Politecnico di Torino

Master's Degree in Biomedical Engineering

A.a. 2025/2026

Real-Time Stress Induction and Mitigation Through Adaptive Multisensory Stimulation in Virtual Reality: A Closed-Loop ECG-Based Approach

Relators:

Prof. Luca Mesin
Prof. Matteo Raggi

Candidate:

Giovanni Talarico

March 2026

List of Figures

Abstract

The exposure to chronic stress has been identified as a key risk factor underlying mental disorders, cognitive impairments, and complex pathologies. In this thesis, a closed-loop stress mitigation system using virtual reality multisensory stimulation and electrocardiography is designed and validated. A racing game was developed using Unity to induce controlled stress through competitive gameplay and a rapidly flashing trajectory. Twenty participants aged 24 ± 2 years were divided into two groups, comparing constant stimulation (Play 1) versus adaptive stimulation (Play 2), with order counterbalancing to eliminate learning effects. ECG signals were acquired with a Polar H10 chest strap and processed in real-time via a TCP/IP connection between Python and MATLAB. Features extraction used windows of 4-second length with 75% overlap from which the instantaneous heart rate was calculated in beats per minute (BPM), standard heart rate variability measures (SDNN, RMSSD, pNN50), and some ECG morphological features like RTr, RPr, QTc.

A Random Forest regressor using 100 trees with 10-fold cross-validation for the baseline and MAT phases was trained, resulting in a median R^2 value of 0.7. The model allowed for the quantification of stress, performed in real time, facilitating the adaptation mechanism for Play 2. The stimulation was modulated at theta rhythms, ranging from 4–8 Hz, initiated at 6 Hz, and varied by 1.2 Hz based on 8-second averages, taking into account reductions following gradient reversals.

Results showed that the effect of adaptive stimulation (Play 2) was significant in reducing stress levels compared to constant stimulation (Play 1). Statistical analysis, specifically Friedman and Wilcoxon tests, showed significant differences in subjective responses such as NASA-TLX and VAS scales, as well as objective parameters such as stress load, mean stress, and BPM. Seventeen out of twenty participants reported that stress levels were higher in the Play 1 scenario than in the Play 2 scenario. Furthermore, the stress load was in the range of 55–65% in Play 1 and 20–45% in the Play 2. By employing the Spearman correlation test, the stimulation effectiveness was established as depending on the model R^2, while it did not depend on the gaming skill level.

The study comes forth with a proof of concept of real-time, non-invasive stress mitigation supported by closed-loop adaptive stimulation in VR environments. The proposed system architecture is seen to effectively combine physiological monitoring with dynamic feedback, with possible clinical applications for anxiety disorders, PTSD, and stress-related disorders. Features to be included in future developments include increasing samples, other physiological parameters like EEG and EDA, and more scenarios of VR environments.

Contents

1 Introduction

1.1 What is stress?

Stress is a term that in engineering is used to indicate the strain to which materials are subjected; following this analogy, the Austrian neuroendocrinologist Hans Selye first introduced the term *stress* in 1936 to describe the tensions to which human beings are exposed. Stress can be classified according to perception, duration, and source of the stimulus [1].

1.1.1 Classification based on perception

Eustress (positive stress): It represents a stimulating form of stress. It makes individuals feel energized, productive, and optimistic, helping the body achieve a new state of equilibrium. It is associated with pleasant challenges that enhance cardiovascular health, physical endurance, and cognitive function.

Distress (negative stress): It occurs when an event is perceived as negative and lacking compensation or reward. It leads to feelings of frustration, dissatisfaction, and reduced well-being.

1.1.2 Classification based on duration

Acute stress: It is an intense, unexpected, and short-term reaction. It activates the immediate fight-or-flight response, leading the psychophysical system to temporarily increase psychomotor performance through the release of adrenaline and noradrenaline, in order to ensure survival.

Chronic stress: It is a long-lasting condition that persists over time. Although the organism may develop a certain degree of habituation, continuous exposure can exceed adaptive capacity, resulting in functional imbalances and increasing the risk of cardiovascular diseases, anxiety, and depression.

1.1.3 Classification based on the source of the stimulus

Psychological stress: It originates in cognitive/emotional responses, for example, perceiving a threat or high personal expectations, or academic demands.

Physiological stress: Represents the organism's response to internal or external factors that change normal physical homeostasis such as illness, injury, poor nutrition, or inadequate sleep.

Traumatic stress: It is the result of being exposed to catastrophic events which lie beyond what a person is capable of adapting to, possibly leading to to Post-Traumatic Stress Disorder (PTSD).

Environmental stress: Is associated with unfavorable living conditions, including excessive noise, pollution, overcrowding, or unsafe environments.

1.2 Nervous system under stress conditions

The part of the nervous system involved in the stress response is the Autonomic Nervous System (ANS), which is divided into three branches: Sympathetic (SNS), Parasympathetic (SNP), and Enteric systems. The sympathetic system is responsible for "fight or flight" responses, increasing heart rate (HR) and blood pressure. The parasympathetic system, on the other hand, promotes "rest and digest" functions, reducing heart rate and facilitating energy recovery through metabolic conservation. The enteric system is partially autonomous and is influenced by which of the sympathetic or parasympathetic systems is dominant. If we go more in detail stress response has 2 components one is slow and it's controlled by the HPA axis and the other,fast, by the SAM axis[2][3].

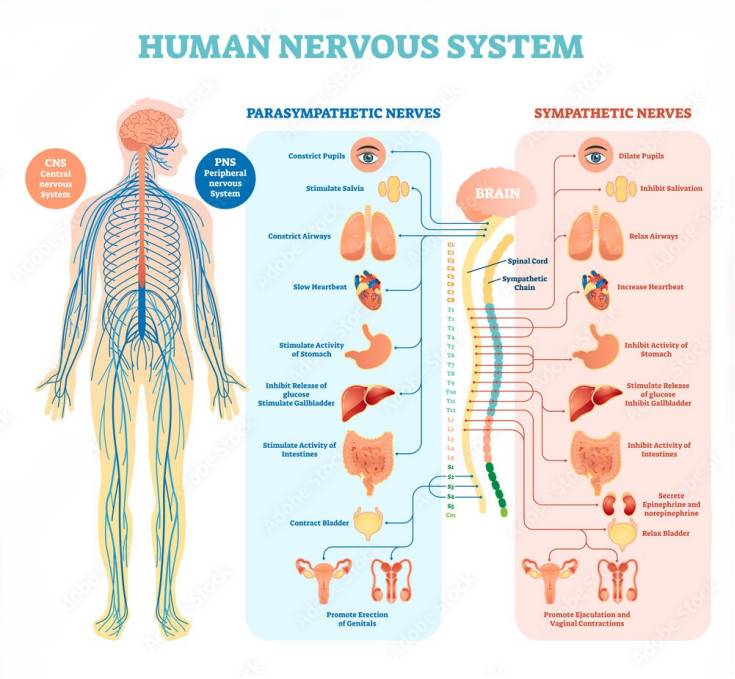


Figure 1: Comparison between parasympathetic and sympathetic systems

1.2.1 HPA Axis

The Hypothalamic–Pituitary–Adrenal axis, as already mentioned before, constitutes the slower response which involves more prolonged effects that can last hours or days. The amygdala, when a stressful event occurs, stimulates

the hypothalamus; the paraventricular nucleus (PVN) synthesizes and releases corticotropin-releasing hormone (CRH). CRH reaches the anterior pituitary, stimulating the release of ACTH (adrenocorticotropic hormone) into the blood. ACTH acts on the adrenal cortex, inducing the release of cortisol.

1.2.2 SAM Axis

The Sympathetic–Adreno–Medullary axis, as said before, represents the rapid stress response, activating within a few seconds after the perception of a stimulus. In this case, the amygdala, following the stressful event, especially if threatening, activates the hypothalamus and the brainstem, in particular regions such as the Locus Coeruleus (LC) and the Paraventricular Nucleus (PVN).[4] These nuclei are important because they are directly connected to sympathetic neurons of the spinal cord, which directly stimulate the cells of the adrenal medulla, leading to the release of catecholamines, such as adrenaline and noradrenaline.

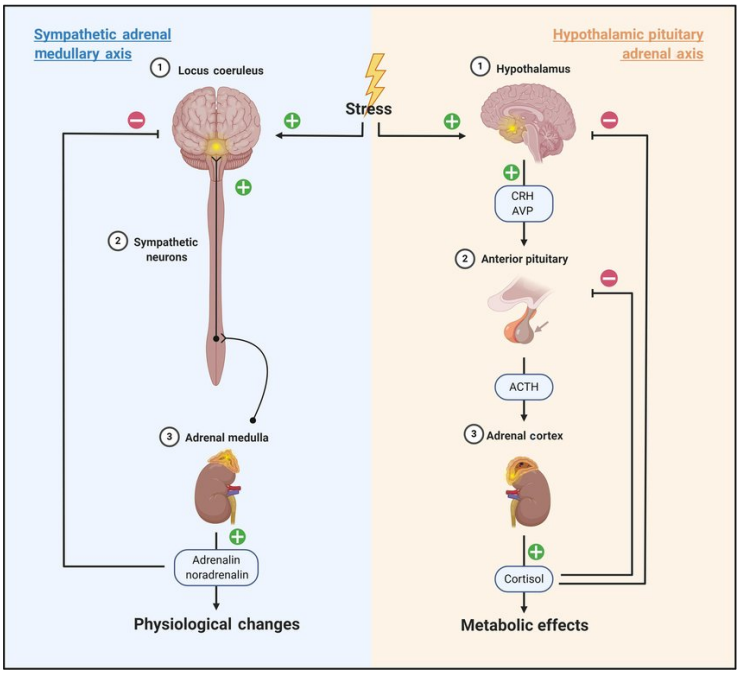


Figure 2: HPA axis vs SAM axis

1.3 Effect of stress on the heart and other body's districts

1.3.1 Heart

The heart is a vital organ and it is the only involuntary striated muscle whose function is to pump blood throughout our entire body. It is divided into the

right heart, which is responsible for pulmonary circulation, and the left heart, which is responsible for systemic circulation. Both parts are composed of atria and ventricles. The atria are synchronized, meaning that they contract and relax together, just as the ventricles do. Contractions are called systoles, and relaxations are called diastoles.

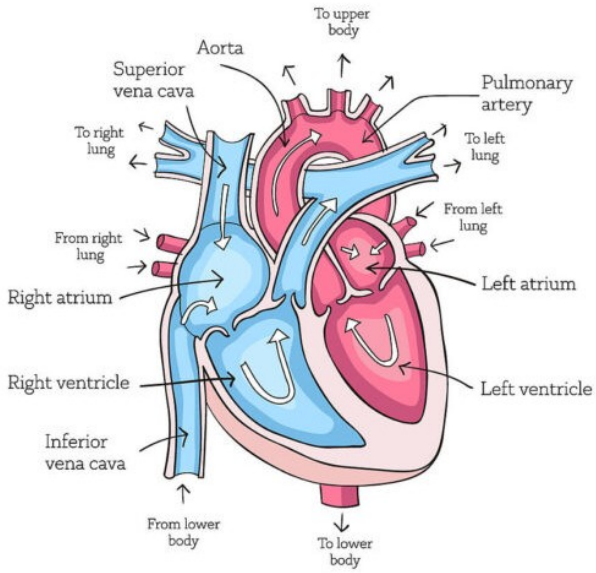


Figure 3: The human Heart

By looking at Figure 3, it can be seen that the ventricles are larger than the atria; therefore, it is impossible for the ventricles to be completely filled solely by atrial systoles. In fact, ventricular filling is said to be passive, meaning that it occurs through the pressure gradient between the atrium and the ventricle and through ventricular relaxation; the atrial systole serves to increase the efficiency of ventricular filling.

Circulation works in this way: blood arrives in the right atrium through the superior and inferior vena cava; from the atrium it passes into the right ventricle through the opening of the tricuspid valve, and from the right ventricle it enters the pulmonary circulation through the pulmonary artery (the only artery that carries deoxygenated blood) after ventricular systole. The blood is then oxygenated in the lungs and enters the left atrium through the pulmonary vein; it then passes into the left ventricle through the bicuspid (mitral) valve, and from the left ventricle it enters the aorta, the largest artery in our body, to then reach all body districts.

The heart is a myogenic organ, meaning that it is able to beat autonomously; however, its rhythm is externally regulated by the autonomic nervous system (sympathetic and parasympathetic).

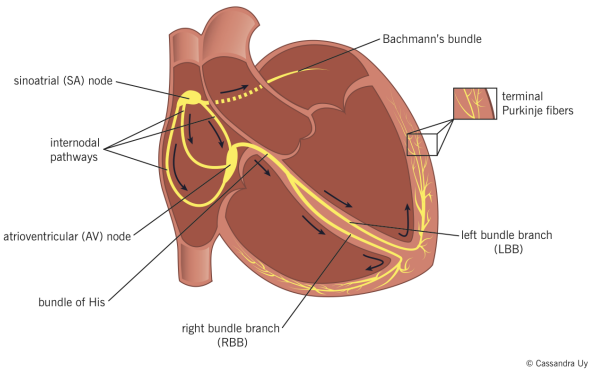


Figure 4: Electrical Conduction in the Heart

The conduction of the electrical signal of the heart begins in the sinoatrial node, which, as shown in Figure 4, is located in the right atrium at the opening of the superior vena cava. It generates impulses with a frequency of 60–100 bpm (beats per minute). From the sinoatrial node originate the internodal pathways, which constitute the conduction pathways through which the conduction signal propagates from the sinoatrial node to the atrioventricular node. The atrioventricular node is located posteriorly on the right side of the interatrial septum. Since the atrial myocardium and the ventricular myocardium are not electrically connected, electrical activity propagates exclusively through the AV node. The AV node has a discharge frequency of 40–60 bpm, and from this its function of delaying the propagation of the impulse between atria and ventricles can be deduced: they cannot contract simultaneously, otherwise blood would not be pumped. After the AV node is the bundle of His, which runs along the right side of the interventricular septum in a subendocardial position for about 1.2 cm. The common trunk of the bundle of His together with the atrioventricular node form the atrioventricular junction. From the common trunk originate the right and left branches, directed to their respective ventricles. And in the ventricles we have the Purkinje fibers, which represent the terminal part of the cardiac conduction system. Stress acts directly on cardiac conduction because, as we have seen before, it acts on the sympathetic nervous system, which is responsible for the control of heart rate, contractile force, and conduction velocity; therefore, when the release of catecholamines occurs, there is an increase in heart rate, an increase in contractile force, an elevation of arterial blood pressure, and a redistribution of blood flow, which is diverted from non-essential organs to voluntary musculature.

1.3.2 ECG signal

We have said that the sinoatrial node generates an electrical signal, called the electrocardiographic signal (ECG). From an engineering point of view, the ECG can be considered a deterministic signal in a first approximation, since it presents a morphology that repeats over time. This characteristic makes it possible to identify different components of the signal:

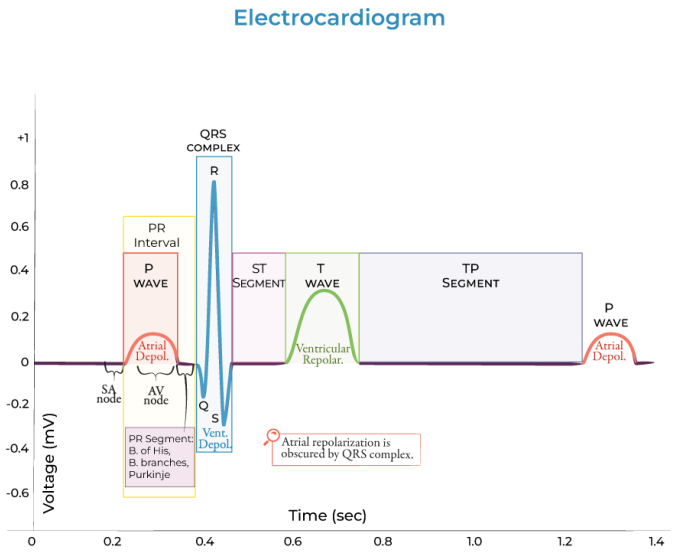


Figure 5: ECG signal

- **P wave**: represents the depolarization of the atria, which push blood into the ventricles. In a normal ECG it is small and rounded, and its duration is about 100–120 ms.
- **QRS complex**: it represents the depolarization of the ventricles. It is a sharp peak that appears after the P wave and has a duration of about 120 ms. As can be observed from Figure 5, the R wave is much higher than the P wave because the ventricles are larger than the atria.
- **T wave**: represents the repolarization of the ventricles, which corresponds to their relaxation phase, preparing for the next heartbeat. The T wave is broader than the P wave.
- **PR interval**: extends from the beginning of the P wave to the beginning

of the QRS complex and represents the atrioventricular conduction time. Its normal duration ranges from 120 to 200 ms.

- **QT interval**: extends from the beginning of the QRS complex to the end of the T wave and represents the total time of ventricular depolarization and repolarization.
- **RR interval**: represents the time between two consecutive R peaks (it's the duration of one heart beat)
- **PR segment**: extends from the end of the P wave to the beginning of the QRS complex and represents the electrical delay in the atrioventricular node.
- **ST segment**: extends from the end of the QRS complex to the beginning of the T wave and corresponds to the plateau phase of ventricular action potentials.
- **TP segment**: extends from the end of the T wave to the beginning of the next P wave. During this interval, all the heart cells are electrically at rest. For this reason, the TP segment is often used as the ECG baseline reference and also represents the moment when the heart is ready to generate the next heartbeat.

1.3.3 Digestive system

Stress and gastrointestinal disorders are correlated. During the alarm phase, the sympathetic nervous system (SNS) considers digestion a non-essential function; norepinephrine and epinephrine reduce peristalsis, causing a delay in gastric emptying. Salivation decreases and nutrient absorption is inhibited [5]. Stress can in fact worsen symptoms related to irritable bowel syndrome and can lead to nausea, gastroesophageal reflux, abdominal cramps, and even ulcers.[6]

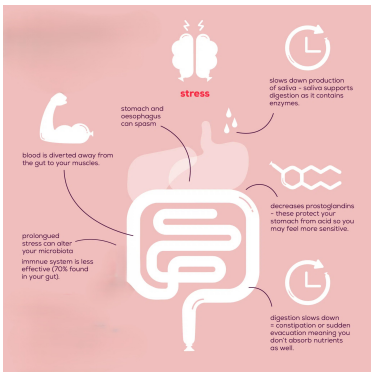


Figure 6: Stress on digestive system

1.3.4 Immune system

Cortisol produced in response to stressors plays a role in controlling inflammatory responses; however, its chronic elevation strongly weakens immune defenses by affecting both the efficiency and the number of T and B lymphocytes and natural killer cells, making the body more vulnerable to respiratory infections and even autoimmune disorders [6]. It may also reduce the effectiveness of tumor immunosurveillance mechanisms.

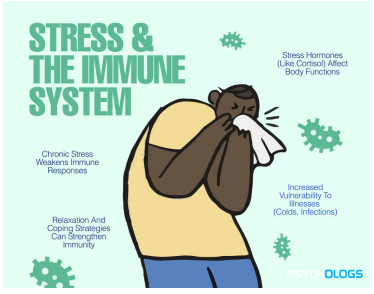


Figure 7: Stress and the immune system

1.3.5 Metabolism and Endocrine System

Stress mobilizes energy reserves to fuel the muscles, inducing significant metabolic reactions such as an increase in blood glucose levels. If stress becomes chronic, it promotes insulin resistance and increases the risk of developing type 2 diabetes. Moreover, cortisol promotes the accumulation of abdominal fat and antagonizes the action of growth hormone, thereby impairing tissue repair.

1.4 Virtual-Reality for stress detection

1.4.1 What is virtual-reality?

"Virtual Reality is electronic simulations of environments experienced via head-mounted eye goggles and wired clothing enabling the end user to interact in realistic three-dimensional situations" (Coates, 1992) [7].

Therefore, it is a type of simulation that allows the user to live an immersive experience that he or she would not be able to experience by simply looking at a monitor, a mouse, and a keyboard, because the user is surrounded by surrounding elements that do not belong to reality but are related to the images that he or she is seeing.

The first visor, called head-mounted display (HMD) [8], was created in 1968 by Ivan Sutherland, a computer scientist at Harvard University. As shown in Figures 8 and 9, the HMD was an elementary and bulky device: a computer sent images to a stereoscopic display, that is, capable of showing two separate images, one for each eye, creating the illusion of depth.

The visor could not be worn freely, since its structure was so heavy that it had to be attached to a mechanical arm suspended from the ceiling; for this reason, it was nicknamed the "Sword of Damocles."

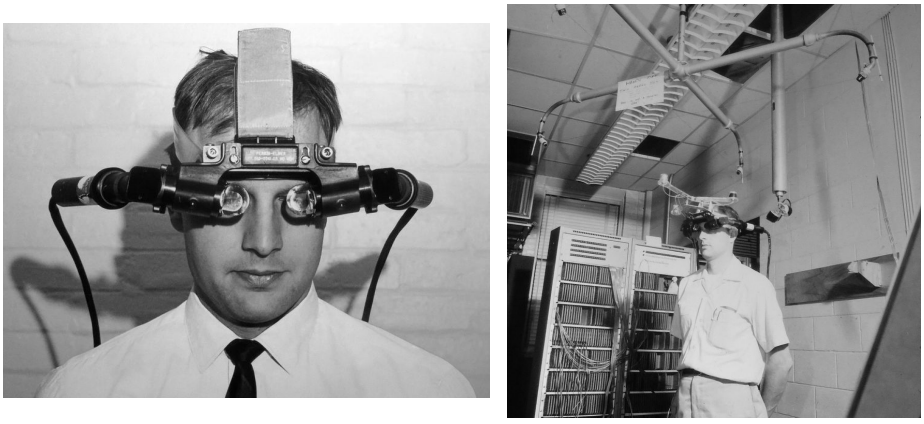


Figure 8: Sutherland's VR headset: frontal view (left) and attached to a mechanical arm suspended from the ceiling (right)

Nowadays, we can say that the idea behind visors is not different from that of Sutherland; what has improved is portability, image quality, and the technology that lies at the base. Therefore, if virtual reality was born for research purposes, gaming is what allowed its diffusion and, consequently, its technological evolution.

It is enough to think that from 2016, with the first Oculus Rift or HTC Vive, which were born as the first visors for gaming, their use was strictly limited both because the games that supported virtual reality were few and because computers had to be very powerful and therefore not very accessible.

A few years later, however, their diffusion allowed this technology to be more or less within the reach of any computer equipped with a mid-range graphics card, and we have arrived at the Meta Quest, which are the visors that are currently used the most.

1.4.2 Virtual Reality in Healthcare

Precisely because of its characteristics of immersivity and involvement, virtual reality today is not used only for gaming but also for many applications, especially in the healthcare sector:

- **Training of doctors and surgeons**: it is used to train future doctors and surgeons: with virtual reality, doctors can practice on specific, complex, and potentially risky cases with the possibility of making mistakes without endangering the patient's life, but also simply to better learn the anatomy of the human body by observing the arrangement and the three-dimensional appearance of the structures of the body.[9]
- **Reducing stress and pain**: it is possible to recreate environments in which the patient/user is subjected to a controlled stress load and to stimuli that allow mitigating the state they are experiencing, in order to help them when similar situations occur in everyday life. The advantage of VR is that, depending on the user's needs, different types of scenarios can be used [10]. This section will be discussed in more detail in the next paragraph.
- **Cognitive and motor rehabilitation**: the use of various environments makes it possible to examine the possibilities related to the risk of cognitive decline and to slow its progression; it has been shown that the integration of VR with motor training in patients affected by Parkinson's disease leads to improvements in gait, balance, and mobility compared to traditional physiotherapy alone, thanks to the use of specific exercises that promote motor learning through continuous feedback and personalized adaptations [11].

1.4.3 Stress detection: State of the art

As already mentioned in the previous paragraphs, in recent years the study of stress has assumed an increasingly important role, since it has been widely demonstrated that prolonged exposure to stressful conditions can promote the onset of mental disorders, cognitive deficits, and complex pathologies. In this context, stress detection is fundamental to identify the causes and triggering factors of stress, allowing it to be monitored, controlled, and its negative effects mitigated. For this reason, several tests exist that allow controlled induction of cognitive and social stress, such as the MIST, TSST, Stroop Word–Colour Test, and MAT, currently among the most widely used in experimental studies.

- **Trier Social Stress Test (TSST)**: After a relaxation period that constitutes the baseline, the test consists of simulating a job interview in which the participant must convince a panel to hire them, and then performing a mathematical task that involves carrying out successive subtractions [12].

Figure 9: Example of TSST setup

- **Montreal Imaging Stress Task (MIST)**: It is a task divided into three phases: rest, control, and experimental. In the rest phase, the participant looks at a static screen where no tasks are shown; in the control phase, the participant must answer arithmetic questions; and in the experimental phase, the participant answers arithmetic questions designed to be beyond their capabilities, thus inducing a cognitive stress condition due to the impossibility of knowing the correct answer. Furthermore, in the experimental condition, the participant sees their individual and average performance so that competitiveness acts as an additional stressor [13].

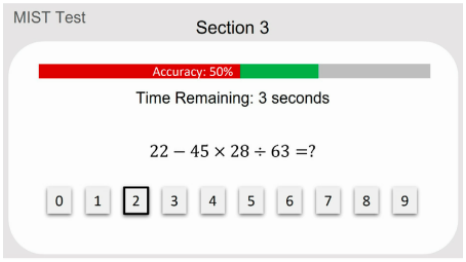


Figure 10: Example of MIST task interface

- **Stroop Word–Colour Test**: It consists of inducing cognitive stress through the presentation of words that indicate colors, printed in an ink color that can be congruent or incongruent with the meaning of the word. The participant is asked to recognize the ink color of the words while avoiding being influenced by the color represented by the text. The limited time to respond to each question acts as a stressor [14].

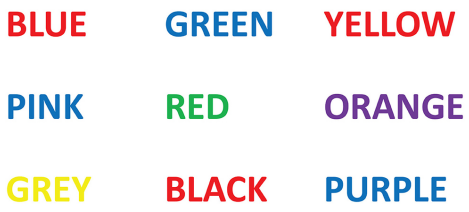


Figure 11: Example of Stroop Word–Colour Test

- **Mental Arithmetic Task (MAT)**: It consists of answering a series of mathematical operations, with a limited time for each response. The task can be customized based on the participant's level, increasing the difficulty of the calculations or providing multiple sessions with increasing difficulty [15].

1.4.4 Physiological Signals

To measure the induced stress state, sensors are used to record physiological signals and machine learning–based classification algorithms, which extract features from the acquired signals to quantify the level of stress to which the subject is exposed.

The main physiological signals considered are: ECG, EEG, GSR/EDA, photoplethysmographic signal (PPG), EDR/Respiratory signal:

Signal	**Band**	**Description**
Electrocardiogram (ECG)	0.05 — 125 Hz (500 Hz)	Records the electrical activity of the heart;
Photoplethysmogram (PPG)	0.01 — 10 Hz (50 Hz)	Measures peripheral blood volume changes;
Galvanic Skin Response / Electrodermal Activity (GSR/EDA)	0.01 — 0.5 Hz (4 Hz)	Measures skin conductance changes reflecting sympathetic nervous system activity related to stress.
Electromyogram (EMG)	1.0 — 500 Hz (2000 Hz)	Records electrical activity of muscles; indicates muscle tension and physical stress.
Electroencephalogram (EEG)	0.1 — 40 Hz (500 Hz)	Records electrical brain activity; useful for analyzing attention, vigilance, and emotional states.
Surface Temperature	0.001 — 0.1 Hz (1 Hz)	Measures superficial skin temperature; variations can indicate physiological stress.
Respiration	0.1 — 1 Hz (4 Hz)	Records thoracic or abdominal movements; used to analyze respiratory rate and depth during stress.

Table 1: Physiological signals characteristics

1.4.5 Machine Learning/Deep Learning Models

The physiological signals analyzed are often combined, as integrating multiple measures provides a more complete view of the stress response. Stress detection is performed using machine learning algorithms that extract features from these signals over time windows of varying lengths, depending on the signal bandwidth, the desired temporal resolution, and the purpose of the experiment. Here's some examples of the most used ones:

Support Vector Machine (SVM) It is a supervised learning algorithm used for both classification and regression. The Support Vector Machine (SVM) consists in determining a hyperplane that maximizes the margin, that is, the distance between the separating hyperplane and the closest points belonging to different classes, called support vectors. SVM is particularly effective when a limited amount of training data is available and is widely used in binary classification problems [16]. If the data cannot be separated linearly, a nonlinear kernel, such as the RBF or a polynomial kernel, can be employed, which transforms data into a domain in which data separation is linear.

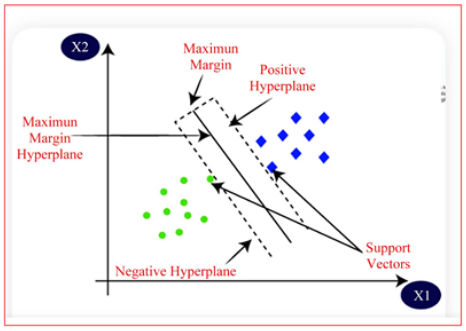


Figure 12: Illustration of a Support Vector Machine showing the hyperplane and support vectors that maximize the margin between classes.

Random Forest (RF) Random Forest is a supervised prediction algorithm that can be applied to classification and regression and involves building multiple predictive models, typically decision trees. The mechanism of training is based on bagging, also called Bootstrap Aggregating. Here, a different dataset is created for each decision tree using random sampling with replacement from the original dataset. Also, all decision trees have equal weight in taking a decision, obtained via majority vote for classification or averaging for regression[17]. To avoid correlation between trees and to enhance diversity, at each node, only a random subset of features is employed. During training, each tree grows branches until at least one of the following conditions is met: all instances in the node belong to the same class, the maximum depth is reached, or the node contains a minimum number of samples. During the prediction phase, each instance of the dataset passes through all trees, following the branches according

to the feature values until reaching the leaves, where each tree provides a decision; these decisions are then aggregated to obtain the model's final prediction.

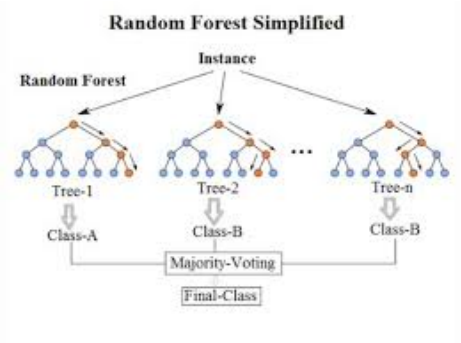


Figure 13: Schematic representation of a Random Forest model, showing multiple decision trees voting to produce a final prediction.

K-Nearest Neighbors (KNN) K-Nearest Neighbors is a supervised classification and regression algorithm that assigns an observation to the class most frequent among its k nearest neighbors in the feature space. The distance is calculated using the Euclidean distance. The disadvantage of this algorithm is that it has a high computational cost.

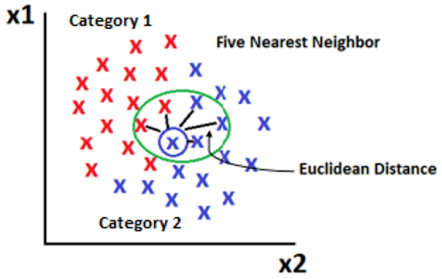


Figure 14: Overview of the K-Nearest Neighbors algorithm illustrating the assignment of a test instance to the class most common among its k nearest neighbors.

Neural Networks (NN) Neural Networks (NN), including deep learning architectures, are models capable of learning complex and nonlinear relationships between inputs and outputs. They consist of layers of artificial neurons, where each neuron performs mathematical operations on the received data (weighted sum + activation function). The data passes through these layers, and as it moves through deeper levels, the network can extract increasingly complex features from multichannel signals, eventually producing an output corresponding to the class of interest (in classification) or a continuous value (in regression)[17].

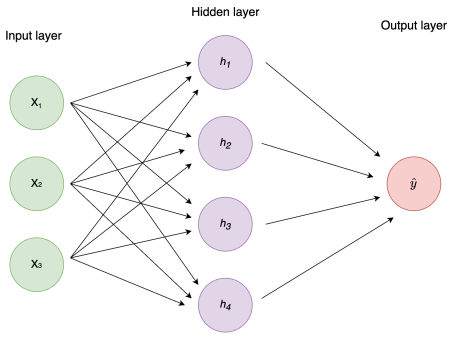


Figure 15: Example of a Neural Network architecture.

1.5 Purpose of the thesis project

This general overview of stress, stress detection techniques, and the use of virtual reality is intended to introduce the main objective of the present thesis project. The work consists in the development of a virtual reality environment, implemented using the Unity game engine, with the aim of inducing a controlled state of stress in the user, quantifying it in real time through a regression algorithm, and subsequently applying targeted sensory stimulation to mitigate the induced stress state.

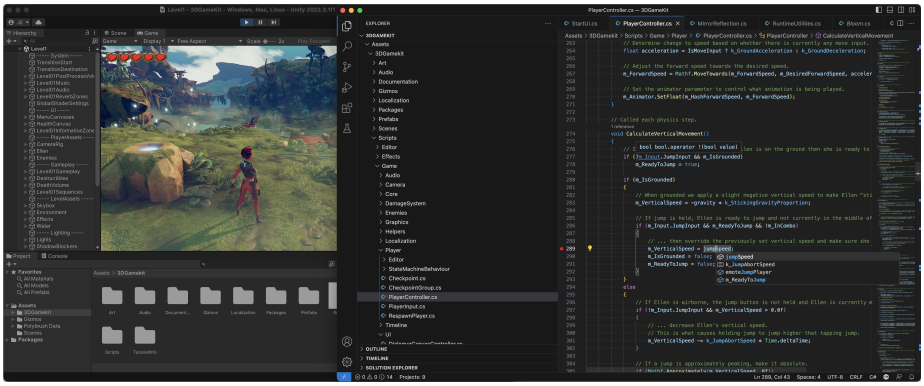


Figure 16: Unity inteface

1.5.1 Unity and the developed game

As previously mentioned, Unity is a software platform widely used for the development of 2D and 3D video games and, in recent years, it has also been extensively adopted in the fields of augmented reality (AR) and virtual reality (VR).Unity primarily employs a programming language called C#. This is an object-oriented programming language in which scene objects are represented by classes. One of the biggest features of Unity is its interface, which provides the ability to create intricate virtual environments even without extensive programming knowledge. There are many functionalities that can be implemented directly via the integrated and graphical interface tools available within Unity, such as collision detection, physics simulations, materials, lighting, animations, audio, user interface elements, and input systems.

In this stressful environment, there is a racing game in VR mode in which the user will compete with four cars driven by artificial intelligence in order to win the race. In this game, the user will be required to follow the trajectory that flashes in order to maintain the speed of the vehicle, which is necessary in order to compete with the vehicles. Failure of the user to follow the trajectory will result in the slowdown of the vehicle.

The flashing trajectory constitutes the main stressor element of the game, as it is difficult to follow due to the high sensitivity of the analog sticks of the controllers (joycons). An additional stress-inducing element is the competitive

nature of the scenario, which leads the user to remain in a constant state of tension driven by the objective of achieving first place. The game includes three operating modes: Training, Play 1, and Play 2.

- **Training mode**: is designed to allow the user to become familiar with the joy-cons, the vehicle dynamics, and the trajectory by completing a lap without competing against other vehicles.
- **Play 1 mode**: involves racing against the other vehicles, during which the user must follow a flashing trajectory at a constant frequency of 6 Hz. The race lasts a total of five laps. In this mode, the joy-cons vibrate at the same fixed frequency used for the flashing trajectory.
- **Play 2 mode**: is structurally similar to Play 1 but introduces an adaptive mechanism: the flashing frequency of the trajectory is dynamically modified based on the user's stress level, as quantified by the regression algorithm. The objective is to mitigate the induced stress by exploring stimulation frequencies in the range of 4 to 8 Hz. The details of this modulation strategy, as well as the frequency range used, will be discussed in the next chapter. The flashing frequency of the trajectory is also synchronized with the vibration frequency of the joy-cons.

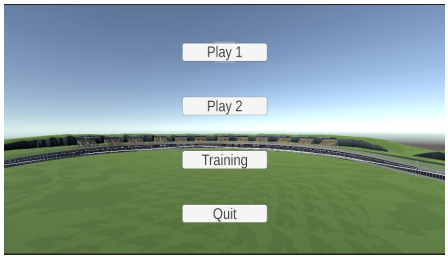


Figure 17: Main menu showing the modes: Training, Play 1, and Play 2

Figure 18: Player view with the vehicle and the trajectory to follow

The sequence in which these modes will be played and the strategy of trajectory blinking and controller vibration will be discussed in the next chapter.

1.5.2 Project details and headset usage

The use of visual stimulation with a blinking trajectory represents a critical aspect, as users with photosensitive epilepsy cannot participate in this experiment. According to the WCAG (Web Content Accessibility Guidelines), flickering at frequencies between 3 and 50 Hz can induce epileptic seizures in photosensitive subjects [18]. The WCAG also specify that the term "blinking" refers to a flicker that is intended to end spontaneously; if it persists over time, it is defined as "flashing". For simplicity, from now on we will use the term "blinking" as a synonym for "flashing".

In the literature, there is no evidence regarding the possible effect of visual stimulation on stress reduction; it is mainly studied in the context of "steady-state-visual-evoked-potential" (ss-VEP) for BCI. For this reason, since there are no articles on this topic, this thesis aims to try to answer the following question: **"Does modulated sensory stimulation have a greater effect than constant stimulation in stress mitigation?"**

The immersive effect of virtual reality has already been discussed, but not its side effects. Prolonged use, in fact, can generate nausea, stomach discomfort, and dizziness [19]. For this reason, the experimental sessions have been programmed to last less than 10 minutes, in order to minimize such effects .

2 Materials and Methods

2.1 Materials

Among the materials used to realize the entire architecture of the project, the following were employed:

- **MSI Cyborg 15 A13VF-1853IT**, gaming notebook equipped with an *Intel Core i7-13620H* processor, *NVIDIA RTX 4060* GPU with 8 GB of GDDR6 memory, and 16 GB of DDR5 RAM.
- **Polar H10**, used for the real-time acquisition of the heart rate signal. It has a sampling frequency of 130 Hz, and it is recommended to moisten it before use in order to reduce the electrode–skin impedance.
- **Meta Quest 2**, which is the headset used by the participants to perform the experiment. It is equipped with two controllers; however, only one controller is used by the participants, with the non-dominant hand.
- **MATLAB 2025b**, used for the processing, analysis, and real-time visualization of the acquired physiological signals.
- **Unity**, software used for the creation of the virtual reality environment (specify the version used).
- **Microsoft Visual Studio**, development environment used for programming Unity scripts in the `C#` language.
- **Visual Studio Code**, code editor used for development in the `Python` language, particularly for managing communication via a `TCP/IP` connection, which is necessary for transferring in real time the heart rate signal acquired by the Polar H10 chest strap to MATLAB.

2.2 Participants

The subjects who participated in the experiment were 20, divided into two groups of 10 each: Group A and Group B. Group A first played the Play 1 mode with constant stimulation, namely trajectory blinking and controller vibration at a frequency that remains fixed for the entire duration of the mode, and then played Play 2 with modulated stimulation. Group B first played the Play 2 mode and then the Play 1 mode. The participants involved in the experiment had a mean age of 24 ± 2 years, all with an educational level ranging from a bachelor's to a master's degree. The experiments were conducted between 10:00 a.m. and 4:00 p.m. Before the experiment, the subjects were advised not to drink coffee, since caffeine can alter the normal heart rate, not allowing the regressor to optimally quantify stress.

2.3 Experimental protocol

The experiment follows a well-defined experimental protocol, divided into the following phases:

- **Setup phase**: setup of all the hardware and software tools required to perform the experiment. During this phase, the participant wears the Polar chest strap and over-ear headphones.
- **Preliminary questionnaire**: administration of a preliminary questionnaire to the participants in order to collect basic demographic data (age, sex, and dominant hand), general health status (in particular the use of medications that may affect heart rate and the presence of epileptic seizures), sensitivity to virtual reality, and daily habits such as caffeine consumption.
- **Baseline**: the participant watches a relaxing video lasting 3 minutes. Since the aim of the experiment is to evaluate stress, the baseline phase is important to minimize the effect of pre-existing stress and anxiety.
- **Mental Arithmetic Task**: phase in which the participant is required to solve arithmetic operations for 3 minutes. As described in Section 1.4.3, this task is used to induce a controlled stress condition.
- **Game task**: the participant wears the virtual reality headset and performs the game task according to the experimental condition defined by the assigned group.
- **Sensor and headset removal**: removal of the chest strap sensor and the virtual reality headset, concluding the experimental session.

After the baseline phase, the Mental Arithmetic Task, and each of the two game sessions, the participant is required to complete a **VAS (Visual Analog Scale)** questionnaire, which evaluates the auto-referred stress level on a scale from 1 to 5.

After the Mental Arithmetic Task and after the Game Task, the participant also completes the **NASA-TLX** questionnaires, which assess the mental and physical workload experienced during the tasks.

At the end of the game task, immediately after completing the NASA-TLX questionnaire, the user fills out an additional questionnaire aimed at providing feedback on the video game. In this questionnaire, the participant describes which elements were perceived as stressful and what sensations were experienced during gameplay. The purpose of this feedback is to determine whether, in addition to the known stressors, there are other elements that were not previously considered and that may have induced additional stress.

2.4 Questionnaires

Among the questionnaires used, as mentioned in the previous section, in addition to the preliminary questionnaire and the video game feedback questionnaire, the VAS (Visual Analogue Scale) and the NASA-TLX (NASA Task Load Index) are employed.
These two questionnaires are commonly used to assess self-reported stress and are mainly intended to verify whether there is consistency between the stress perceived by the user and that estimated by the regression model.

The **VAS** is a psychometric instrument used to quantify the intensity of a subjective perception, such as pain or stress [20]. For pain assessment, a horizontal 10 cm line is typically used, with two completely opposite conditions placed at the extremes, such as no pain and worst possible pain. No intermediate markers or emojis are included, as the scale is intended to remain as analog as possible.

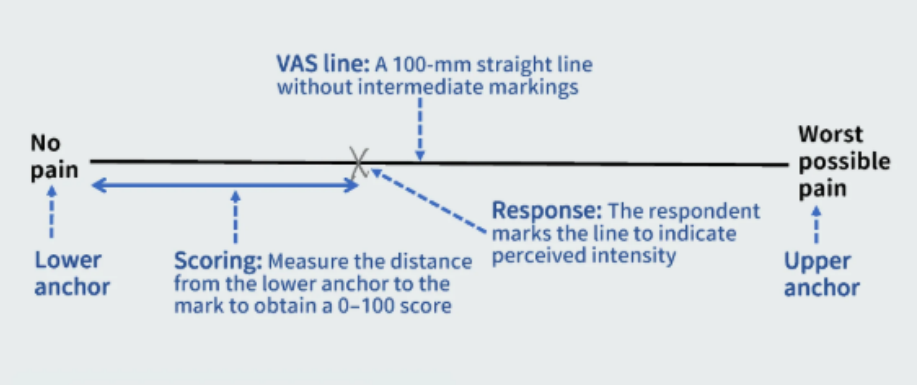


Figure 19: A typical Visual Analogue Scale (VAS) for pain assessment

For stress assessment, a structure similar to a numerical scale is preferred, usually ranging from 1 to 10, where 1 corresponds to no stress and 10 to high stress. In this experiment, the scale was set from 0 to 4. The VAS is a highly versatile questionnaire, as it can be administered at any stage of the experiment; it is easy to understand and unambiguous.

Figure 20: Experimental Visual Analogue Scale (VAS) used for stress assessment, where 0 indicates low stress and 4 indicates high stress

The NASA Task Load Index (NASA-TLX) is used to evaluate stress workload from both a physical and a mental perspective by analyzing six dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration [21][22].
The task consists of two parts. In the first, the six dimensions are coupled in pairs with each other, resulting in 15 combinations. For each pair, the user is asked to indicate which of the two dimensions contributed more to the perceived workload. The number of times a dimension is chosen weights it. In the second part, the rating that the user gives to each dimension ranges from 0 to 100. The final score of the test then is calculated as:

$$\text{NASA-TLX} = \frac{\sum_{i=1}^{6} R_i \cdot W_i}{\sum_{i=1}^{6} W_i} = \frac{1}{15} \sum_{i=1}^{6} R_i \cdot W_i \tag{1}$$

where:

- R_i represents the score (*rating*) assigned to the i-th dimension;
- W_i represents the weight associated with the i-th dimension, derived from the pairwise comparisons.

The final score ranges from 0 to 100; for example, a value of 80 indicates a relatively high level of perceived stress. For this experiment, the rating scale ranges from 0 to 4. The formula used to calculate the NASA-TLX score remains the same as in the original method, with the only difference being the numerical range of the ratings.

2.4.1 Baseline and Mental Arithmetic Task

During the baseline phase, the participant is shown a relaxing video that, throughout its entire duration, proposes breathing and meditation exercises guided by a narrated voice. The objective is to induce a state of relaxation. During the Mental Arithmetic Task (MAT), instead, the participant is required to repeatedly subtract 17 from a randomly generated four-digit starting number. The task has a total duration of 3 minutes, and for each operation the participant is given 8 seconds to provide the answer, which must be entered using the keyboard. The choice of an 8-second time limit is not random, but represents a compromise aimed at making the task sufficiently challenging without inducing an excessive level of difficulty that could lead the participant to prematurely abandon the task, due to the inability to correctly complete the input of the answer within the given time, compromising the extraction of features from the ECG signal that the regressor will use.

In the event of an incorrect response, the four-digit starting number is randomly regenerated and the participant's score is reset to zero. This mechanism is intended to induce psychological pressure on the participant, as an incorrect response at a later stage nullifies all previously accumulated progress. Furthermore, the participant is additionally stressed by the presence of an acoustic signal (buzzer), which is activated both in the case of an incorrect response and when the allotted time expires without a response being provided.

2.4.2 Game Task

In this phase, as specified in the previous sections, participants are required to first complete the Training mode and subsequently play the two experimental modes (Play 1 and Play 2), following the order determined by their assigned experimental group.

During the Training mode, participants are initially shown an instructional image depicting both joy-cons, illustrating the available controls. Although the game is played using a single controller, the task must be performed with the non-dominant hand; therefore, displaying the controls of both joy-cons allows both right-handed and left-handed users to have an appropriate reference.After that, the interface appears, which helps the participants place the camera at their desired location. Then, the training process ends by running a single lap, whereby the route remains stationary and does not have a flashing feature, and no other vehicles to compete with during the race.

The structural form of the modes Play 1 and Play 2 is the same as that of the modes described in 1.5.1, differing only in the trajectory flashing modality and the vibration of the controllers. At the start of both races, background music is played to enhance immersion and environmental noise, intentionally introduced as an additional stressor. Simultaneously, the user interface is activated, displaying the current position in the race, the number of completed laps out of the total, and a timer. These elements are explicitly designed to stimulate competitiveness and performance pressure, thereby contributing to stress

induction.

Both races terminate when the participant crosses the finish line for the fifth time, completing a total of five laps.

2.5 Methods

This section is fundamental to describe all the mechanisms of signal processing and the algorithms implemented to manage feature extraction, model training, real-time regression, and the stimulation modulation strategy, creating a real-time closed-loop system.

2.5.1 ECG processing

During the **baseline** phase and the **MAT** phase, the electrocardiographic (ECG) signal is acquired through a Polar chest strap and then saved in CSV format on Excel. The signal is then imported into MATLAB and divided into time windows of 4 seconds, with a **75% overlap** between consecutive windows.

Each extracted signal segment is provided as input to a customized MATLAB function called detect_PQRST.m. Inside this function, the signal is initially **detrended** to remove the low-frequency contribution.

Subsequently, two recursive **Butterworth** filters are applied:

- a **high-pass** filter of order 3 with a cutoff frequency of 100 mHz,
- a **low-pass** filter of order 4 with a cutoff frequency of 40 Hz.

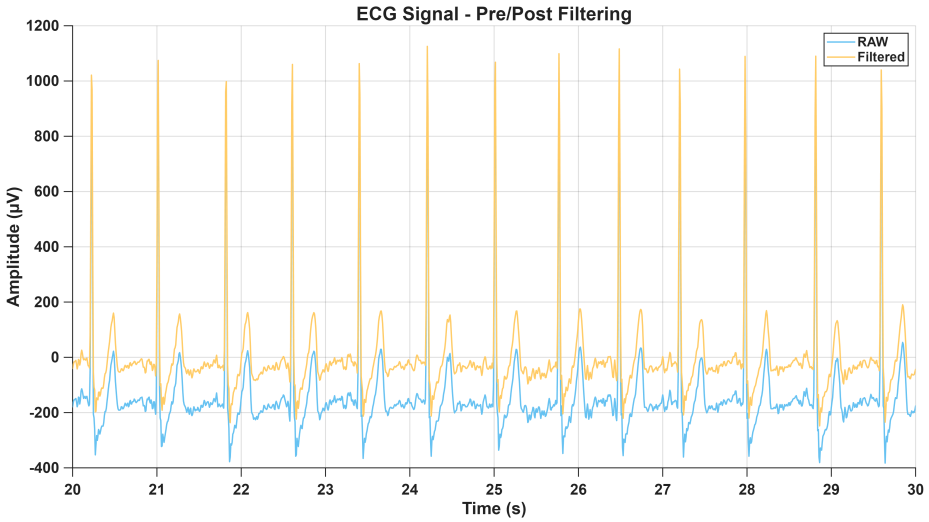


Figure 21: Raw ECG signal vs filtered ECG signal

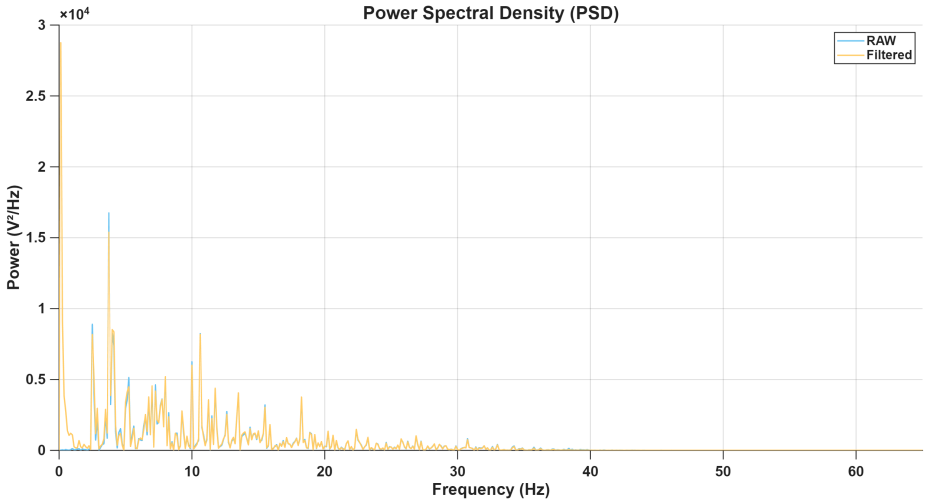


Figure 22: Power Spectral Density of raw and filtered ECG

The ECG signal filtered in this way is then analyzed using the MATLAB function findpeaks.m, which is used to detect the **R peaks**. In particular, the peak detection is performed by imposing:

- a minimum peak height equal to

$$\text{MinPeakHeight} = 1.5 \cdot \sigma_{ECG}$$

  where σ_{ECG} represents the standard deviation of the filtered ECG signal;

- a minimum distance between consecutive peaks of 300 ms.

Once the R peaks are detected, the **RR intervals** are calculated as the time difference between two consecutive peaks:

$$RR_i = t_{R_i} - t_{R_{i-1}}$$

From these intervals, the **mean RR** of the analysis window is calculated, which is used to estimate the **mean heart rate**:

$$HR_{\text{mean}} = \frac{60}{\overline{RR}}$$

Identification of P, Q, S, and T waves

The waves are identified with temporal reference to the R peak, according to physiologically motivated search windows [23]:

- **Q wave:** local minimum in the interval

$$[R - 82 \text{ ms},\ R]$$

- **S wave:** local minimum in the interval

$$[R,\ R + 82 \text{ ms}]$$

- **P wave:** local maximum in the interval

$$[Q - 198 \text{ ms},\ Q]$$

- **T wave:** local maximum in the interval

$$[S,\ S + 200 \text{ ms}]$$

Adaptation of windows to heart rate

For the first heartbeat, the above time windows are defined assuming a **reference duration**:

$$RR_{\text{ref}} = 0.8 \text{ s}$$

corresponding to a reference heart rate of:

$$HR_{\text{ref}} = \frac{60}{0.8} = 75 \text{ bpm}.$$

For subsequent heartbeats, in order to account for local variations in heart rate, all search windows are adapted through a **scaling factor** α, defined as:

$$\alpha = \frac{RR_{\text{local}}}{RR_{\text{ref}}} = \frac{RR_{\text{local}}}{0.8}$$

where RR_{local} represents the RR interval calculated between the current and the previous heartbeat. The time window sizes are then scaled by α, enabling the P, Q, S, and T wave detection process to dynamically adapt to local variations in heart rate, including bradycardia and tachycardia.

As illustrated in Figure 23, the detect_PQRST.m function allows for the automatic identification of the main characteristic events of the cardiac cycle, returning for each heartbeat the temporal indices and amplitudes of the P, Q, R, S, and T peaks. Based on this information, the physiological features used for training the regression model are subsequently extracted. In particular, the extracted features are the following:

- **BPM (Beats Per Minute)**: average heart rate, computed from the RR intervals:

$$\text{BPM} = \frac{60}{\overline{RR}} \tag{2}$$

  where $\overline{RR}$ represents the mean of the RR intervals expressed in seconds.

- **RTr**: ratio between the mean amplitude of the R-wave and the amplitude of the subsequent T-wave. Under stress conditions, the R-wave tends to be more pronounced and vigorous, increasing the ratio compared to relaxed conditions.

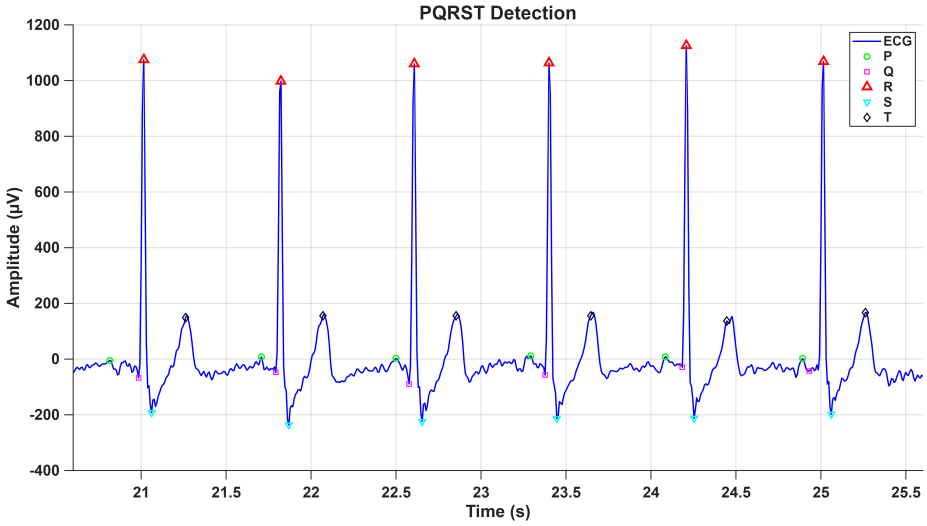


Figure 23: Output Plot of the function detect_PQRST.m

- **PRr**: ratio between the mean amplitude of the R-wave and the amplitude of the preceding P-wave. Under stress conditions, the R-wave tends to be more pronounced, resulting in an increase of this ratio compared to relaxed conditions.
- **Mean QT ($QT_{\mathbf{mean}}$)**: average QT interval computed over 4-second windows:

$$QT_{\text{mean}} = \frac{1}{N} \sum_{i=1}^{N} QT_i \tag{3}$$

- **QTc**: QT interval corrected for heart rate using Bazett's correction:

$$QT_c = \frac{QT}{\sqrt{RR}} \tag{4}$$

  where QT and RR are expressed in seconds.

- **SDNN**: standard deviation of successive RR intervals:

$$\text{SDNN} = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (RR_i - \overline{RR})^2} \tag{5}$$

- **pNN50**: percentage of successive RR intervals differing by more than 50 ms:

$$\text{pNN50} = \frac{1}{N-1} \sum_{i=1}^{N-1} I\left(|RR_{i+1} - RR_i| > 50\,\text{ms}\right) \times 100 \tag{6}$$

- **RMSSD**: root mean square of successive differences between RR intervals:

$$\text{RMSSD} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N-1}(RR_{i+1} - RR_i)^2} \tag{7}$$

.

SDNN, pNN50, and RMSSD are features of Heart Rate Variability (HRV), which reflects how the autonomic nervous system responds to various conditions, including stress. These metrics can be computed from the tachogram, providing a non-invasive assessment of autonomic regulation.

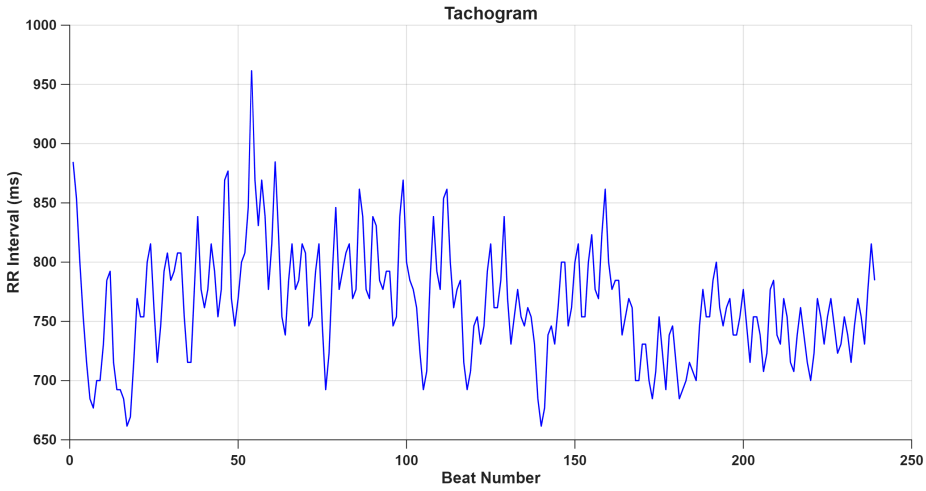


Figure 24: Tachogram

Feature	Description	Expected Trend
BPM	Average heart rate	↑
RTr	R-wave / T-wave ratio	↑
PRr	R-wave / P-wave ratio	↑
Mean QT	Average QT interval	↓ / stable
QTc	Heart-rate corrected QT	↓ / stable
SDNN	Standard deviation of RR intervals	↓
pNN50	Percentage of RR differences >50 ms	↓
RMSSD	Root mean square of successive RR differences	↓

Table 2: ECG features used for regression and their expected variation under stress conditions. Arrows indicate the physiological trend compared to relaxed conditions.

The features are normalized using the z-score, i.e., from each feature the mean of the corresponding features computed over all segments is subtracted, and the result is divided by the standard deviation. This operation ensures that the resulting features have zero mean and unit variance. The z-score normalization is defined as:

$$z = \frac{x - \mu}{\sigma}$$

where x is the value of the original feature, μ is the average feature calculated on each segment and σ is the realtive standard deviation

2.5.2 Regressor

The choice of using a regressor instead of a classifier derives from the fact that the regressor provides a continuous output, whereas the classifier returns binary values. For the task under consideration, it is important to obtain a continuous coefficient between 0 and 1, as this allows a more realistic representation of the transition of a stress-inducing event toward a non-stressful condition. On the contrary, a classifier would provide discrete values (for example 1 for stress and 0 for no stress), which would be poorly informative in intermediate phases. By using a regressor, it is instead possible to capture both the direction and the intensity of the influence of the event on the subject's response.

With regard to the model, a *Random Forest* was chosen, which is widely used in the literature for stress detection problems due to its high stability and low probability of overfitting [24]. As discussed in Paragraph 1.4.5, in the regression setting the output of each tree is averaged in order to produce, in this case, the

stress coefficient, which lies within the interval $[0, 1]$, where 0 corresponds to no stress and 1 corresponds to stress.

The model configuration is defined by the following MATLAB instruction:

```
Model = fitrensemble(features, label_win, 'Method', 'Bag', ...
'NumLearningCycles', 100, 'Learners', 'tree', 'FResample', 1.0);
```

The model hyperparameters are the following:

- **Features**: usually, for regression tasks, a number of features equal to one third of the total should be used; considering that there are 8 features and therefore few features per tree, all of them were used.
- **Number of trees (`NumLearningCycles` = 100)**: this value represents the so-called sweet spot between low variance and contained training and inference times.
- **Ensemble method (`Method` = Bag)**: bootstrap aggregation is used to improve model robustness and reduce overfitting, and `FResample` = 1.0 indicates that each tree operates on a dataset with the same size as the original one.
- **Node splitting criterion**: Mean Squared Error (MSE), which is the default choice for regression problems.
- **Maximum tree depth**: unlimited, in order to allow trees to split into many branches until the leaves are reached (samples that all belong to the same class).

To evaluate the model performance, a **10-fold cross-validation** procedure is adopted. The dataset is divided into ten equally sized subsets; at each iteration, one subset is used as the **test set**, while the remaining nine form the **training set**. This process is repeated for a total of ten iterations, ensuring that each fold is used exactly once as the test set.

For each fold, the coefficient of determination R^2 is computed, providing a measure of the model's ability to explain the variance of the extracted features. The coefficient R^2 is defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2} \tag{8}$$

where y_i represents the ground truth value, $\hat{y}_i$ the model prediction, and $\bar{y}$ the mean of the ground truth values. The model associated with the highest R^2 value is selected as the optimal model, while the average R^2 value is used to assess the overall goodness of fit.

In addition to the coefficient of determination, it is good practice to compute the **Root Mean Square Error** (RMSE), which provides a quantitative measure of the prediction error and is therefore informative of how close the model predictions are to the ideal target values. The RMSE is defined as:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{9}$$

In general, a high R^2 value, corresponding to a good explanation of the **Variance Accounted For** (VAF), is often associated with a low RMSE. However, this relationship is not always guaranteed: in the presence of highly variable data—although this is not the case in the present study, as the input data were normalized using z-score normalization—a model may accurately explain the overall variance of the dataset while still exhibiting a relatively high prediction error. This occurs because R^2 is a dimensionless metric, whereas RMSE depends on the scale of the target variable.

Lastly, for the chosen model, its feature importance is calculated, followed by the determination of the top five features based on that importance. Then, the model is retrained using these features, with the objective of simplifying the chosen model for expected improvements in its generalization performance.

2.5.3 Real-Time Stress Detection Implementation

The model that is chosen is stored and used to conduct the *game task* for real-time stress detection. For the implementation of the ECG signal, it is obtained through the use of a Polar chest strap and is sent to the MATLAB interface through a TCP/IP connection by the use of Python as a bridge.

The fact that a TCP/IP connection is being utilized results in efficient and low-latency data processing as opposed to using a strategy that implements continuous signal storage in `.csv` files and simultaneous reading via MATLAB. This strategy performs a signal processing and classification pipeline in a near-real-time manner.

The ECG signal is stored in a sliding buffer with 520 samples, representing a window of 4 seconds, based on the sampling frequency provided by the Polar device at 130 Hz.

Within each second, the sampling window updates its content with 130 samples, representing a 75% overlap between successive windows.

The same process for the entire pipeline of preprocessing and feature extraction is carried out, just like in the training phase. This is fundamental to ensuring that predictions are made properly, considering that the model was trained under certain conditions, and any discrepancy in the preprocessing process could automatically result in unreliable predictions.

Finally, for stability and safety reasons, the stress coefficients obtained as output of the regressor are clipped to values in [0,1] so as to be coherent with the given labels.

Furthermore, the output of the regressor is always subject to a *post-processing* step which uses information provided by the last step. Specifically, the estimated coefficient of the current step is computed as a weighted average of the current output and the two last predictions, following the structure:

$$y_{i_{\text{final}}} = \begin{cases} y_i, & \text{if } i = 1 \\ 0.7\, y_i + 0.3\, y_{i-1}, & \text{if } i = 2 \\ 0.5\, y_i + 0.3\, y_{i-1} + 0.2\, y_{i-2}, & \text{if } i \geq 3 \end{cases}$$

where y_i denotes the regressor prediction at iteration i, and $y_{i_{\text{final}}}$ represents the final value after post-processing. This scheme allows the estimate to be stabilized by reducing the influence of possible local fluctuations in the model output.

2.5.4 Close loop system

The model predictions, as previously described, are used to estimate the user's stress state, identifying whether the user is in a stressed condition, a no-stress condition, or in a transitional state between the two. During the game task, the user completes the three game modes described in the previous paragraphs. In all modes, the regressor keeps track of the user's state; however, it plays a particularly significant role only in Play 2 mode, as the predictions are used to modulate the vibration frequency of the controllers and the blinking of the trajectory, directly influencing the gaming experience based on the measured stress.

However it is important to monitor the user's state in all modes. The training mode primarily serves to allow the user to become familiar with the game and the virtual environment; what is relevant for the experiment is to compare the stress load between Play 1 and Play 2, in order to assess whether stress-modulated stimulation is more effective than continuous stimulation that is independent from the measured stress.

The inversion of the execution order between the two groups of users is intended to eliminate the learning bias: if all users start with Play 1 and then continue with Play 2, a potentially lower stress level recorded in Play 2 may instead be attributable to the user's previous familiarity with the race, rather than to the effect of the modulated stimulation.

2.5.5 Stimulation Modulation Strategy

The chosen stimulation frequency range is between 4 and 8 Hz, corresponding to theta waves, which synchronize in the brain during states of relaxation and deep sleep [25]. In both modes, stimulation starts at 6 Hz, representing the center of the stimulation band.

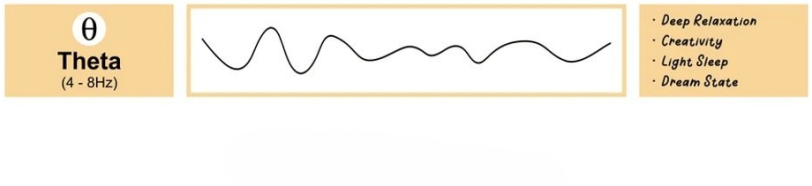


Figure 25: Theta Wave

During the first 8 time windows of 4 seconds each, the user's average stress is calculated. Starting from 6 Hz, the stimulation frequency is adaptively modulated along a random direction, with increments of 1.2 Hz corresponding to the variation step. After the following 8 windows, the average stress is recalculated according to the following logic:

- If stress increases, the stimulation is ineffective and the direction of the step is reversed;
- If stress decreases, it means that the optimal frequency is being approached, so the same direction is continued.

In the case that three consecutive gradient reversals occur, the variation *step* is reduced by 20%, in order to progressively refine the search for the frequency that induces relaxation in the user.

It is important to underline that the optimal frequency is highly subjective and may vary over time, since the stress perceived by the user can fluctuate due to signal noise or the alternation of the two psycho-emotional states. Consequently, reductions of the step or gradient reversals do not always occur: the system is designed to adapt dynamically and guarantee a personalized modulation of stimulation throughout the entire task.

2.6 Statistical analysis

All the results obtained from the acquisitions were subsequently analyzed using two non-parametric ANOVA-based methods. The first one is the Friedman test, a non-parametric statistical test particularly suited for detecting statistically significant differences among 3 or more, repeated measurements collected from the same subjects. This test is commonly adopted when the assumptions of the corresponding parametric test, namely the repeated-measures ANOVA, are not satisfied.

The main assumptions of repeated-measures ANOVA are data sphericity and homogeneity of variances across experimental conditions, which were considered too restrictive for the type of experiment conducted in this study.

The Friedman test is based on two hypotheses: the null hypothesis H_0, which states that there are no statistically significant differences among the median

ranks of the groups and that any observed differences are due to chance; and the alternative hypothesis H_1, which states that at least one group exhibits a median rank that differs significantly from the others [26].

In practice, the Friedman test compares whether there is a global difference between the extracted feature values of the mathematical task, training, Play 1, and Play 2 conditions. However, it is important to understand that the Friedman test does not describe which conditions show significant differences. It only helps identify whether there is a significant difference.

The result obtained from the Friedman test performed in MATLAB takes the form of *p-value*, and values below 0.05 are those in which statistical differences exist among the conditions being analyzed.

At this point, when a *p-value* lower than 0.05 is detected and a statistically significant difference is identified, it is necessary to determine where this difference occurs. For this purpose, post-hoc analyses were performed using the Wilcoxon signed-rank test.

The Wilcoxon signed-rank test is a non-parametric alternative to the paired t-test and is employed when the assumption of data normality is not satisfied. The test requires paired observations originating from the same subject. As in the Friedman test, the analysis is based on a null hypothesis H_0, stating that no significant difference exists between the paired conditions, and an alternative hypothesis H_1, stating that a statistically significant difference is present [27].

At the end of the test, a *p-value* is computed and statistical significance is conventionally reported using asterisk notation: three asterisks (***) for $p < 0.001$, two asterisks (**) for $p < 0.01$, and one asterisk (*) for $p < 0.05$.

The *Wilcoxon signed-rank test* is a post-hoc test with Holm-Bonferroni correction. The Holm-Bonferroni procedure is a step-down (or sequential) method for controlling the Family-Wise Error Rate (FWER), or the likelihood of committing a Type I Error in multiple comparison tests. It works by arranging the *p-values* in ascending order and comparing them with increasingly permissive *p-value* levels or significance levels. In contrast to the classical Bonferroni method, the Holm-Bonferroni procedure tests all possible comparisons with decreasing levels of correction, maximizing the statistical power.

For the i-th comparison after ordering the *p-values*, the adjusted *p-value* is computed as:

$$p'_{(i)} = \min\left(p_{(i)} \times (m - i + 1),\, 1\right),$$

where m denotes the total number of comparisons. Monotonicity is enforced by:

$$p'_{(i)} = \max\left(p'_{(i)},\, p'_{(i-1)}\right).$$

The null hypothesis H_0 is rejected for the i-th comparison if:

$$p'_{(i)} < \alpha,$$

with $\alpha = 0.05$.

In any case, even if the Friedman test fails to produce statistically significant results, the Wilcoxon test is applied nonetheless, and the results from this test

must be treated with extreme caution. A distinction will be made within the figures, indicating black asterisks if the Wilcoxon test was applied after the Friedman test, or red if applied without the Friedman test.

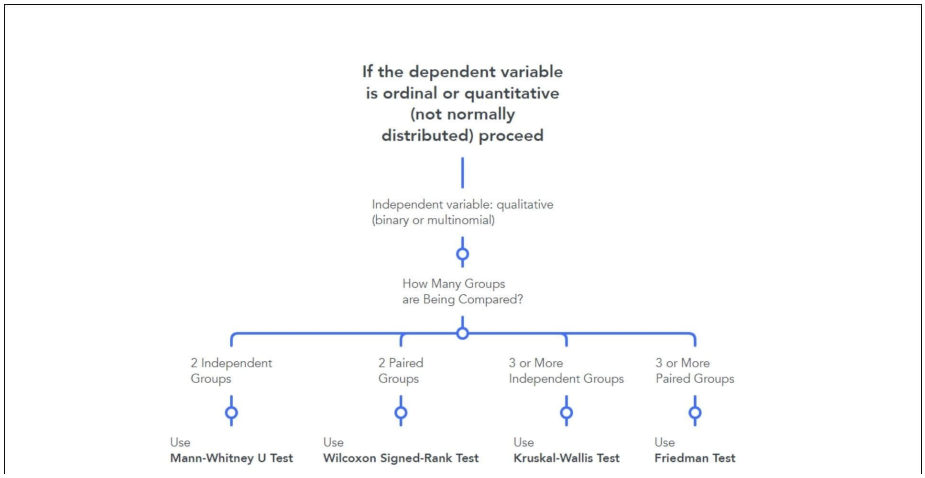
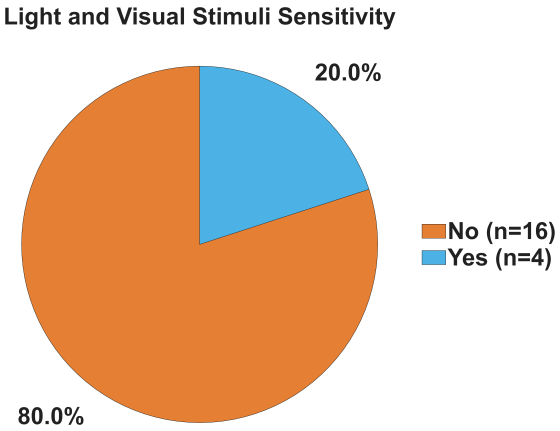


Figure 26: Statistical Analysis overview

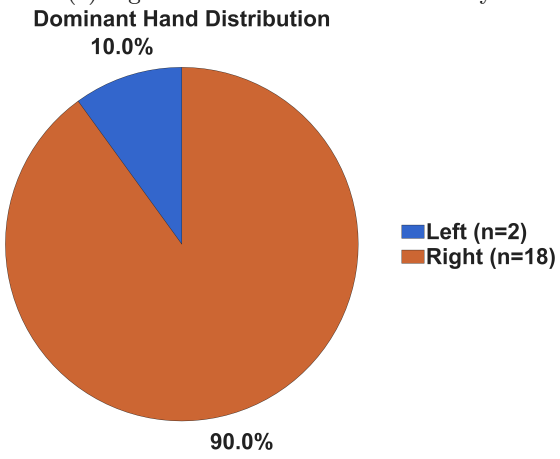
3 Results

3.1 Questionnaires

As described in Section 2.4, before the experiment each participant was asked to complete a preliminary questionnaire. The most relevant information extracted from this questionnaire is summarized below using pie charts:



(a) Light and Visual Stimuli Sensitivity

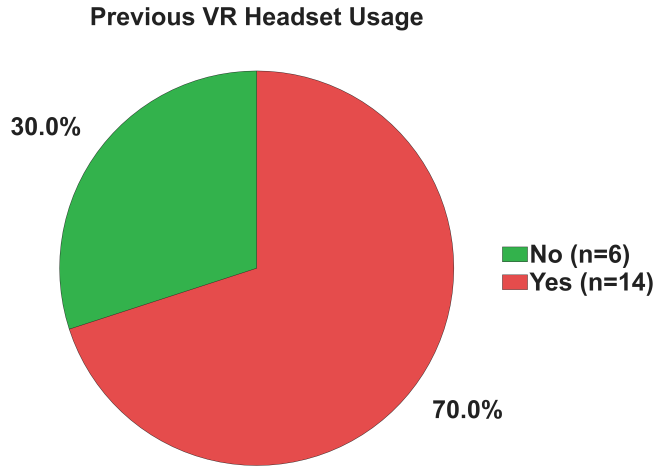


(b) Dominant Hand Distribution

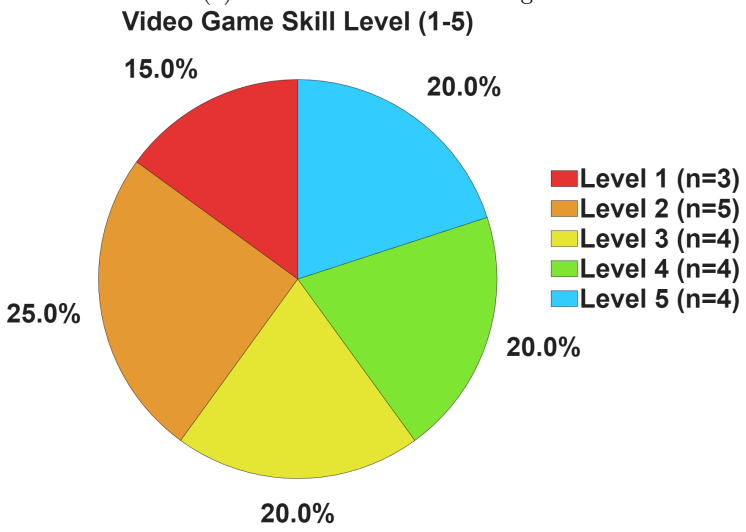
Figure 27: Participant sensitivity and hand dominance characteristics

Figure 27.a represents a pie chart showing the distribution of subjects who are sensitive to light and repetitive stimuli, serving only to understand whether the experiment causes problems for those subjects who are susceptible to these stimuli. Figure 27.b shows that there are 18 right-handed subjects and 2 left-handed subjects, which is an important indication since the experiment is per-

formed with the non-dominant hand.



(a) Previous VR Headset Usage

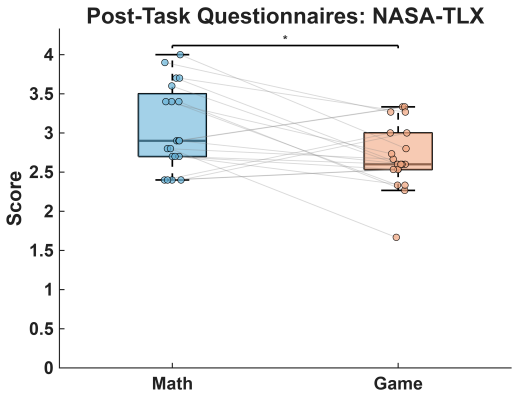


(b) Video Game Skill Level (1-5)

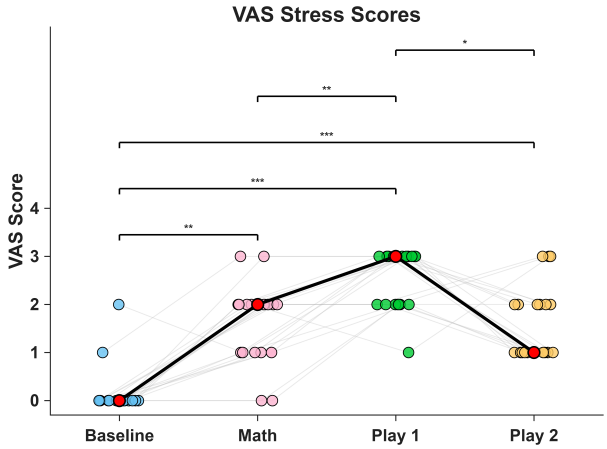
Figure 28: Participant gaming experience and VR familiarity

Figure 28.a shows how many subjects had already used VR, finding that 16 had previously used VR and 6 had not, and then Figure 28.b shows the skill level from 1 to 5 (1 being low and 5 being high) that each user believes they have. This information will be used later to see if there is a correlation between subjective skill level and the effectiveness of the adapted multimodal stimulus. During the experiment, Visual Analogue Scales (VAS) were administered between one condition and the other in order to continuously monitor the perceived stress level of each participant. At the end of the Mathematical Task (MAT) and after

the entire VR gaming experience, participants completed the NASA Task Load Index (NASA-TLX) questionnaires, which were used to assess the cognitive workload associated with the two stressful tasks. The NASA-TLX scores were computed according to the procedures described in the previous sections.



(a) NASA-TLX results. * indicates p-values between 0.05 and 0.01, ** between 0.01 and 0.001, and *** $p < 0.001$.



(b) VAS results. * indicates p-values between 0.05 and 0.01, ** between 0.01 and 0.001, and *** $p < 0.001$.

Figure 29: Comparison of NASA-TLX and VAS results across experimental conditions

What the NASA-TLX shows in Figure 29.a is that the mathematical task is perceived as significantly more cognitively demanding compared to the game, but as seen in Figure 29.b, the mode with constant stimulation (Play 1) is perceived as more stressful than the mathematical task, while Play 2 proves to be less stressful. Therefore, what is seen in the NASA-TLX boxplot for the

game is simply an average behavior between Play 1 and Play 2, which makes it appear significantly less stressful than the mathematical task.

Figure 30: Play 1(Stimolazione costante) vs Play 2(Stimolazione modulata)

Figure 30 shows a pie chart that further illustrates, at the individual level, which experimental modality was perceived as the most stressful. Seventeen subjects reported Play 1 as the most stressful condition, whereas three subjects identified Play 2 as the most stressful: Subject 1 belonging to Group A and Subjects 4 and 8 belonging to Group B.

3.2 ECG Features Analysis

In this section, the results of the electrocardiographic features are presented. First, attention is focused on the most significant feature, namely heart rate expressed in beats per minute (BPM). Subsequently, the trends of the features normalized with respect to the baseline are analyzed.

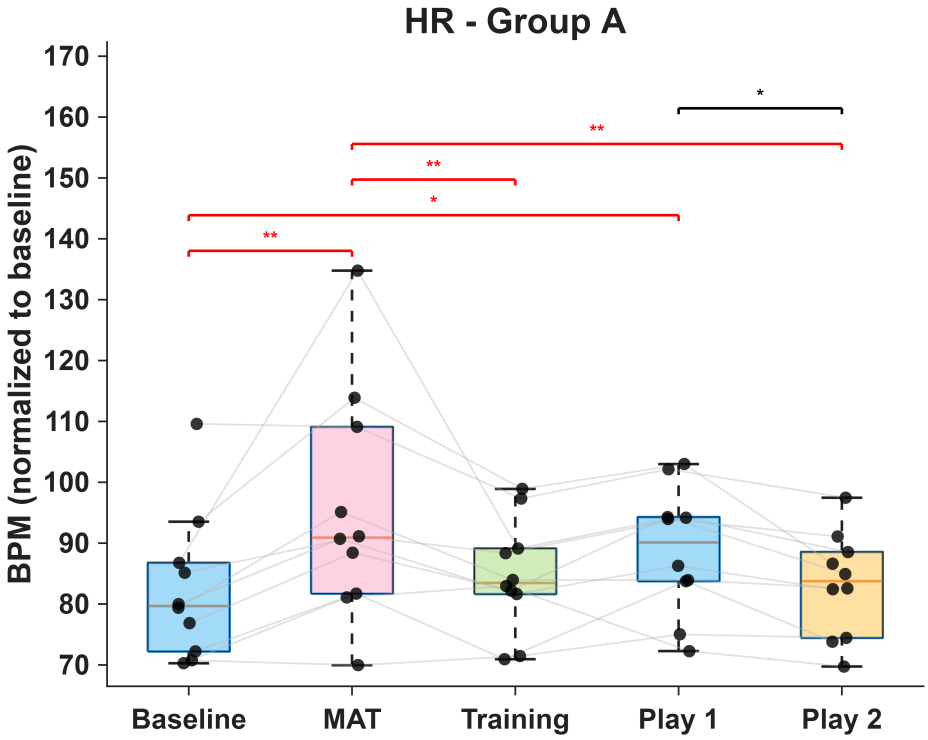


Figure 31: BPM in Group A across all experimental phases. * indicates p-values between 0.05 and 0.01, ** between 0.01 and 0.001. Black segments and asterisks indicate that the Wilcoxon test was applied after the Friedman test, while red segments and asterisks indicate that it was applied without the Friedman test.

What can be seen in Figure 31 is that, at a global level, the Friedman test concluded that the overall difference in HR is localized between Play 1 and Play 2, while the Wilcoxon test applied a priori found significant differences, some clearly visible, such as between Baseline and MAT, marked with **, and between Baseline and Play 1, marked with *. Other differences are less evident, but they allow us to state that Baseline, Training, and Play 2 do not show statistically significant differences in HR, and similarly, MAT and Play 1 do not.

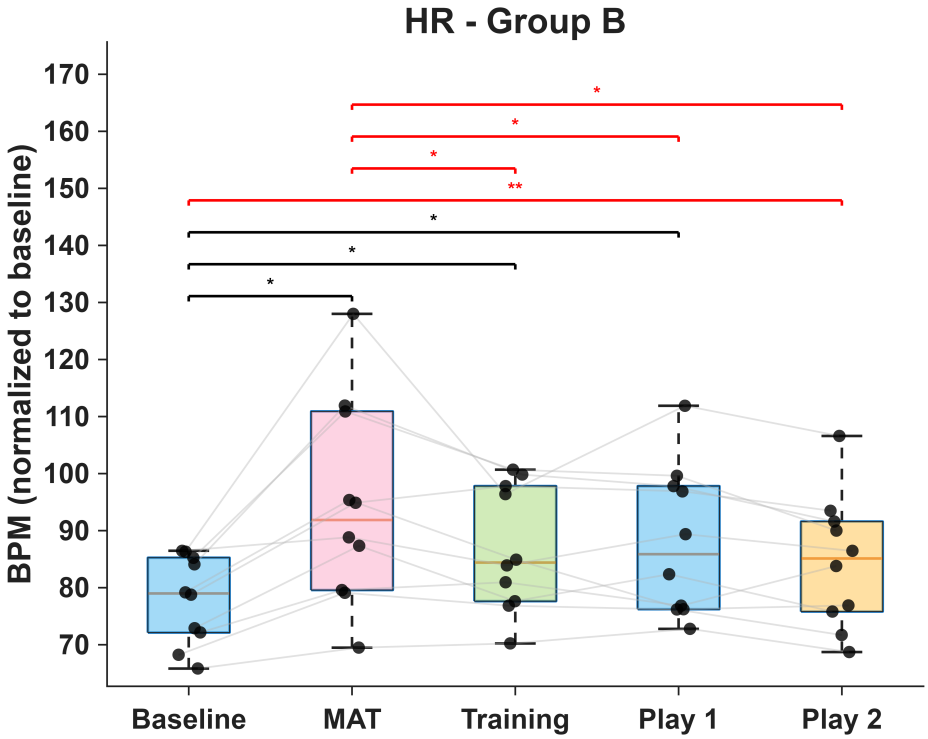


Figure 32: BPM in Group B across all experimental phases.* indicates p-values between 0.05 and 0.01, ** between 0.01 and 0.001. Black segments and asterisks indicate that the Wilcoxon test was applied after the Friedman test, while red segments and asterisks indicate that it was applied without the Friedman test.

.

In Figure 32, the Friedman test with Wilcoxon post hoc indicates that the difference is localized between Baseline–MAT, Baseline–Training, and Baseline–Play 1, suggesting no statistically significant difference with Play 2. On the other hand, the Wilcoxon test applied a priori indicates a much more significant difference between Baseline and Play 2 than the Wilcoxon test applied after Friedman, which reports no difference. Therefore, the a priori Wilcoxon result should not be considered.

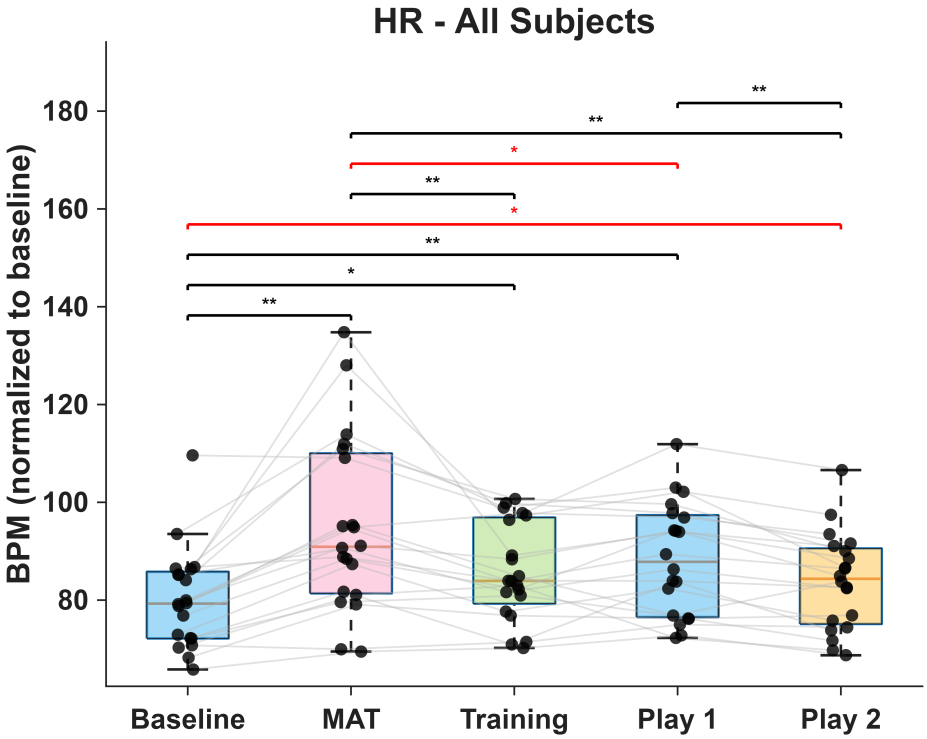


Figure 33: BPM across all subjects during all experimental phases.* indicates p-values between 0.05 and 0.01, ** between 0.01 and 0.001. Black segments and asterisks indicate that the Wilcoxon test was applied after the Friedman test, while red segments and asterisks indicate that it was applied without the Friedman test.

In Figure 33, the previously observed differences in the two preceding figures are largely confirmed. It is worth noting that this analysis was performed across all subjects, and Play 1 and Play 2 are significantly different, especially with HR being higher in Play 1 compared to Play 2, consistent with the subjects being more stressed during Play 1, as also shown in Figure 30 in which the subjects expressed their subjective opinion.

3.2.1 Features trend analisys

To analyze the trend of the features across the different experimental phases, each feature mean value of each subject computed for each segment was averaged and normalized with respect to the corresponding baseline mean value, defined as the phase in which the user watched the relaxing video. Baseline normalization was applied only to HRV features (BPM, SDNN, pNN50, RMSSD), as they capture short-term autonomic modulation. ECG temporal features (QT interval, RT_r, RP_r), which exhibit limited intra-subject variability and reflect more stable electrophysiological properties, were therefore excluded from this normalization procedure.

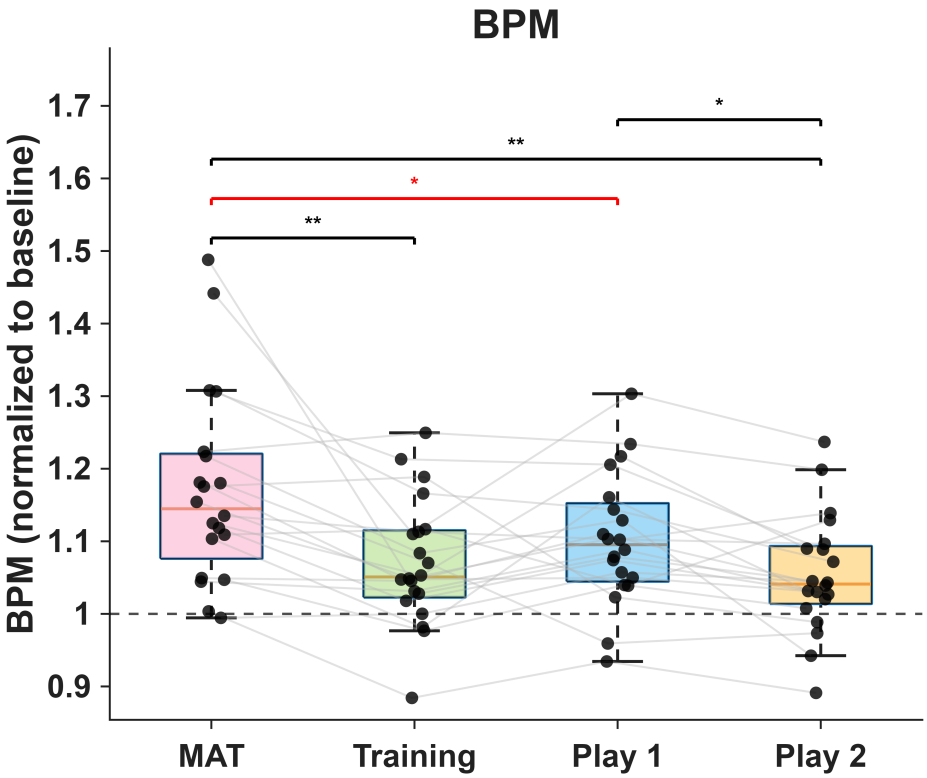


Figure 34: Feature Ratio BPM. * indicates p-values between 0.05 and 0.01, ** between 0.01 and 0.001.Black segments and asterisks indicate that the Wilcoxon test was applied after the Friedman test, while red segment and asterisc indicate that it was applied without the Friedman test.

In Figure 34, the Friedman test with Wilcoxon post hoc highlights a significant difference between MAT and Training, and between MAT and Play 2, marked with **, as well as a difference marked with * between Play 1 and Play 2. This suggests a behavior similar to that observed in Figure 33. Additionally,

the difference between MAT and Play 1 highlighted by the a priori Wilcoxon test appears inconsistent and should therefore be interpreted with caution.

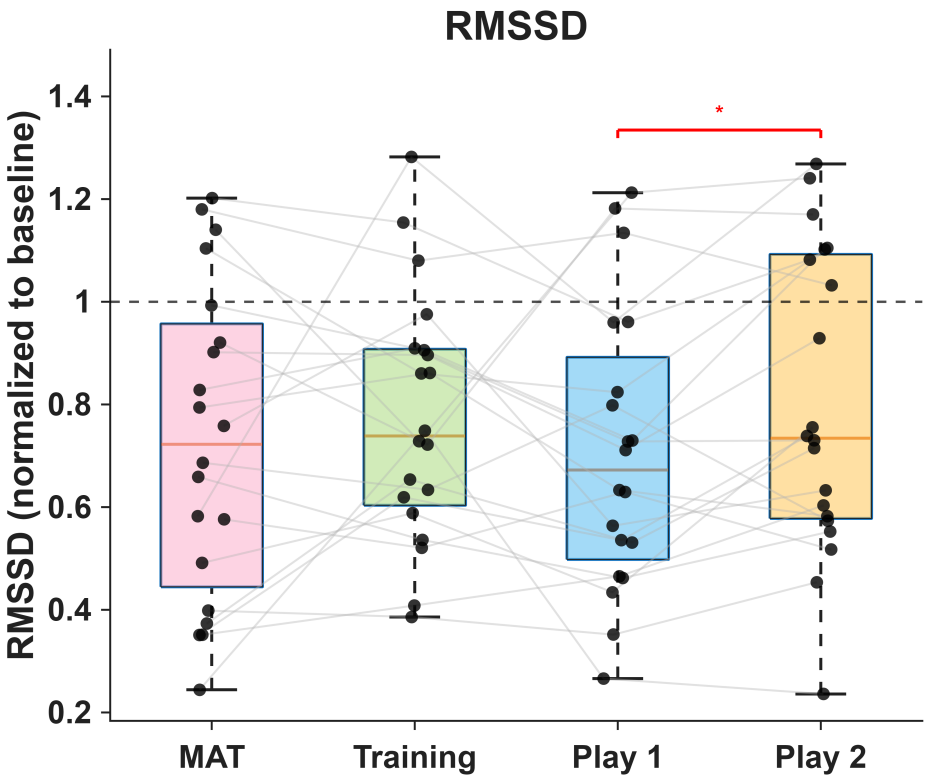


Figure 35: Feature Ratio RMSSD.* indicates p-values between 0.05 and 0.01. Red segment and asterisk indicate that it was applied without the Friedman test.

In this figure, the Friedman test did not identify any global differences. However, the a priori Wilcoxon test highlighted relevant information: considering that Figure 34 showed a significant difference between Play 1 and Play 2 in BPM, a significant difference in RMSSD between Play 1 and Play 2 is consistent with an autonomic nervous system response to a stress-inducing event. In this context, a lower RMSSD can be interpreted as an indication of stress, as explained in Table 2.

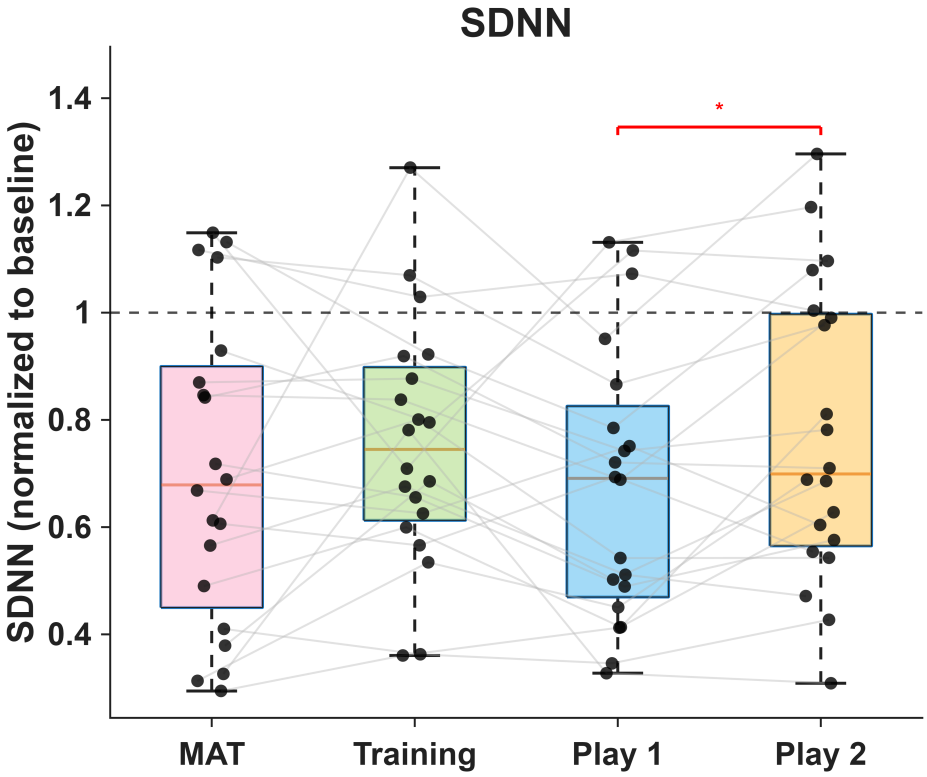


Figure 36: Feature Ratio SDNN.* indicates p-values between 0.05 and 0.01. Red segment and asterisk indicate that it was applied without the Friedman test

Similarly to Figure 35, Figure 36 shows the same behavior, which can also be attributed to the strong correlation between RMSSD and SDNN.

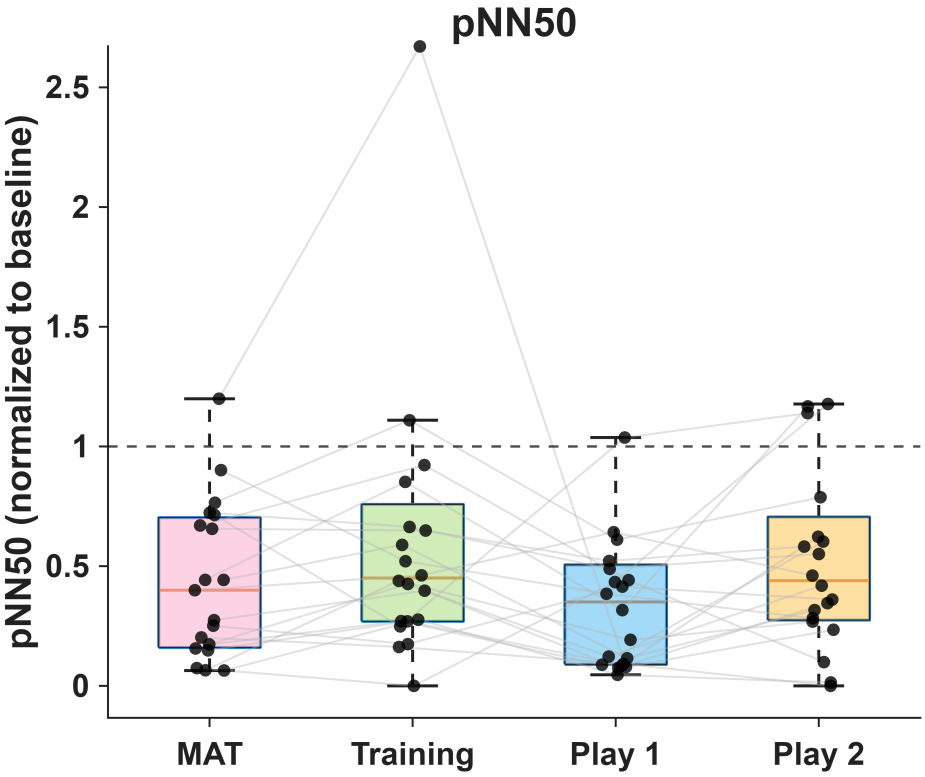


Figure 37: Feature Ratio pNN50.

In Figure 37, neither the Friedman test nor the a priori Wilcoxon test highlights any differences.

3.3 Regressor analisys and results

To evaluate the performance of the regressor, two metrics were considered, namely the coefficient of determination R^2 and the Root Mean Square Error (RMSE), as described in detail in Section 2.5.2. For each subject and for each model, a 10-fold cross-validation procedure was applied. The average R^2 value across the folds was used as an indicator of the model's ability to explain data variance (VAF), thus providing a measure of the overall quality and stability of the regressor.

Model selection was instead performed by choosing the model exhibiting the highest R^2 value. In this framework, the average R^2 was used to assess how well the model explains the variance of the data, while the best R^2 value was adopted as the criterion for model selection.

Since the Random Forest algorithm allows the estimation of feature importance, and given that eight input features were available, a ranking of the features based on their relative importance was generated. The selected model was then retrained using only the top five most relevant features.

The following figures illustrate how the mentioned criteria were applied to all subjects, taking Subject 14 as a representative example.

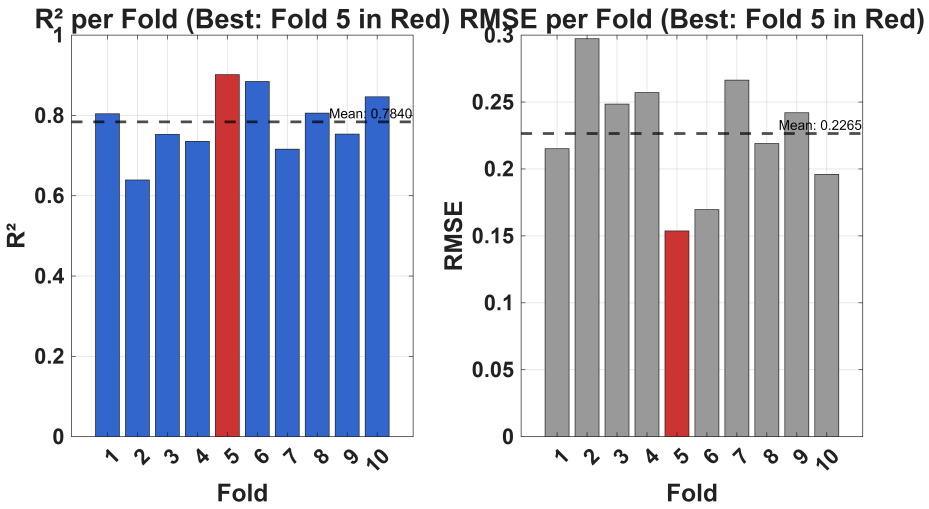


Figure 38: R^2 and RMSE values for each fold

In Figure 38, the red bars indicate the best-performing fold, namely the one characterized by the highest R^2 value. The dashed line represents the average R^2 computed across all folds, which is used to evaluate the overall quality and stability of the model.

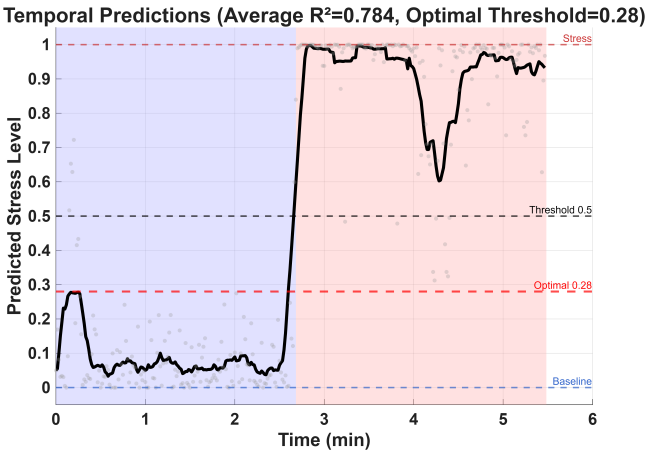


Figure 39: Temporal prediction during the training phase

Figure 39 shows how each 4-second window was classified by the regressor during the training phase. The blue background on the left indicates the baseline condition (relaxing video viewing), whereas the red background on the right corresponds to the MAT phase. Two dashed horizontal lines are shown at values 0 and 1, representing the labels for non-stress and stress conditions, respectively. The solid black line represents the smoothed prediction trend over time, while the dashed black line indicates the default threshold used for classification. The dashed red line corresponds to the optimal threshold computed to maximize the regressor's accuracy.

The optimal threshold was not used during the inference phase and was introduced exclusively for subsequent data analysis purposes.

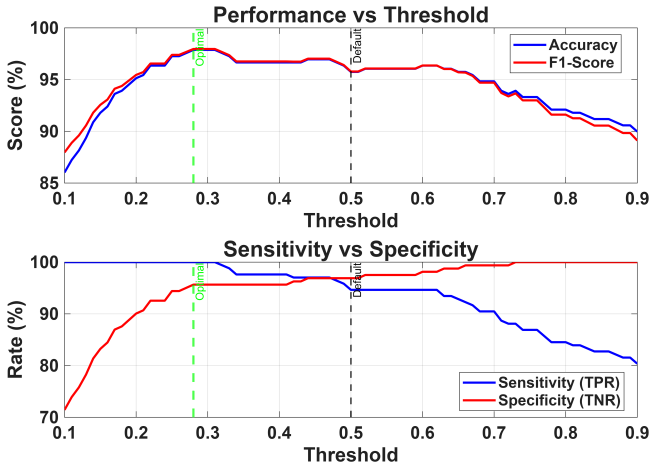


Figure 40: Threshold Analysis

Figure 40 presents the scheme followed for the selection of the optimum threshold. The threshold value that corresponds to the peak in classification accuracy is chosen. Other performance metrics, such as the F1-score, sensitivity, and specificity, are also presented as supplementary measures to better describe the behavior of the classifier. .

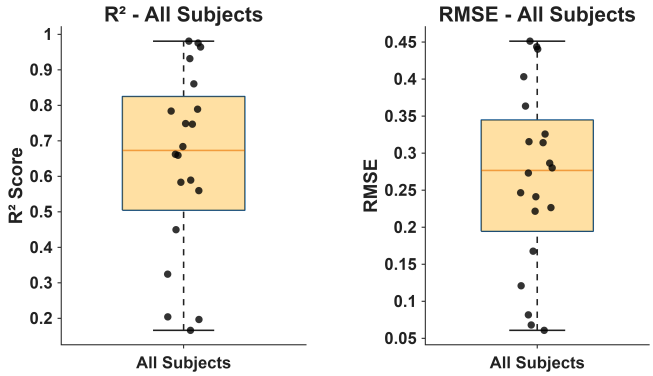


Figure 41: R^2 and RMSE overall

Figure 41 shows the distribution of the average R^2 values and the corresponding RMSE. A median R^2 value of approximately 0.7 is observed, with three subjects exhibiting an average R^2 lower than 0.25. Overall, these results indicate that the model performs well for the majority of subjects, particularly for those in whom an increase in heart rate is observed between the baseline and MAT phases.

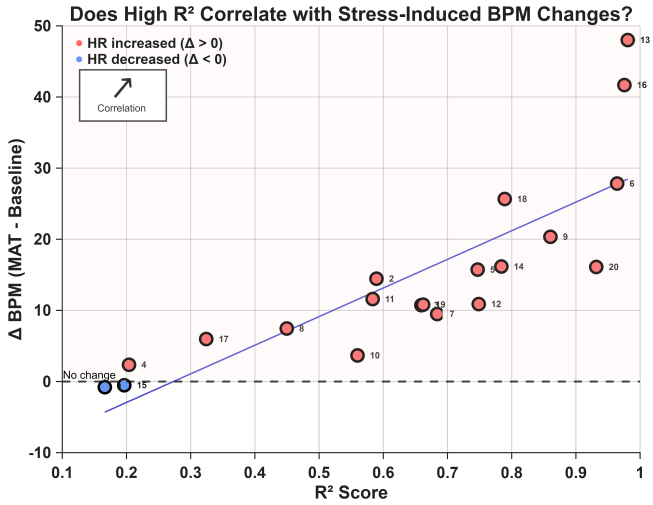


Figure 42: R^2 correlation with delta BPM

Figure 42 illustrates that an increase in the heart rate delta between the MAT and baseline phases is associated with an increase in the R^2 value, indicating an improvement in the model's ability to explain data variance. In the upper-left corner, an arrow indicates the direction of the correlation, computed using Spearman's rank correlation coefficient.

Spearman's correlation was selected as it is particularly suitable for assessing monotonic relationships between variables, even in the absence of linearity, and provides a robust measure of the dependence of one variable on another.

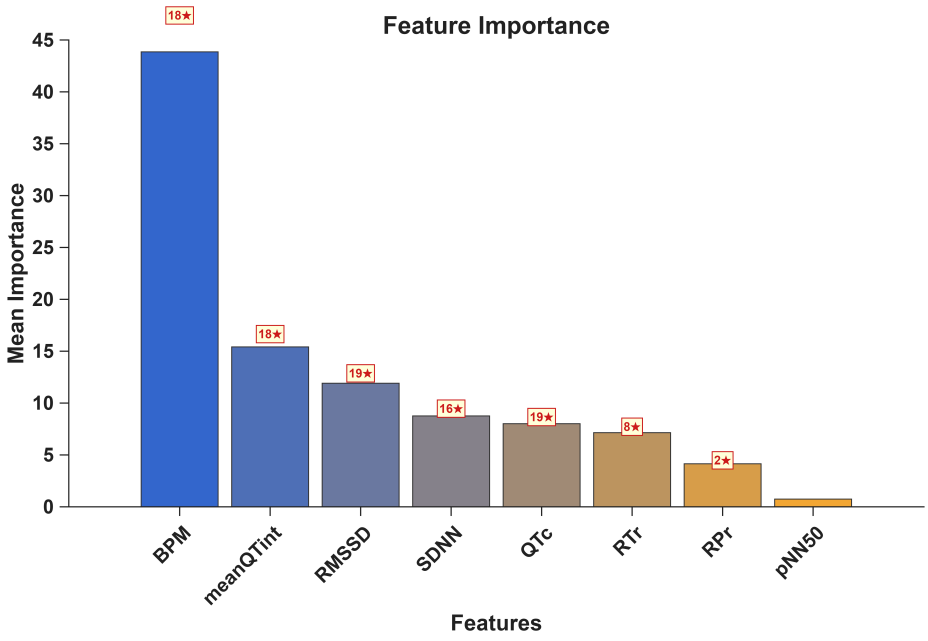


Figure 43: Feature importance

Figure 43 illustrates the overall importance of the features extracted from the ECG. Above each bar in the bar chart, a number accompanied by a star indicates how many times that feature ranked among the top five. As can be observed, heart rate (BPM) emerges as the most important feature in terms of percentage contribution.

3.4 Stimulation results

This section presents the results related to the stress predicted by the regression model during the gameplay phase for each participant under the two gaming conditions: Play 1 with constant stimulation and Play 2 with adaptive stimulation. These results aim to determine which of the two modalities was more effective in mitigating stress, allowing an assessment of the stimulation strategy that proved to be the most effective.

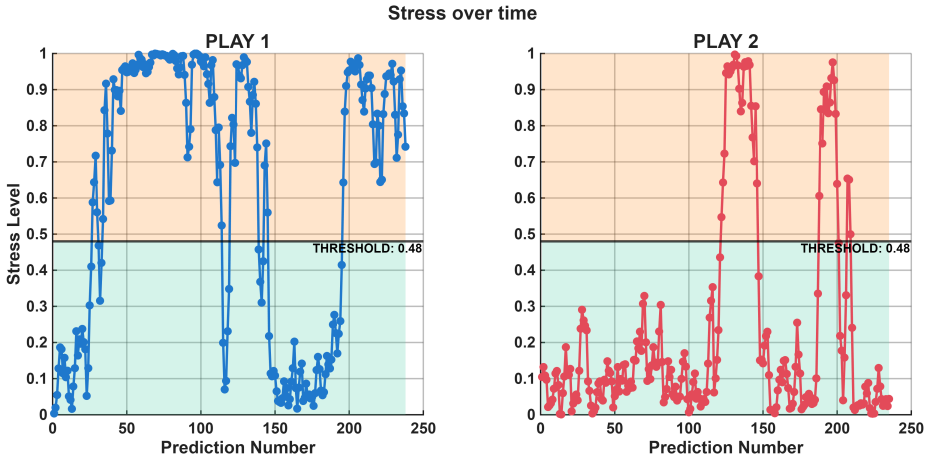


Figure 44: Stress over time in Play 1 and Play 2 modes for Subject 7

Before analyzing which stimulation strategy resulted to be the most effective, it is necessary to present two figures showing the regressor predictions over time for the two stimulation modalities and how the stimulation frequency varied over time in the Play 2 mode. The following two figures refer to Subject 7.

Figure 44 shows the temporal evolution of the stress coefficients (S). The stress threshold is also reported, which for Subject 7 is equal to 0.48. On the x-axis, the prediction index is reported, corresponding to one prediction per second, considering a window length of 4 seconds with a 75% overlap. On the y-axis, the stress value S predicted by the regressor is shown.

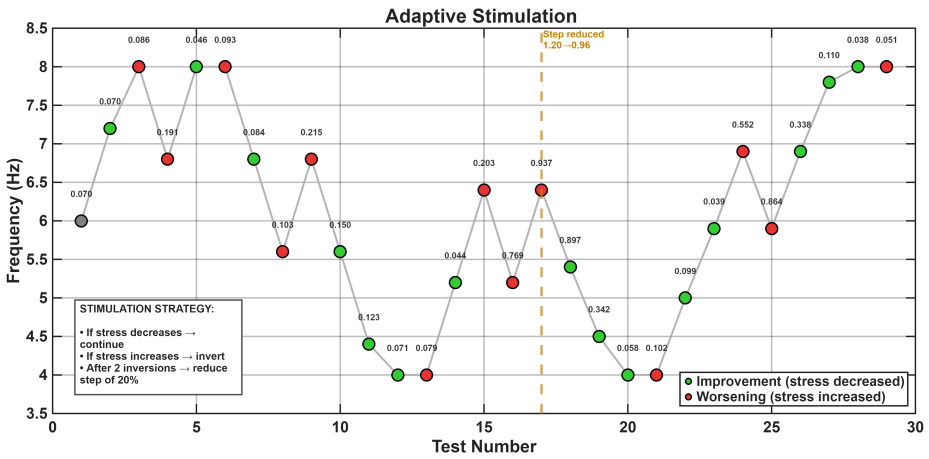
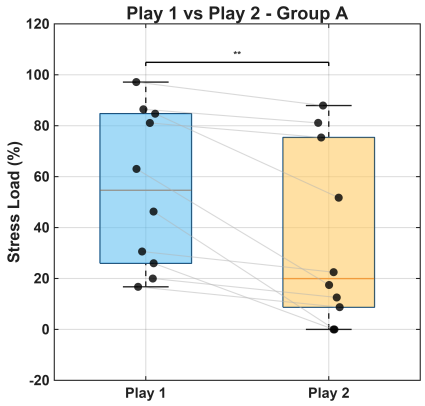


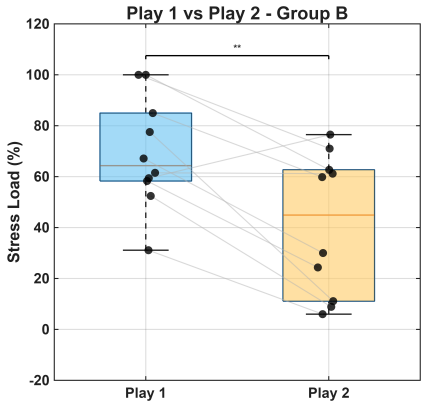
Figure 45: Evolution of stimulation frequency during adaptive stimulation for Subject 7

Figure 45 shows how the stimulation frequency varies for Subject 7 as a function of the average stress computed over 8 consecutive windows (corresponding to 8 seconds). As previously described in Section 2.5.5, the stimulation starts at 6 Hz, and the frequency is then varied by 1.2 Hz in a random direction. If the average stress decreases, the frequency continues to change in the same direction; conversely, if the stress increases, the direction is inverted. After two consecutive inversions, the step size is reduced by 20%, as also reported in the annotation box in the figure. The x-axis represents the test number (one test every 8 seconds), while the y-axis shows the stimulation frequency.

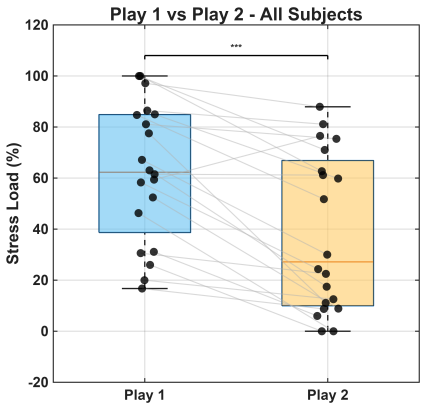
The following figures show which stimulation modality resulted to be more effective, first at the group level and then from a global perspective.



(a) Stress Load – Group A



(b) Stress Load – Group B



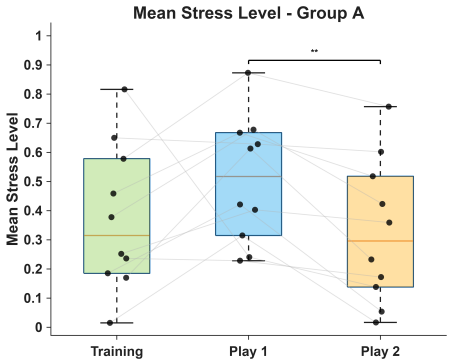
(c) Play 1 vs Play 2 – Aggregate

Figure 46: Comparison of stress load across groups and global overview. ** indicates p-values between 0.01 and 0.001, while *** indicates $p < 0.001$.

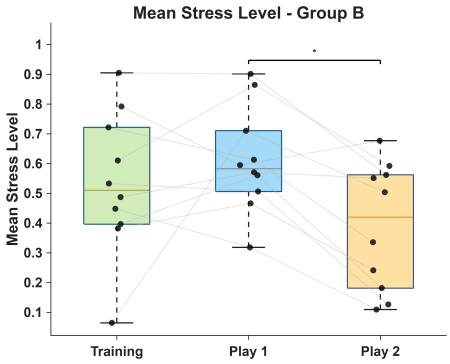
Figures 46.a and 46.b each show two boxplots corresponding to the two stimulation modalities, where each point represents the stress load measured for a given modality. At this stage, the optimal threshold introduced in Figure 40 is applied: all stress values S above the threshold are assigned a value of 1, whereas values below the threshold are assigned a value of 0. The stress load is then computed as the percentage of ones over all predictions for each modality.

The results show that, for both groups, the Wilcoxon test identified a statistically significant difference (**) between Play 1 and Play 2. Specifically, Play 1 exhibits a higher median stress load, with values of approximately 55% for Group A and 65% for Group B, compared to median values of about 20% for Group A and 45% for Group B in Play 2. For both groups, Play 1 resulted to be the more stressful stimulation modality.

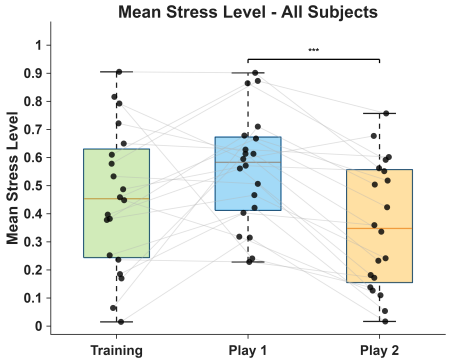
This outcome is further confirmed by Figure 46.c, which highlights the same statistically significant difference at the global level (***).



(a) Mean Stress Level



(b) Mean Stress Level – Group B



(c) Mean Stress Level – Aggregate

Figure 47: Mean stress level comparison across groups and global overview.* indicates p-values between 0.05 and 0.01, ** between 0.01 and 0.001, while *** indicates $p < 0.001$

In Figure 47, the comparison of the mean stress levels is shown for Group A, Group B, and for all subjects. Unlike Figure 46, the training modality—intended as a preliminary trial round representing the user's familiarization phase with the game—has also been included. Although it does not represent an experimental condition, this phase allows for a clearer observation of the evolution of the mean stress level for each subject, from the familiarization phase to the two stimulation modalities.

Specifically, in Figure 47.a, the Friedman test and the Wilcoxon post-hoc test highlight a statistically significant difference (**) between Play 1 and Play 2, similarly to the stress load results reported in Figure 46.a. In Figure 47.b, a statistically significant difference (*) is observed, although less pronounced than the one reported for stress load in Figure 46.b. Finally, Figure 47.c shows the global behavior of the mean stress level across all subjects, confirming a highly significant statistical difference (***) consistently with the results reported in Figure 46.c.

The a priori Wilcoxon test yielded the same outcome as the post-hoc Wilcoxon test.

The final result indicates that adaptive stimulation is more effective than constant-frequency stimulation.

The next figure is intended to investigate whether, for some subjects, the ineffectiveness of the stimulation could be attributed to a low R^2 of the regression model, which would indicate a poor understanding of the subject's state and, consequently, a less effective stimulation.

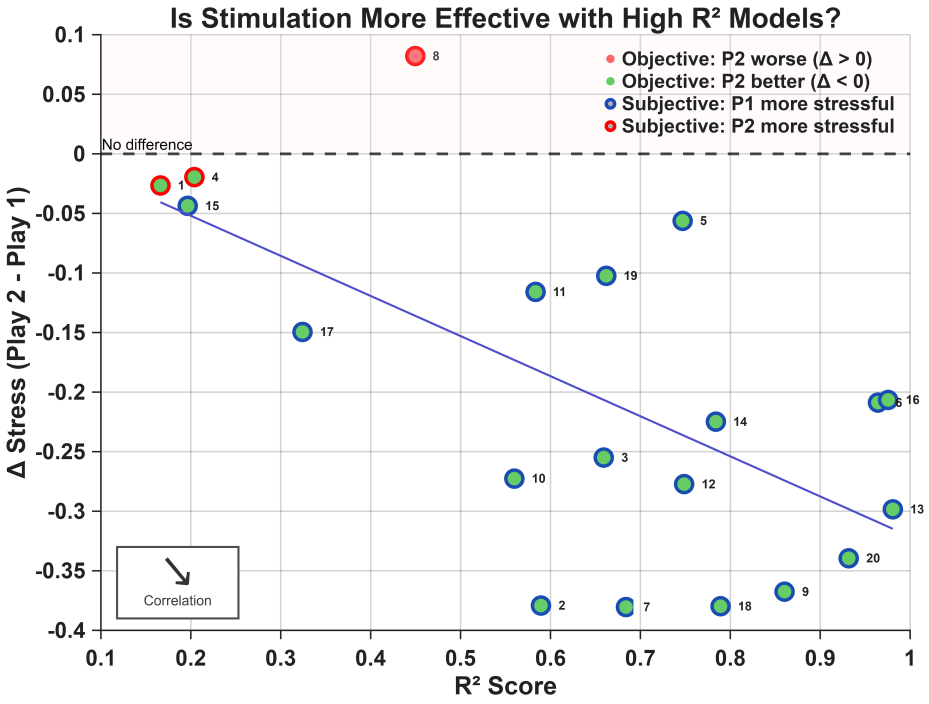


Figure 48: Is stimulation effectiveness dependent on R^2?

Figure 48 shows, on the x-axis, the R^2 of the regression model and, on the y-axis, the stress difference between Play 2 and Play 1. Red markers indicate subjects with a positive delta, while green markers indicate a negative delta. Blue borders surround markers of subjects who reported Play 1 as more stressful, whereas dark red borders indicate subjects who reported Play 2 as more stressful.

The figure highlights three subjects (1, 4, and 15) with very low R^2. In particular, for subjects 1 and 4, the model predicted higher stress during Play 1, whereas objectively they experienced more stress during Play 2. For subject 15, the situation is more nuanced: although this subject reported Play 1 (constant stimulation) as more stressful than Play 2 (adaptive stimulation), which is consistent with the model's prediction, their inclusion in the same low-R^2 group as subjects 1 and 4 indicates an inconsistent result, especially since the delta stress is very close to zero for all three cases.

Moreover, Spearman correlation analysis suggests a descending relationship between R^2 and stimulation effectiveness, indicating that higher R^2 values are associated with a more significant stress reduction delta.

Since the stressful environment used during the inference phase was a video game and each participant has a different gaming skill level, all subjects were asked, prior to the experiment, to rate their gaming ability on a scale from 1 to 5. This was done to investigate whether the stimulation had a different effect

depending on the subject's skill level, in order to assess a potential relationship between gaming ability and stimulation effectiveness.

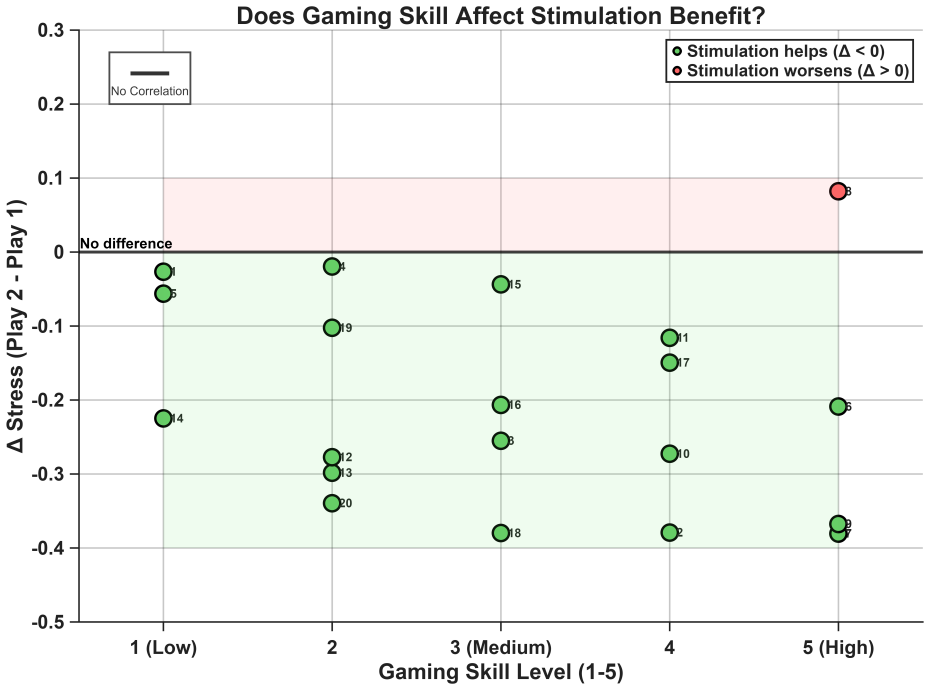


Figure 49: Is gaming skill level related to stimulation effectiveness?

Figure 49 shows the distribution of subjects as a function of their self-reported skill level, from 1 to 5. On the x-axis is the skill level, and on the y-axis is the stress delta. Spearman correlation analysis revealed no significant correlation, indicating that the stimulation effectiveness is independent of the participant's gaming ability. Thus, regardless of skill level, the efficacy of the stimulation appears to depend solely on the accuracy of the regression model, as previously observed in Figure 48.

4 Discussion

The conducted experiment demonstrated that adaptive stimulation is indeed more effective in mitigating induced stress compared to constant stimulation. This result is supported not only by the questionnaire data but also by the outcomes based on the classifier's predictions, which show a consistent relationship between the measured stress levels and the self-reported stress.

Overall, the presented work exhibits points of strenght, while also being affected by a number of limitations.

4.1 Limits

Starting from the limitations, the first aspect to consider is that the study was conducted on a sample of 20 subjects. Although the statistical tests applied had been able to point out, with a certain degree of determinism, a statistically significant difference between the two stimulation modalities, the sample size is limited and does not fully ensure the generalizability of the results.

Another limitation concerns inter-individual variability in the regressor response. Indeed, as shown by the R^2 value analysis, some subjects (1, 4, and 15) revealed very low performance; thus, the model could not achieve an adequate prediction of these subjects' stress levels. As shown in Figure 41, the low R^2 values are related to the fact that heart rate did not significantly vary between the baseline condition and the MAT. This observation introduces an additional limitation, namely that individuals can exhibit markedly different physiological responses when exposed to stressful conditions; for some subjects, such conditions may even be perceived as non-stressful and therefore may not induce any appreciable alteration in heart rate.

The fact that a single sensor is employed could be considered a limitation as well. Usually, this will impose a restriction on the number of features that can be obtained, and it is not always guaranteed to be informative enough, especially considering the windowing strategy itself. The integration of more signals, such as EEG (electroencephalography) or EDA (electrodermal activity), would have helped the system to be more efficient by increasing the number of features, hence improving R^2 performance and efficacy of this stimulation. However, this would have come at the cost of increased computational complexity, both during model training and real-time inference, as well as a longer and more demanding experimental setup. In particular, mounting an EEG cap is time-consuming and intrusive compared to wearing a chest strap, which typically requires no more than two minutes.

The choice of a 4-second window with a 75% overlap, although representing a sweet spot for real-time predictions, does not allow for a robust computation of HRV features. Therefore, if real-time stress prediction is to be maintained, these features should be computed over longer windows and used as secondary features. This approach would be feasible provided that a sufficient number of features is extracted from additional physiological signals.

Another limitation of the study is that it relied on multimodal stimulation

which, although it produced convincing results, did not allow us to disentangle the individual contribution of each stimulation modality or to determine whether the combined use was more effective than the single modalities.

4.2 Strengths of the Experiment

One of the main strengths of the experiment is the implementation of a closed-loop system capable of dynamically modulating the stimulation based on the stress state measured through the ECG signal. Unlike many studies that are limited to offline classification or open-loop approaches, this work demonstrates the feasibility of a real-time framework in which the stimulation is continuously adapted in an automatic and non-invasive manner.

Yet another strength is the developed system architecture that can seamlessly integrate the Polar H10 device with any software environment, including Python and MATLAB, as it uses TCP/IP as the medium of interaction between the two. In such an architecture, Python is used for the acquisition of the ECG signal, while signal processing is carried out in the MATLAB environment, all within the context of the virtual reality environment developed using the Unity framework. With such an architecture, the latency is ensured to be less than 5 seconds between the two events.

Splitting the group into two groups, i.e., group A and group B, helped to properly validate the effectiveness of both conditions of stimulation by eliminating bias that may have occurred as a result of learning.

It is evident that the multimodal stimulation approach, incorporating both visual and tactile stimuli, is an innovative feature of the proposed research. Furthermore, the majority of proposed studies rely on the binaural beats (BB) and consequently auditory stimuli. On the other hand, as mentioned in the previous section, the exact contribution of the individual modality of stimulation remains unclear.

4.3 Future Developments

A first future development involves increasing the sample size in order to enhance the statistical power of the study and improve the generalizability of the results.

It would also be appropriate to extend the study to specific clinical populations, such as individuals affected by anxiety disorders, ADHD, or post-traumatic stress disorder (PTSD), in order to verify whether adaptive stimulation produces effects comparable to those observed in healthy subjects and to assess its potential use as a therapeutic intervention.

Another possible future development concerns the inclusion of additional physiological signals, such as EEG or EDA/GSR, in order to better define and differentiate primary features (e.g., BPM) from secondary features such as SDNN, RMSSD, and pNN50, which should be computed over time windows longer than 4 seconds.

Further developments may include testing real-time adaptive stimulation in different VR scenarios, such as public speaking simulations similar to the Trier

Social Stress Test (TSST), exposure to specific phobias such as claustrophobia, or emergency situations such as a building on fire.

Moreover, VR environments specifically designed for relaxation and meditation could be developed, in which adaptive stimulation is not used to mitigate induced stress but rather to enhance a state of well-being, with the potential for future applications in clinical settings.

4.4 Conclusions

In conclusion, this thesis demonstrates the feasibility and effectiveness of an integrated real-time stress detection system based on ECG signals and adaptive multisensory stimulation within a virtual reality environment. The obtained results confirm that stimulation modulated according to the subject's stress state is significantly more effective than constant-frequency stimulation in reducing both subjectively perceived and objectively measured stress.

Bibliografy

[1] Brian Chu, Kanwaljit Marwaha, and Tomas Sanvictores. "Physiology, Stress Reaction". In: *StatPearls*. Updated May 7, 2024. Treasure Island (FL): StatPearls Publishing, 2025. URL: https://www.ncbi.nlm.nih.gov/books/NBK541120/ (visited on 01/15/2025).

[2] X. Guo et al. "The HPA and SAM axis mediate the impairment of creativity under stress". In: *Psychophysiology* 61.3 (2024). Epub 2023 Nov 15, e14472. DOI: 10.1111/psyp.14472.

[3] Matteo Raggi, Liam J. Moore, and Luca Mesin. "Inducing and Assessing Acute Mental Stress in Controlled Conditions: Topical Review and Guidelines for Effective Experimental Protocols". In: *IEEE Access* 13 (2025), pp. 10022–10042. DOI: 10.1109/ACCESS.2025.3528518.

[4] L. D. Godoy et al. "A Comprehensive Overview on Stress Neurobiology: Basic Concepts and Clinical Implications". In: *Frontiers in Behavioral Neuroscience* 12 (2018), p. 127. DOI: 10.3389/fnbeh.2018.00127. URL: https://www.frontiersin.org/journals/behavioral-neuroscience/articles/10.3389/fnbeh.2018.00127/full.

[7] Anitha S. Pillai and Prabha Susy Mathew. "Impact of Virtual Reality in Healthcare: A Review". In: *Virtual and Augmented Reality: Concepts, Methodologies, Tools, and Applications*. IGI Global, 2019. Chap. 2. DOI: 10.4018/978-1-5225-7168-1.ch002.

[10] Sermo Team. *Ecco come la VR nella formazione medica trasforma il trattamento dei pazienti*. https://www.sermo.com/it/resources/ecco-come-la-vr-nella-formazione-medica-trasforma-il-trattamento-dei-pazienti/. Accessed: 2025-01-01. 2025.

[11] Mattia Miccoli and Redazione Santagostino. *Riabilitazione Cognitiva e realtà virtuale: benefici e applicazioni*. https://www.santagostino.it/magazine-psiche/realta-virtuale-riabilitazione-cognitiva/. Accessed: 2026-01-01. 2025.

[12] "Biological and psychological markers of stress in humans: Focus on the Trier Social Stress Test". In: *Neuroscience & Biobehavioral Reviews* 38 (2014), pp. 94–124. ISSN: 0149-7634. DOI: https://doi.org/10.1016/j.neubiorev.2013.11.005. URL: https://www.sciencedirect.com/science/article/pii/S0149763413002728.

[13] Katarina Dedovic et al. "The Montreal Imaging Stress Task: using functional imaging to investigate the effects of perceiving and processing psychosocial stress in the human brain". In: *Journal of Psychiatry & Neuroscience* 30.5 (2005), pp. 319–325.

[14] Fatimah Abdul Hamid, M. Naufal M. Saad, and Aamir Saeed Malik. "Characterization stress reactions to stroop color-word test using spectral analysis". In: *Materials Today: Proceedings* 16 (2019). Conference on Biomedical and Advanced Materials, 28-29 November 2017, pp. 1949–1958. ISSN: 2214-7853. DOI: https://doi.org/10.1016/j.matpr.2019.06.073. URL: https://www.sciencedirect.com/science/article/pii/S2214785319313057.

[15] Manjeet Yadav and Nilesh Kumar Sahu. "Understanding Stress: A Web Interface for Mental Arithmetic Tasks in a Trier Social Stress Test". In: *arXiv preprint arXiv:2403.10356* (2024). Accessed: 2025-01-02. URL: https://arxiv.org/abs/2403.10356.

[16] IBM. *Support Vector Machine (SVM)*. https://www.ibm.com/it-it/think/topics/support-vector-machine. Accessed: 2026-01-03. 2026.

[17] E. Y. Boateng, J. Otoo, and D. A. Abaye. "Basic Tenets of Classification Algorithms K-Nearest-Neighbor, Support Vector Machine, Random Forest and Neural Network: A Review". In: *Journal of Data Analysis and Information Processing* 8 (2020), pp. 341–357. DOI: 10.4236/jdaip.2020.84020. URL: https://doi.org/10.4236/jdaip.2020.84020.

[20] Malcolm Koo and Shih-Wei Yang. "Visual Analogue Scale". In: *Encyclopedia* 5.4 (2025). ISSN: 2673-8392. DOI: 10.3390/encyclopedia5040190. URL: https://www.mdpi.com/2673-8392/5/4/190.

[21] Sandra G. Hart. "NASA-Task Load Index (NASA-TLX): 20 Years Later". In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 50. 9. Los Angeles, CA, USA: Human Factors and Ergonomics Society, Jan. 2006, pp. 904–908.

[22] Ting Wei, Weiwei Wang, and Suihuai Yu. "Analysis of the Cognitive Load of Employees Working from Home and the Construction of the Telecommuting Experience Balance Model". In: *Sustainability* 14.18 (2022). ISSN: 2071-1050. DOI: 10.3390/su141811722. URL: https://www.mdpi.com/2071-1050/14/18/11722.

[23] Anas Fouad Ahmed, Ammar A. Al-Hamadani, and Mohammed K. Al-Obaidi. "Efficient and Real-Time Approach for PQRST and Atrial Fibrillation Detection of ECG Signal". In: *2023 46th International Conference on Telecommunications and Signal Processing (TSP)*. 2023, pp. 91–96. DOI: 10.1109/TSP59544.2023.10197821.

[24] Kamana Dahal, Brian Bogue-Jimenez, and Doblas Ana. "Global Stress Detection Framework Combining a Reduced Set of HRV Features and Random Forest Model". In: *Sensors* 23.11 (2023), p. 5220. DOI: 10.3390/s23115220.

Sitografy

[5] Psicosomatica.eu. *Di cosa parliamo quando parliamo di stress*. s.d. URL: https://www.psicosomatica.eu/di-cosa-parliamo-quando-parliamo-di-stress/ (visited on 12/30/2025).

[6] My-personaltrainer.it. *Quali sono gli effetti dello stress sul corpo?* 2025. URL: https://www.my-personaltrainer.it/salute/effetti-dello-stress.html (visited on 12/30/2025).

[8] Immersive Archive. *Ivan Sutherland's Head-Mounted Display (1968)*. Accessed: 31 December 2025. 1968. URL: https://immersivearchive.org/ivan-hmd.html.

[9] Sermo. *Ecco come la VR nella formazione medica trasforma il trattamento dei pazienti*. Accessed: 1 January 2026. 2024. URL: https://www.sermo.com/it/resources/ecco-come-la-vr-nella-formazione-medica-trasforma-il-trattamento-dei-pazienti/.

[18] World Wide Web Consortium (W3C). *Understanding Success Criterion2.3.1: Three Flashes or Below Threshold*. Accessed: 2025-01-05. W3C Web Accessibility Initiative (WAI). 2023. URL: https://www.w3.org/TR/UNDERSTANDING-WCAG20/seizure-does-not-violate.html.

[19] Mariam Mdoreuli. *Realtà Virtuale e Cervello: Quali sono gli Effetti della VR sul nostro Cervello?* Italian. Accessed: 2025-01-05. Komete XR. July 29, 2025. URL: https://komete-xr.com/it/blogs/infos/quali-sono-gli-effetti-della-vr-sul-nostro-cervello?srsltid=AfmBOorllW3xBCTI6fEQ_hBigIbkosQ0-oBRYBoBFAv_D45f0sF8ppa9.

[25] Daniela Mazzitelli. *Le onde Theta*. Consultato il 12 gennaio 2026. Thetalab. 2020. URL: https://www.thetalab.it/le-onde-theta?utm_source.

[26] Adriano Gilardone. *Come e quando usare il test di Friedman*. Last modified 6 April 2025, posted 16 December 2024. Adriano Gilardone Blog. 2024. URL: https://adrianogilardone.com/test-di-friedman/.

[27] Adriano Gilardone. *Test di Wilcoxon dei segni: una guida completa*. Last modified 12 March 2025, posted 4 November 2024. Adriano Gilardone Blog. 2024. URL: https://adrianogilardone.com/test-di-wilcoxon-dei-segni/.