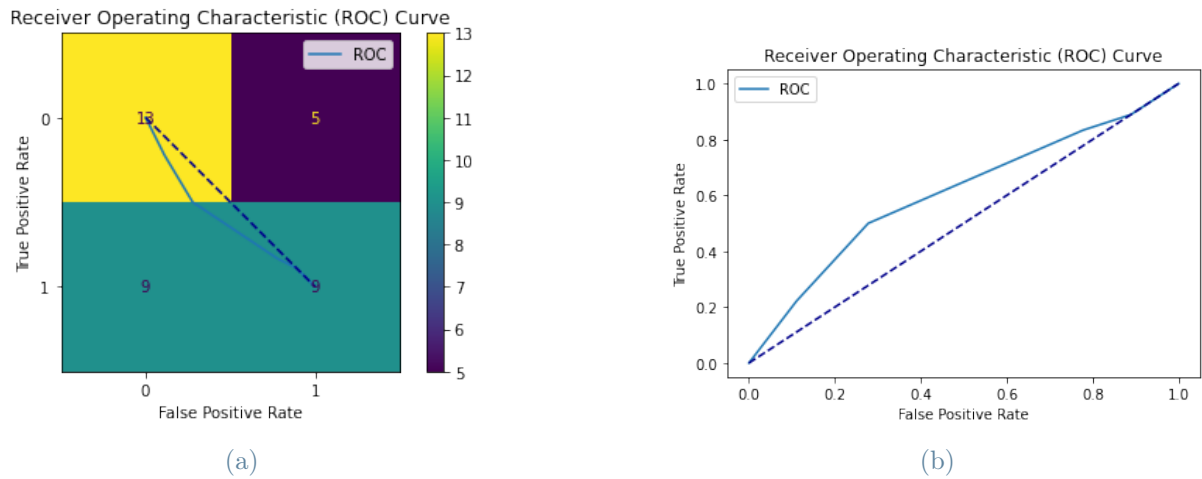
Figure 3.33: Classification matrix (a) and ROC curve (b) of the Arousal classification task for FS data

As for the second task, which was classification of the level of arousal, the maximum accuracy obtained was of 61.1%, obtained using 5 neighbors (fig. 3.33) and the obtained AUC value was 0.61.

The third task was classification of the level of valence, for which the best accuracy was of 47.2% using 13 neighbors (fig. 3.34) and the obtained AUC value was 0.47.

These results highlighted a difficulty in separating valence levels from physiological signals, while arousal separation is easier.

(a)   (b)
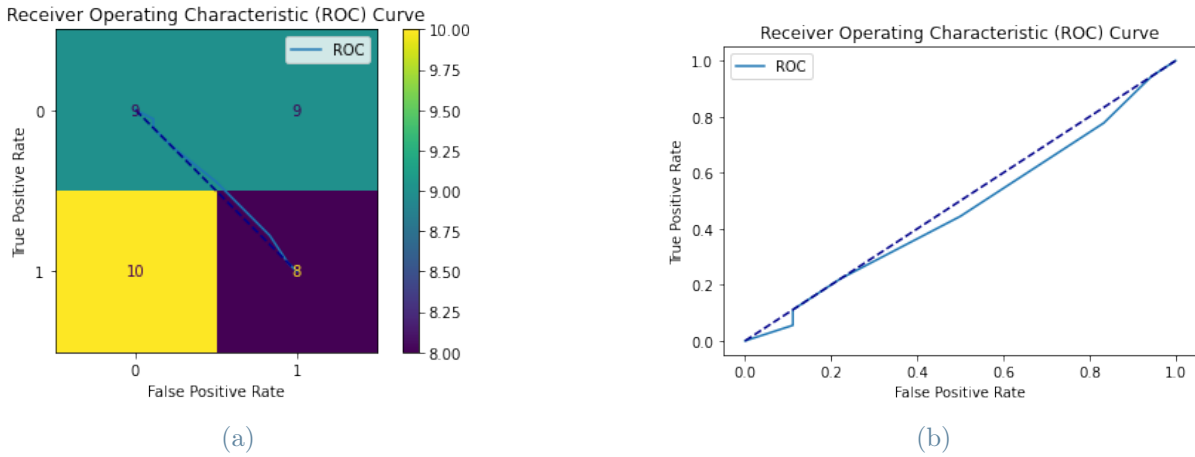
Figure 3.34: Classification matrix (a) and ROC curve (b) of the Valence classification task for FS data.

3.5.2.   Results of Classification of Virtual Reality data

The best K-Neighbors model for the multiclass separation achieved provided an accuracy of 48.5% and considered 9 neighbors to classify the data. The resulting classification matrix can be seen in figure 3.35.
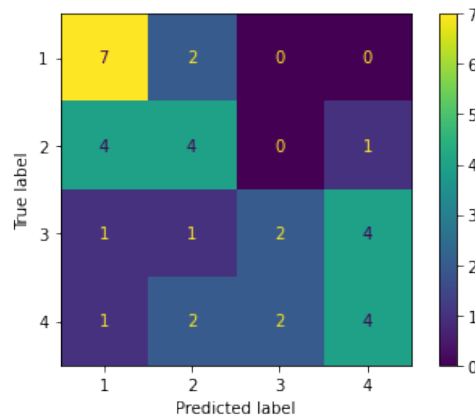
Figure 3.35: Classification matrix of multiclass classification for VR data.

As for the separation of the arousal level, K-Neighbors achieved an accuracy of 80% (9 neighbors), the highest of all tasks for both datasets (fig. 3.36) and the obtained AUC value was 0.87.
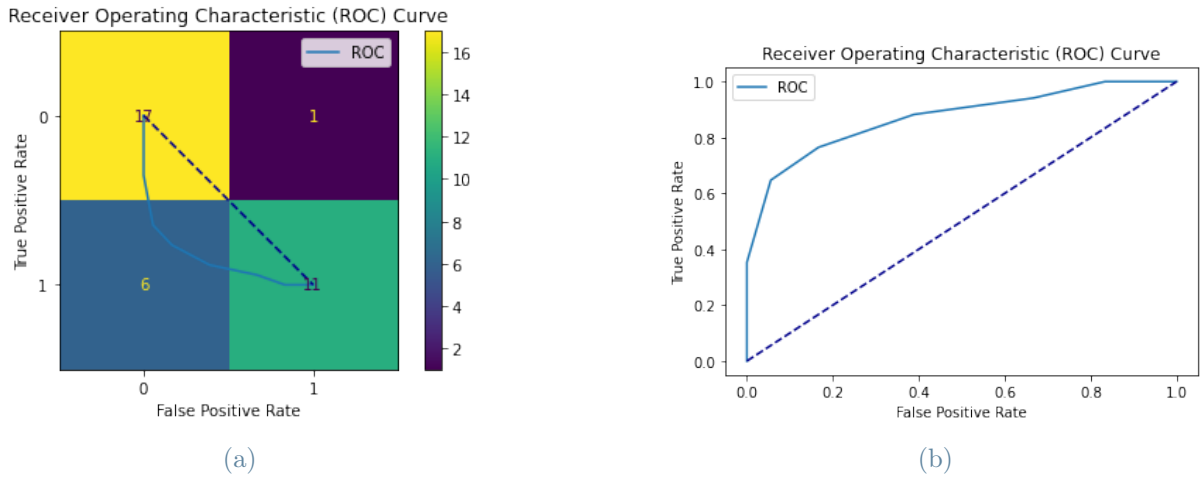
Figure 3.36: Classification matrix (a) and ROC curve (b) of the Arousal classification task for VR data

Finally, the third task was separating the valence level. The highest obtained accuracy was of 65.7% (4 neighbors) and the obtained AUC value was 0.69, confirming the observation made in the previous paragraph (fig. 3.37).
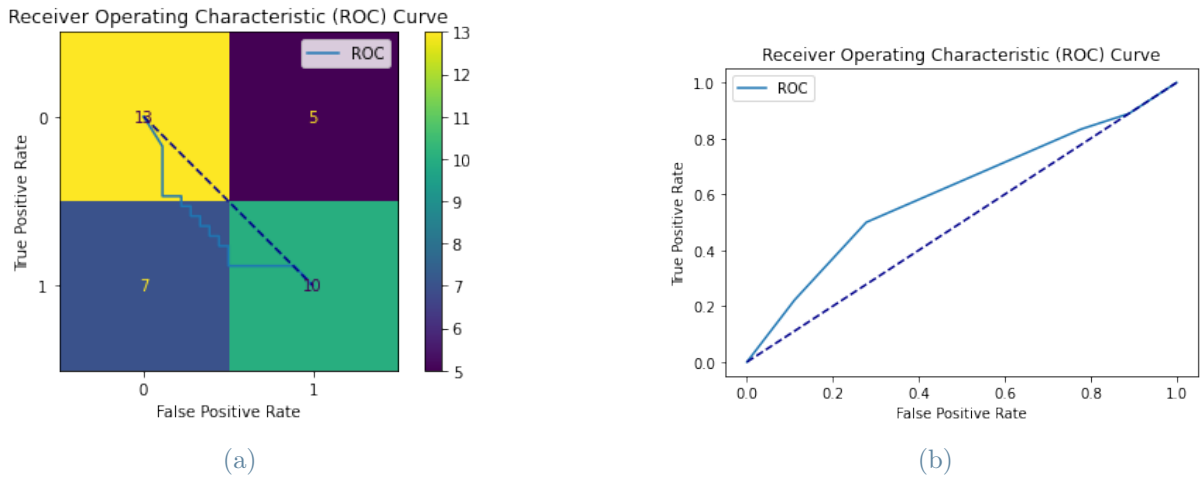
Figure 3.37: Classification matrix (a) and ROC curve (b) of the Valence classification task for VR data

3.6. Summary of Results

In this section the results of the thesis were reported.

- **Questionnaire's Results**: SAM and PAM results and their comparison (among FS data and OASIS values for FS, and among VR and FS data for VR) were

shown and evaluated. While the flat screen method evoked responses with lower arousal than expected, data coming from the post-VR questionnaire indicate that VR doesn't have this limitation; the elicited responses seem to be more centered in the respective quadrant of the Circumplex Plane.

- **Physiological Response Results**: the median trends of a selection of features was reported. The trends were coherent with the expected response and showed that VR-induced variations in the features were generally more pronounced than the FS-induced ones, and that the VR stimulation generated a higher level of arousal in the subjects.

- **Statistical Analysis**: the results of the Friedman test and the Multcompare test with Bonferroni correction were reported. These tests identified 17 features that significantly differed in specific comparisons between two different emotive states. Notably, all these features were from the VR dataset, while none were relevant in the flat screen dataset. The study also reported the results of the Wilcoxon test.

- **Feature Selection Results**: using different thresholds for the FS and VR datasets, the Square method was applied in order to find the best three features for each elicitating method that are capable of separating the 4 emotional states. The results were plotted in a 3-dimensional graph, and also their separation capacity for arousal level and valence level was reported. The feature set identified for VR separates the 4 emotional states in a more clear and definite way than the FS feature set (figs. 3.22 and 3.26).

- **Machine Learning Classification Results**: after having identified a different set of best features for FS and VR, found applying the Square method to find more than just three features, they were used to train a number of machine learning models. The best one, for all the three tasks, was always a K-Neighbors. For both datasets, the most accurate classification task was the one based on arousal level. VR trained models always performed better than FS trained ones.

4 | Discussion and Conclusions

4.1. Discussion

This work explores the integration of Virtual Reality and Affective Computing using rigorous quantitative methods, such as advanced signal processing, to extract and analyse high-quality features. A novel experimental protocol was developed in a fully immersive virtual environment to test the validity of the VR technology as medium for targeted emotional triggers, against an already validated picture-based method for emotion elicitation.

One of the main hypothesis that were formulated in the early stages of this work is that the traditional methods for laboratory-based emotion elicitation are to be considered outdated. The quantity and variety of stimuli that the modern occidental human being takes as input every day, from an ever increasing quantity of devices, mediums and content creators, could be limiting our responses to non-immersive emotional stimulations.
In this thesis work both self-reported and physiological data have been acquired from 49 subjects, 5 of this acquisitions were later discarded.
For the flat screen stimulation part of the experiment, pictures taken from the OASIS database were used; this database was chosen instead of the more widely used IAPS for its more contemporary selection of images, up-to-date ratings and for its free access.

The average perceived protocol duration is extremely on point for the FS stimulation section of the protocol and about 5 minutes less than the actual duration for the VR elicitation part; this suggests a lower level of boredom for the VR stimulation, despite it being longer than the FS section, and possible application of VR stimulation for longer protocols.

The collected results from the post FS stimulation assessment show that in no case, among the 4 different emotional parts, the mean of the expected value of valence-arousal is even inside of the 95% confidence interval, built with the self-reported data of the subjects, mirroring the evaluating process of the OASIS database itself (fig. 3.20). In the specific case of happiness, not only the evoked response is far from the expected, but even falls

into the quadrant associated with lower levels of arousal, that identifies emotional states such as Relaxation and Contentment. This could hint to a difficulty in inducing high arousal states, except if the selected pictures have graphic content, like in the case of the fear stimulation, which elicited an on-point response.

While the Flat Screen method evoked responses with lower arousal than expected, PAM data coming from the post-VR questionnaire indicate that VR doesn't have this kind of limitation. As also the median plots of many physiological features indicate, VR elicitating parts were generally more arousing. As stated in [3], arousal level and the perception of being present in a virtual environment are correlated. Thus, a slightly higher arousal may be linked with the immersiveness of the VR environments and events, and not just to the specific content. Indeed, looking at Table 3.1 it is clear that the VR valence-arousal plots on the Circumplex plane are more centered in their targeted emotional quadrant then their FS counterpart.

Physiologically, the extracted features (both in time and frequency domain) are consistent with the literature. All used signals (HRV, GSR, BVP and RESP) follow the expected trends of ANS activation in respect with the target emotion.
Signals acquired during the VR stimulation are still coherent with the expected trends, but show a higher variability than the ones induced by the FS method. This variability is even more accentuated in those features that are directly linked with arousal, like those derived from the GSR.

Friedman's test, followed by a post-hoc with Bonferroni correction, revealed that no features filled by data acquired during the FS stimulation were useful in separating emotional states.
However, for VR derived data, these tests identified a set of 17 features that are significantly capable of separating specific emotional states. These features are particularly effective at distinguishing between Fear and Relaxation, which are opposite emotions and achieved the highest distance on the CMA graph that displayed the questionnaire's SAM data (fig. 3.20). Also Fear and Sadness resulted separable by 6 features, highlighting that a separation by arousal level may be generally more simple, especially after a VR stimulation that, as mentioned before, may induce a more aroused and variable response in the subjects. This insight was then confirmed by analysing the machine learning's outcomes, in which binary arousal-based classification achieved the highest accuracy among all three tasks for both datasets.
Binary classification based on valence had lower accuracy than the arousal-level separation, and again this result was already predicted by the Friedman's post-hoc test, where valence level was significantly identified only by 4 features.

As for the performed multiclass classification keeping the four elicitated emotions as targets, the small dimensions of the dataset have to be taken into consideration, only 35 instances for each target class were available to train the model. Nonetheless the outcomes of the VR trained model have an accuracy score of 48.5% against an accuracy of 25% of the FS derived model, that is the limit value for random classification for four target classes.

4.2. Main Innovations

The main innovations of this work can be summarized by the following points:

- This thesis proposes a novel Virtual Reality experimental framework for targeted emotion elicitation, validating the use of this new technology in the field of Affective Computing.

- One of the key innovations is that this study directly compares a traditional emotion elicitation method and a Virtual Reality-based one, through a comprehensive approach that uses coherent stimuli, subjective assessment, advanced signal processing, and machine learning.

- The environments used in this study for the VR stimulation were designed to create lifelike scenarios by implementing audio and visual cues, both static and dynamic (timed event-related), rather than relying on single isolated stimuli as vastly done in literature. This approach aims to provide a more immersive and realistic experience for the subjects, which may elicits specific emotional responses.

- It's one of the first times that, in the field of Affective Computing, Point-Process algorithms are used for the extraction of cardiac features over time.

- This study introduces a novel definition of physiological feature sets that can differentiate not only between high and low arousal or valence, but also between specific pairs of emotional states among Sadness, Relax, Happiness and Fear. This insight will be extremely useful in building emotional classifiers.

4.3. Limitations and Future Developments

The study presents a preliminary method for emotion elicitation using a VR device, but acknowledges a limitation in the potential effect of sound on the subjects' response during VR stimulation. As sound wasn't present in the Flat Screen part, it is possible that at least part of the subjects' response was modified by its presence in the VR method. Aware

of this bias, the researchers considered two alternatives: adding sound in the FS section, or removing it from the VR part. The first option was rejected because of the limited data in literature on emotion elicitation using images and sounds simultaneously. After evaluation, the second possibility was also rejected, as a soundless virtual environment could have reduced the desired feeling of presence, thereby limiting the full potential of the chosen stimulation medium. A possible area of future studies could be introducing sound stimulation in the Flat screen elicitation portion of the protocol, to better assess the effect of sound in addition to the visual stimuli. Also a new comparison between a real life scenario and a VR immersive environment replicating it could be studied.

Another limitation of this study stands in the number of subjects the protocol was experimented on. The statistical analysis, feature extraction and classification modeling were based on data derived from 44 subjects only. More experiments and data are needed in order to quantitatively assess the capabilities, applications and limitations of the proposed protocol. Accordingly, more acquisition phases and comprehensive studies have already been planned.

4.4.   Conclusions

This study aimed to develop new VR-based methods for affective computing to elicit, identify, and distinguish various emotional states compared to the traditional Flat Screen elicitation method that uses validated pictures as stimuli. The entire protocol was developed using Unity.

To conduct both qualitative and quantitative analyses, self-report data from post-elicitation questionnaires and physiological signals were collected. Physiological signals, including ECG, PPG, GSR, and Respiration, were acquired using a ProComp device. Different features were extracted from the signals, and in-depth analyses were performed. Statistical analysis was carried out to evaluate the effects of emotions on physiological indexes calculated from the signals. Machine learning algorithms were employed to classify emotions based on their level of arousal, valence, and position in the four quadrants of the CMA plane.

Although further studies are needed to confirm and expand on the obtained insights, this thesis demonstrates the efficacy of Virtual Reality as an emotional elicitation tool and its potential in enhancing the field of Affective Computing. This claim is justified by the more pronounced amplitudes and higher variability of the VR-extracted features, that better explain the changes in the arousal of the subjects and, on a smaller scale, also in in their valence. In conclusion, this work proposes a novel Virtual Reality experimental

framework for targeted emotion elicitation; this innovative approach proved its ability to better elicit specific emotional responses in the subjects, compared with a traditional Flat Screen picture-based method. VR provides a more immersive and realistic experience for the subjects that is sure to enhance the field of Affective Computing, opening up exciting new opportunities.

Bibliography

[1] Rafael Calvo, Jacqueline Kory, and Sidney D Mello. Affect Elicitation for Affective Computing. 2014.

[2] Anna Felnhofer, Oswald David Kothgassner, Mareike Schmidt, Anna-Katharina Heinzle, Leon Beutl, Helmut Hlavacs, and Ilse Kryspin-Exner. Is virtual reality emotionally arousing? investigating five emotion inducing virtual park scenarios. *Int. J. Hum. Comput. Stud.*, 82:48–56, 2015.

[3] Javier Marín-Morales, Carmen Llinares, Jaime Guixeres, and Mariano Alcañiz. Emotion recognition in immersive virtual reality: From statistics to affective computing. *Sensors (Switzerland)*, 20(18):1–26, 2020. ISSN 14248220. doi: 10.3390/s20185163.

[4] Javier Marín-Morales, Juan Luis Higuera-Trujillo, Alberto Greco, Jaime Guixeres, Carmen Llinares, Claudio Gentili, Enzo Pasquale Scilingo, Mariano Alcañiz, and Gaetano Valenza. Real vs. immersive-virtual emotional experience: Analysis of psychophysiological patterns in a free exploration of an art museum. *PLoS ONE*, 14, 2019.

[5] Rukshani Somarathna, Tomasz Bednarz, and Gelareh Mohammadi. Virtual Reality for Emotion Elicitation-A Review. Technical report.

[6] Rosalind W Picard, Singapore-mit Alliance, Walter Bender, Cynthia Breazeal, and David Cavallo. Affective learning — a manifesto. (October), 2004. doi: 10.1023/B.

[7] Rafael A. Calvo and Sidney D'Mello. Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing*, 1(1):18–37, 2010. ISSN 19493045. doi: 10.1109/T-AFFC.2010.1.

[8] Abhishek Chunawale and Manngesh Bedekar. Human Emotion Recognition using Physiological Signals: A Survey. *SSRN Electronic Journal*, 2020. doi: 10.2139/ssrn. 3645402.

[9] Laurie Kelly McCorry. Physiology of the Autonomic Nervous System. *Acta Anaesthesiologica Scandinavica*, 8(4):17–20, 1964. ISSN 13996576. doi: 10.1111/j.1399-6576. 1964.tb00252.x.

[10] Sylvia D. Kreibig. Autonomic nervous system activity in emotion: A review. *Biological Psychology*, 84(3):394–421, 2010. ISSN 03010511. doi: 10.1016/j.biopsycho.2010.03.010.

[11] Lisa Feldman Barrett. *How emotions are made: The secret life of the brain*. Picador, 2020.

[12] James A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178, 1980. ISSN 00223514. doi: 10.1037/h0077714.

[13] Albert Mehrabian and James A Russell. *An approach to environmental psychology*. The MIT Press, Cambridge, MA, US, 1974. ISBN 0262130904.

[14] P. M.A. Desmet, M. H. Vastenburg, D. Van Bel, and N. Romero. Pick-A-Mood development and application of a pictorial mood-reporting instrument. *8th International Conference on Design and Emotion: Out of Control - Proceedings*, (May 2014), 2012.

[15] Alice M. Isen, Barbara Means, and Robert Patrick. Some factors influencing decision-making strategy and risk taking. 1974.

[16] Margaret M. Bradley and Peter J. Lang. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1):49–59, 1994. ISSN 00057916. doi: 10.1016/0005-7916(94)90063-9.

[17] Rensis Likert. A technique for the measurement of attitude scales. 1932.

[18] Peter J Lang, Margaret M Bradley, and Bruce N Cuthbert. International affective picture system (IAPS): Technical manual and affective ratings. *NIMH Center for the Study of Emotion and Attention*, pages 39–58, 1997.

[19] Benedek Kurdi, Shayn Lozano, and Mahzarin R. Banaji. Introducing the Open Affective Standardized Image Set (OASIS). *Behavior Research Methods*, 49(2):457–470, 2017. ISSN 15543528. doi: 10.3758/s13428-016-0715-3.

[20] Ben Meuleman and David Rudrauf. Induction and profiling of strong multi-componential emotions in virtual reality. *IEEE Transactions on Affective Computing*, 12:189–202, 2018.

[21] Regina Kaplan-Rakowski, Karen R. Johnson, and Tomasz Wojdynski. The impact of virtual reality meditation on college students' exam performance. *Smart Learning Environments*, 8(1), 2021. doi: 10.1186/s40561-021-00166-7.

[22] Daniel R. Mestre and Jean-Louis Vercher. Immersion and presence. 2011.

[23] Muhammad Anas Hasnul, Nor Azlina Ab Aziz, Salem Alelyani, Mohamed Mohana, and Azlan Abd Aziz. *Electrocardiogram-based emotion recognition systems and their applications in healthcare—a review*, volume 21. 2021. ISBN 1141126389. doi: 10.3390/s21155015.

[24] Mohammad Ghamari. A review on wearable photoplethysmography sensors and their potential future applications in health care. *International Journal of Biosensors Bioelectronics*, 4(4):195–202, 2018. ISSN 2573-2838. doi: 10.15406/ijbsbe.2018.04.00125.

[25] Mathias Benedek and Christian Kaernbach. A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods*, 190(1):80–91, 2010. ISSN 01650270. doi: 10.1016/j.jneumeth.2010.04.028.

[26] Ravinder Jerath and Connor Beveridge. Respiratory Rhythm, Autonomic Modulation, and the Spectrum of Emotions: The Future of Emotion Recognition and Modulation. *Frontiers in Psychology*, 11(August):1–7, 2020. ISSN 16641078. doi: 10.3389/fpsyg.2020.01980.

[27] Frans A. Boiten. The effects of emotional behaviour on components of the respiratory cycle. *Biological Psychology*, 49(1-2):29–51, 1998. ISSN 03010511. doi: 10.1016/S0301-0511(98)00025-8.

[28] Qiang Zhang, Xianxiang Chen, Qingyuan Zhan, Ting Yang, and Shanhong Xia. Respiration-based emotion recognition with deep learning. *Computers in Industry*, 92-93:84–90, 2017. ISSN 01663615. doi: 10.1016/j.compind.2017.04.005.

[29] Marcus Toftedahl and Henrik Engström. A taxonomy of game engines and the tools that drive the industry. In *DiGRA Conference*, 2019.

[30] George Trumbull Ladd. Direct control of the retinal field. *Psychological Review*, 1(4):351–355, 1894. doi: 10.1037/h0068980.

[31] Andres Pinilla, Jaime Garcia, William Raffe, Jan-Niklas Voigt-Antons, Robert P. Spang, and Sebastian Möller. Affective visualization in virtual reality: An integrative review. *Frontiers in Virtual Reality*, 2, 2021. doi: 10.3389/frvir.2021.630731.

[32] Alberto Valdes Rey. A novel affective computing frame- work based on VR elicitation and advanced signal processing emotion recognition assessment., 2022.

[33] Eun Yeong Choe, Anna Jorgensen, and David Sheffield. Simulated natural environments bolster the effectiveness of a mindfulness programme: A comparison with

a relaxation-based intervention. *Journal of Environmental Psychology*, 67:101382, 2020. ISSN 0272-4944. doi: https://doi.org/10.1016/j.jenvp.2019.101382.

[34] Xiao-Ping Gao, John H. Xin, Tetsuya Sato, Aran Hansuebsai, Marcello Scalzo, Kanji Kajiwara, Shing-Sheng Guan, J. Valldeperas, Manuel José Lis, Monica Billger, and et al. Analysis of cross-cultural color emotion. *Color Research amp; Application*, 32 (3):223–229, 2007. doi: 10.1002/col.20321.

[35] Konstantinos Drossos, Andreas Floros, Andreas Giannakoulopoulos, and Nikolaos Kanellopoulos. Investigating the impact of sound angular position on the listener affective state. *IEEE Transactions on Affective Computing*, 6(1):27–42, 2015. doi: 10.1109/taffc.2015.2392768.

[36] Jorn Bakker, Mykola Pechenizkiy, and Natalia Sidorova. What's your current stress level? Detection of stress patterns from GSR sensor data. *Proceedings - IEEE International Conference on Data Mining, ICDM*, (1):573–580, 2011. ISSN 15504786. doi: 10.1109/ICDMW.2011.178.

[37] Maximiliano Mollura, Edoardo Maria Polo, Li-wei Lehman, and Riccardo Barbieri. Assessment of heart rate variability derived from blood pressure pulse recordings in intensive care unit patients. *2020 Computing in Cardiology Conference (CinC)*, 2020. doi: 10.22489/cinc.2020.423.

[38] Maximiliano Mollura, Stefano Romano, Giulio Mantoan, Li-wei Lehman, and Riccardo Barbieri. Prediction of septic shock onset in icu by instantaneous monitoring of vital signs. *2020 42nd Annual International Conference of the IEEE Engineering in Medicine amp; Biology Society (EMBC)*, 2020. doi: 10.1109/embc44109.2020. 9176276.

[39] Maximiliano Mollura, Li-Wei H. Lehman, Roger G. Mark, and Riccardo Barbieri. A novel artificial intelligence based intensive care unit monitoring system: Using physiological waveforms to identify sepsis. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379(2212), 2021. doi: 10.1098/rsta.2020.0252.

[40] D. J. Daley and D. Vere-Jones. Basic properties of the poisson process. *Springer Series in Statistics*, page 18–37, 1998. doi: 10.1007/978-1-4757-2001-3_2.

[41] Riccardo Barbieri, Eric C. Matten, Abdul Rasheed A. Alabi, and Emery N. Brown. A point-process model of human heartbeat intervals: new definitions of heart rate and heart rate variability. *American journal of physiology. Heart and circulatory physiology*, 288, 1 2005. ISSN 0363-6135. doi: 10.1152/AJPHEART.00482.2003.

[42] Edoardo Maria Polo, Maximiliano Mollura, Marco Zanet, Marta Lenatti, Alessia Paglialonga, and Riccardo Barbieri. Analysis of the Effect of Emotion Elicitation on the Cardiovascular System. *Computing in Cardiology*, 2021-September:2021–2024, 2021. ISSN 2325887X. doi: 10.23919/CinC53138.2021.9662859.

[43] Sunil Gupta, Kamal Saluja, Ankur Goyal, Amit Vajpayee, and Vipin Tiwari. Comparing the performance of machine learning algorithms using estimated accuracy. *Measurement: Sensors*, 24:100432, 2022. ISSN 2665-9174. doi: https://doi.org/10.1016/j.measen.2022.100432.

[44] Trevor J. Hastie, Robert Tibshirani, and Jerome H. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd Edition*. 2005.

List of Figures

List of Tables