POLITECNICO DI TORINO

Master of science program Biomedical Engineering

Master Degree Thesis

Quantitatively Measuring User-experience with Connected and Autonomous Vehicles in Simulated Virtual Reality-based Environments

Supervisor
Prof. Fabrizio Lamberti

Co-Supervisor
Dr.ssa Lia Morra

Candidate
Salvatore La Rosa

April 2019

To my family.
For always.
For ever.

Abstract

The field of autonomous driving is sharply growing. One of the main issues related to such a context is the possible user's lack of trust toward the autonomous vehicle. The area of Human-Machine Interaction (HMI) is expected to provide support in this context. Furthermore, physiological measures can help to obtain a real-time characterisation of the user's physiological state, by letting us understand emotions like the stress level related to a specific situation and or task. Based on the above considerations, the aim of this thesis is to compare two different user interfaces which differ in the amount of information presented to the user: one is referred to "omni-comprehensive", since it presents all the information which are expected to be available in connected and autonomous driving scenarios; the other one is named "selective", as it exploits only a subset of that information. Physiologic measures, i.e. electrodermal activity (EDA) and heart rate variability (HRV), are used as indicators of affective states, and a driving experience in virtual reality is used to elicit changes in user's stress levels. In particular, during the experience, seven unexpected events have been programmed to occur, representing different hazardous situations. A user study has been carried out, including also a pre-post questionnaire survey. The most substantial outcomes of the study show statistically significant differences in EDA between the two interfaces ($p = .036$) when analysing features extracted from EDA during the 10 seconds before and after the unexpected events and in several other control events. Unexpected events have a significant effect on EDA in both user interfaces ($p < .001$). Each user interface has shown up a strong relationship ($p < .05$ in six unexpected events) between the user's subjective response and EDA. Data suggest that, even though the amount of the information provided by the omni-comprehensive interface has been considered as excessive compared to the selective one ($p < .05$), it can be regarded as contributing at reducing the stress response related to a pre-programmed hazardous event.

Contents

Chapter 1

Introduction

The automotive industry is exponentially growing in terms of new technologies, leading digital innovation on board the vehicles and focusing in particular on *connected* and *autonomous* cars. Security and control represent the main common interest of the sector worldwide.

With the term autonomous car, we are referring to a machine which is able to take over tasks and responsibilities from its driver. Five classification level are defined in term of autonomous car; starting from the driver assistance, which does not take the control of the car, and reaching the fully automated system which completely replaces the human driver with the autonomous system.

Instead, with the term connected car, we are referring to vehicles which are connected to the Internet becoming an Internet-of-Things (IoT), enabling the car to send and received data. Such a connection between cars is especially useful with respect to safety concerns. A study suggests that the connected car market will grow at an annual rate of about 10%, in particular in 2016-2021 the connected car market will grow 204% to a value of 122.6 billion euros [1]. Studies concerning safety issues are the most numerous because of the significant number of investments involved.

Companies are increasing attention to invest in such automotive sector, because of advantages that the large-scale adoption of self-driving cars could offer. As above mentioned, the first obvious advantage concerns safety issues; automated vehicles can react in a shorter time than the human driver, potentially leading to a substantial decrease in road accidents. Ideally, thanks to the connection to the Internet, the automated car knows about the other vehicles present in the proximity; therefore it can modify its behaviour considering the current state of the environment like, for instance, a traffic jam. Additionally, the lack of parking issue would be solved, as the car would accompany the passenger at the requested point and find a parking space in another area, and then return to pick him up when required. Therefore, the number of advantages related to autonomous driving is very high.

Nowadays, in the market, it is possible to find semi-automated vehicles.

The main concern is not only related to technological challenges, but it relies on social lack of trust towards the partially or fully automated car.

Human-Machine Interaction (HMI) represents the point of communication between the user and the machine. In autonomous driving context, it has been ascertained that HMI can aid to build relationship based on confidence between the human and the system. Therefore, designing an adequate HMI is a priority for automated vehicles.

Research active in HMI design mainly concerns semi-automated vehicles for which it is necessary to maintain the driver attention towards the road context to react when requested, taking control of the car. Instead, research in the HMI related to fully automated driving is less active. In a high-automation level vehicle, the user is no longer requested to intervene; therefore the utilisation of HMI can create the relationship of trust required to promote acceptance of the new technology.

Physiological measures can also provide support into the identification of emotional states during driving. In fact, many studies focused on identifying the current driver's mental state, for such applications like drowsiness detection or stress evaluation [2][3].

On the other side, driving simulators have become an increasingly attractive tool to evaluate and assess topics related to the driving context. Moreover, since the field of autonomous driving is fastly growing, the driving simulator can provide a controlled environment to test the performance, giving the possibility to save time and money.

Considering the limited literature regarding the HMI in a fully automated car, it is interesting to understand if the user generally prefers to receive driving-related feedback from the driver-less vehicle and to what extent. Thus, the focus of this thesis is *to quantitatively evaluate the user-response when interfaced to a HMI which provide a relatively high driving context information with a HMI that strongly reduces such information.* In other words, to compare a more informative user interface with its counterpart which includes just a subset of information. Such techniques could provide support for HMI designer to develop optimal user interfaces to build trust in a fully automated vehicle.

The thesis is organised as follows: an initial introduction of the scientific literature related to the autonomous field, HMI and driving simulators, following an overview of the physiological signals which can provide support toward the emotional state evaluation. Then, a description of the virtual driving experience is present. Continuing, the detailed working procedure is illustrated with the description of the statistical analysis needed. Finally, result and conclusions are reported at the end.

Chapter 2

State of Art

The following chapter is composed of three main components: autonomous driving, Human-Machine Interaction and driving simulators. The first refers to a general overview of the autonomous driving context, while the second refers to the HMI sectors which should provide support towards the acceptance of such a new technology and, lastly, an overview of the virtual driving simulators which provides a safe and controlled testing environment.

2.1 Autonomous Driving

Nowadays, the field of Autonomous Driving is drawing attention from a very large amount of companies. From the dictionary, the term *a*utonomous refers to the capacity of self-government. An autonomous vehicle, for the National Highway Traffic Safety Administration (NHTSA), is those in which the decision making of the driving operation is not necessarily done directly by the driver. During self-driving mode, a decision like braking, steering or accelerating is taken by the autonomous vehicle [4]. Therefore autonomous cars, which are also named as a driver-less car or self-driving car, are vehicles able to sense inputs coming from the environment and driving without human actions [5]. The transition between non autonomous car and a fully automated car is described in the in the rest of this section.

2.1.1 Classification of Autonomous Vehicle

There are several classification of autonomous driving systems, the most used are the NHTSA and the SAE (Society of Automotive Engineers). In September 2016, the NHTSA adopted the SAE classification described in SAE J3016 [6]. It is a six-levels classification which is divided in:

- **Level 0**
  No automotion: the total dynamic driving task is performed by the driver.

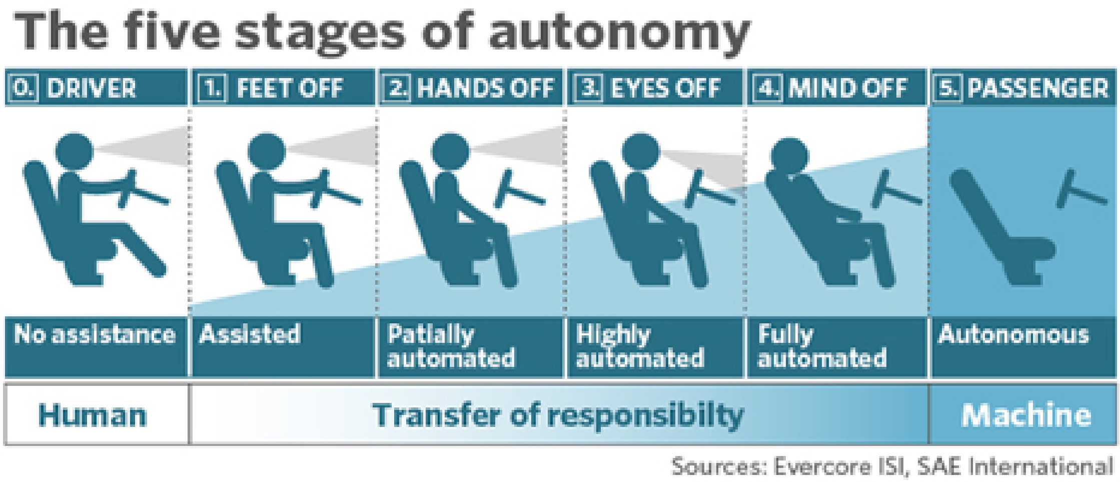


Figure 2.1: Picture showing the level of automotion described by the SAE J3016 standard.

- **Level 1**
  Driver assistance: in some particular situation the system performs the longitudinal or the lateral motion control, and the driver executes the other tasks. Adaptive cruise control represents an example.

- **Level 2**
  Partial driving automation: in some particular situation the system performs both the longitudinal and lateral motion control, and the driver oversees the system and executes left tasks, for instance, lane keeping cruise control.

- **Level 3**
  Conditional driving automation: in some particular situation the systems perform all the driving task activity like steering, accelerating and breaking and other tasks like line changing. However, the driver must be ready to request to intervene or in case of system failure. The driver remains the sole responsible for any accident.

- **Level 4**
  High driving automation: in some particular situation, the system performs the complete dynamic driving task even if the user does not respond to a request to intervene.

- **Level 5**
  Full driving automation: no particular situation is required, the system carries out all the driving aspects, except for the starting of the systems and setting the destination.

More specifically, level 1 autonomous cars are equipped with some modality like LDWS (Lane Departure Warning System) and ACC (Adaptive Cruise

Control). The LDWS is a system that alerts the line changing of the car, for instance in such a situation of driver drowsiness. The ACC system modulate the longitudinal motion control in term of acceleration and deceleration depending on the traffic conditions. In order to give the self-government ability to the autonomous vehicle, sensors are fundamental to sense external input and present them in a comprehensible way to the calculator which represent the brain of the autonomous machine. Both LDWS and ACC systems require ultra-sonar sensors, long-range radar (LRR) and eventually also video-cameras. The ultra-sonar sensors utilise mechanical waves to sense objects located in the proximity of the car. The LRR sensor technology is based on electromagnetic (EM) waves, an antenna transmits these EM waves which are reflected by the objects, from the time between emission and reflection is possible to individuate the object location. Same general idea rules the computing of the distance ultra-sonar sensor technology.

Level 2 autonomous cars, should provide in addition to the technologies of level 1, the LKA (Lane Keep Assist). This functionality acoustically alerts when the car is leaving the lane and, while the ACC is activated, a slight counter-steering to keep the automobile in the trajectory. A further functionality is represented by the PA (Parking Assist) which provides a warning alarm for the driver who is carrying out a parking manoeuvre. It is also available the AEB (Automatic Emergency Braking) system which helps to avoid impact with potential obstacles in such a critical situation, and for inevitable accidents, reducing the gravity of the situation by means of the reduction of the speed and of the predisposition of the car to the impact. Required sensors for these functionalities are similar to level 1 autonomous car. In addition, short radar range sensor is used to enhance better accuracy in short-distance objects.

Following the SAE standard, level 3 autonomous vehicles still need the driver attention, mainly to respond to a request to intervene communicated by the system. However, this level of automation offers major functionality than previous levels. The TJA (Traffic Jam Assist) is a modality which makes the car partially independent in dense traffic condition, the system takes over the lateral and longitudinal control of the vehicle. Therefore, in such a situation, the car is able to drive off, brake, steer and accelerate autonomously. However, the driver has to continuously oversee the system and be prepared to take over the total control of the car at any moment. Another functionality is represented by the DM (Driver Monitoring), capable of evaluating the attention and the fatigue of the driver. Further sensor technology is required for this level of automation, the LIDAR. The light detection and ranging, also known as LIDAR, sensors are based on the emission of optical light, mainly the infrared laser light. To compute the object distance, the approach is similar to the radar sensors, from the temporal distance between the emission and reflection of a pulse is calculated the position. A great advantage of this technology is that works independently of the light condition.

Between the level 3 and level automated vehicles reside the most significant difference in term of autonomy. Level 4 autonomous cars are able to solve the particular situation in which a level 3 autonomous car would have waited for the driver to intervene. Technology sensors used for these vehicles are similar to the previous level. However, the information derived by the sensor are combined and the global information is more informative compared to the one obtained by the single sources, it is the 'sensor fusion' approach [7].

In conclusion, level 5 is the one in which the car handles any problem and in any environmental conditions. Sensors and technologies are seen as a single entity, perfectly combined and able to communicate with each other to drive the car safely and without intervention human along the whole trip [4].

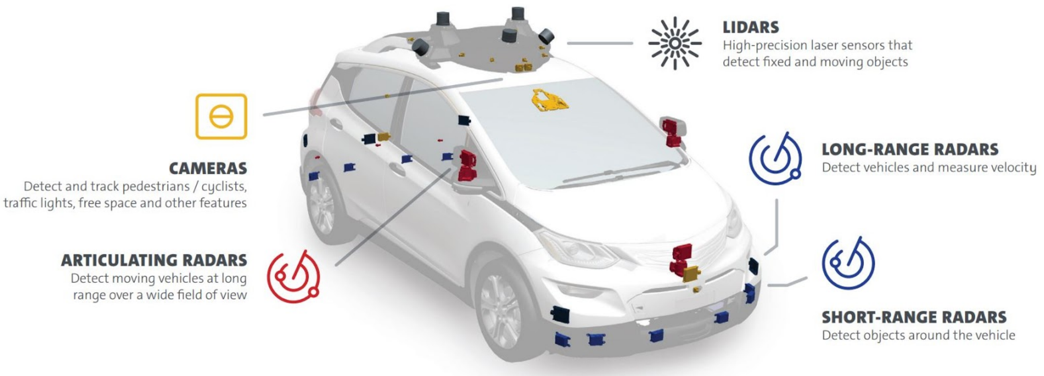


Figure 2.2: General picture representing the typical set up for an autonomous vehicle [8].

2.1.2 Lack of Trust

An unresolved problem of this upcoming technology is the social issue related to automation in daily-life. An American research conducted on 2017 reports that a slight majority of Americans, which answered to the survey (about 4000 individuals), would not ride a self-driving car if given the chance, lack of trust is the main concern, figure 2.3 depicts the different percentage and relative concerns. 56% of Americans declare they would not want to travel in an autonomous vehicle if given the chance, despite the 44% declare they would do so if they had the opportunity. Among those who would not ride a driverless car, more specifically the 42%, the major concern is an unwillingness to cede the decision making to a computer in a possibly life-threatening condition or a lack of trust towards the autonomous vehicle, however, the 30% refers to safety concerns of different types. In addition, the 65% Americans who took part into the survey, also expressed worries about their safety if they were to share the street with driverless freight trucks, more in detail the 32% would feel in danger [9]. Hence, there is a discordant view between Americans, who took part in this survey, regarding the field of autonomous driving. A large number of researches have been done in order to investigate deeply the concept of trust.

The percentage of people who would trust a self-driving car is quite low. The majority of the participants, however, have a certain level of initial confidence that would lead them to try and use a semi-autonomous car.

Jack Weast, a senior engineer at the head of the Autonomous Guide division of Intel, claims that it could even be capable of building a perfect autonomous vehicle from the point of technological view, nevertheless, if customers do not feel psychologically safe, they would not use the services of that car and therefore would not buy it. Hence, it is essential to create a relationship of trust between the user and the machine, which allows the person to perceive a sense of security and well-being both physical and psychological [10].

An experiment conducted by Intel gave the opportunity to take a ride in an experimental autonomous car, in order to get feedback on that experience. The participants were then interviewed both before and after the test session. The main worries of the participants which took part in that Intel experiment were comparable with the main concerns expressed by the American survey conducted in 2017 previously discussed. The machine decision making, the absence of human judgement in the driverless car, the lack of human typical ability to trigger mental processes of analysis and evaluation about unexpected events and situations, such as pedestrians crossing in unauthorized areas, or other drivers that cut off the road to the autonomous car, these represents the main anxieties of the Intel testers. However, they also expressed the belief that the autonomous cars of the future will be safer compared to the human driver. The participant's attention was focused, as well as on the confidence in the car, on the understanding of human-machine interfaces like touch screens, displays, voice signals and more, interaction modalities between the subject and the autonomous vehicle. Results obtained were unanimous: "every single participant experienced a huge increase in the level of trust after completing the trip". Every tester was excited about the growth of this market, showing themselves convinced that autonomous cars will be a safer means of transport. Most participants believed it would take time to become familiar with the autonomous driving system, although once confidence is built, some of the warnings and communications could become bothersome and invasive [11].

2.1.3 Building Trust

The study by Ekman et al.[12] represents one of the most interesting works in on the issue of trust in the context of autonomous cars. The authors have developed a framework that provides general guidelines for HMI designers for autonomous vehicles. The framework collects the ideas proposed by others studies summerized in the paper and define the 11 identified trust-affecting factors:

- **Mental model:**
  An approximate representation of functionalities and skills of the system

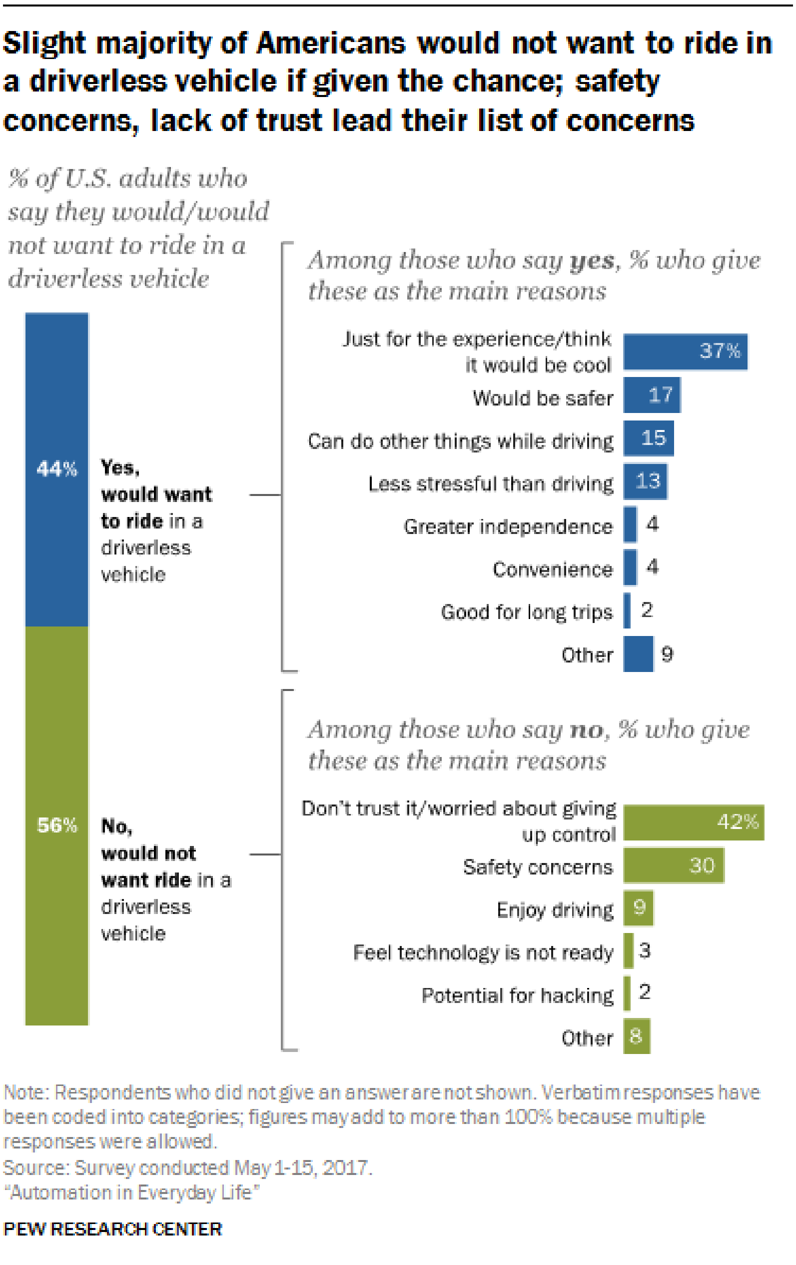


Figure 2.3: Survey conducted in 2017 [9].

that helps the user to understand the intentions and the system decisions;

- **Expert/Reputable**:
  The system is described as a skilled agent, for example through an aesthetically pleasing interface;

- **Common goals**:
  The system proposes goals to the user, for example the possibility of choosing between different driving styles;

- **Training**:
  A learning phase to be performed previously and after the first use of the vehicle, allows increasing the the degree of user knowledge of the system functionality;

- **Anthropomorphism**:
  The attribution of human characteristics to the system, for example through the name, sex, voice;

- **Feedback**:
  The ability of the system to provide continuous feedback to the user, ideally addressing two or more senses;

- **Adaptive automation**:
  The ability of the system to adapt on the basis of the user's physical and psychological state;

- **Customization**:
  The ability to regulate non-critical features of the system based on user preferences. It is based on automation adaptive in the fact that the latter is based on the ability of the system to learn based on user preferences while personalization an active choice.

- **Uncertainty information**:
  Information relating to the system's inability to handle properly certain situations, for example, the inefficient operation of sensors or GPS;

- **Why and How information**:
  Information regarding upcoming actions that the system is about to complete. A message of the type of "how" it describes how the system can solve a certain task while a message of the type "why" provides an explanation of why the system is about to be implemented that action;

- **Error information**
  Information communicated following an error or an accident to explain why is verified and to what extent the performance of the system has been compromised.

This framework takes multiple factors into consideration simultaneously. Trust is indeed influenced by many aspects. The intentions of Ekman et al, developing this framework, were to develop a framework that is simple to understand, as regards the aspect of trust, and able to help designers in the task of calibrating user confidence in the context of human machine interface for autonomous vehicles. The human-machine interface represents, in fact, the point of communication between the user and the autonomous car, therefore it represents a human factor involved in the process of adopting autonomous cars. The study by Ekman et al supports how a good level of trust can be achieved through a well-studied HMI.

2.2 HMI - Autonomous Driving

In order to build trust in automated vehicles, the technology named human machine interface (HMI) covers a fundamental role. The HMI incorporates whether software and device which allows the user to interact with a machine. Such technology can be as easy and omnipresent as a common single-touch display equipped on a machine or as technologically innovative as smart-phones or smartwatches connected to the machine.

2.2.1 The Role of HMI in Different Automation Level

As previously mentioned, the NHTSA taxonomy defines different automation levels, for each of them is required an adequate interface to allow the driver to be conscious of what is going on outside the vehicle, i.e. to establish an accurate situation awareness [13]. For instance, semi-automated level 3 autonomous car comprises two particular phases:

- The autonomous driving phase where the decision making is performed by the machine. Consequently, the driver can carry out some secondary activities. This driver condition is named out-of-the-loop.

- The transition from autonomous to manual driving where the human driver should react to a request to intervene. This is a crucial phase, if not performed correctly, can lead to accidents.

Both phases require the adequate amount of information in order to involve the user on the driving task, holding the driver attention to correctly respond to a intervene request. The human driver must also be conscious of what is happening inside the vehicle in order to fully understand the driving action performed by the autonomous system. To this end, it is essential to define and display the right information at the right time [14].

Therefore, the role of the HMI is critical within the semi-automatic vehicles nowadays available in the state of the art because it represents the point of communication between the automatic system and the driver.

However, the human-machine interface is also important for the future prospect, vehicles with a high level of autonomy, for instance, level 4 and 5, the role of the human user on board changes, the human driver becomes the passenger and the HMI needs a redesign, in order to make the travelling experience comfortable for the vehicle occupants. Nevertheless, the main thought is not to completely replace the standards currently in force which, for example, require the presence of the steering wheel on board the car or a dashboard but to modify them when required [15].

2.2.2 HMI - Display

A fundamental element regarding HMI concerns the choice of how to communicate information to the user. The visual mode represents the primary and widely used in all vehicle interfaces and, consequently, represents the most consistent channel of communication. It is interesting, therefore, to assess which of the various display devices are most effective in the automotive sector in general, and specifically in the autonomous car domain.

There are several types of display devices that can be used inside a vehicle but fundamentally these can be grouped into three categories. Head-down display (HDD), whose main advantage is that not to block the view of the real world for the user, but lead to distracting the driver attention from the road in order to be able to observe the display. Head-up display (HUD) which, on the other hand, allows the use of the necessary information by maintaining the look towards the road context; however, this type of display poses some realizations challenges, difficulties due to the low level of visibility in direct light conditions and it means that the elements of the real world are obstructed by the virtual elements projected on the windshield. The last category is composed by the head-mounted displays (HMD), which offer a virtual vision ad-personam, which is superimposed on the world real; the main advantage is to allow information to be viewed regardless from where the user is directing the look. However, the device it could obstruct or at least cover a certain portion of the world's view[16].

Figure 2.4: Head down display [17].

.

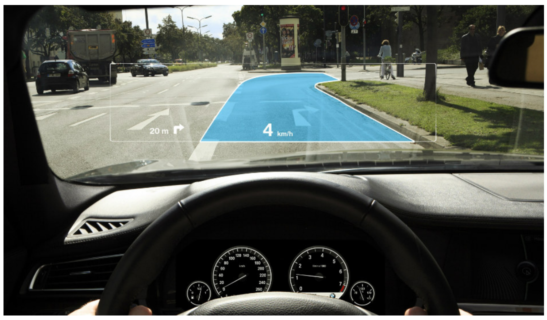


Figure 2.5: Head up display [18].

.

In the literature, there are several comparative studies. An example is provided by the study of Doshi et al. [19] conducted research using the windshield as a display and used augmented reality representations to send alerts to the driver when it exceeded the speed limit. The authors found that the user showed fewer reaction times using the HUD instead of the HDD, with a reaction time shorter than 1 second. Moreover, Medenica et al. [20] worked with a driving simulator to measure user performance with an HDD and HUD that provided two different types of navigation aids. Through eye-tracking measurements, the authors found that motorists were able to keep their eyes on the road for a time significantly higher than that achieved with HDD and experienced a significant decrease in cognitive load.

Therefore, generally, HUD systems are evaluated as more user-comfortable HMI compared to HDD.

Augmented Reality HUD

Through the use of a HUD, the information can be displayed in a spatially dissociated manner to the real environment, or in contact-analog mode, that is in a manner strictly linked to the real environment, Figure 2.6.

The research is very active in the use of AR-HUD (Augmented Reality Head Up Display) semi-autonomous vehicles. AR-HUDs, through the projection of information directly on the windscreen, can contribute to increasing the driving experience and road safety, as well as increasing confidence in the nowadays cars but also in the autonomous ones [21].

Heaeuslshmid et al. [22] have shown how a WSD (Windshield Display), or a HUD that projected on the windscreen information in contact-analog mode, compared to the absence of information, implies a capacity to identify obstacles faster and more effectively.

Concerning the context of the localisation of pedestrians, the study by Phan et al. [23] establishes how a driver assistance system that offers AR aids to the user improves driving safety. A bounding box highlights the pedestrian's position; this visual indication in augmented reality increases the visibility of the pedestrian who is more easily located and makes the driver more vigilant to drive.

In the context of autonomous driving, research into the use of AR-HUDs is less active. Bonte, managing director and vice president at ABI Research, argues that this type of technology can help consumers in the transition process towards autonomous cars [24]. During this process, he believes it is essential to keep the user informed, showing contextual information, that is what is happening around the car; the information should show how the vehicle can manage risks and emergencies. Displaying this information should help to gain confidence in autonomous cars capabilities. However, an overload of information can be dangerous, especially if combined with commercial information or entertainment. Nevertheless, accurately projecting three-dimensional information into the driver's field of view requires a high level of precision and

Figure 2.6: On the left non contact-analog mode, on the right contact-analog mode[25].

sophisticated image processing software. Reaching such a level of accuracy is the limit which slows down the marketing of displays in augmented reality.

2.3 Virtual Driving Simulator

The field of simulation is becoming a mainstream of the nowadays engineering production. It enables to test the prototype before to proceeding towards the production, therefore, limiting the errors and saving money. In the autonomous vehicles field, such context covers a big portion of the whole studies; the driver simulators. The "simulation is a key technology for developing and validating autonomous vehicles on a large scale in parallel to expansive road tests" [26] said Danny Atsmon, CEO of artificial intelligence and deep learning Israelian company, Cognata. For example, the Cognata platform reproduces a simulated test scenario in which there is a traffic model based on artificial intelligence, real-world objects such as streets, buildings, and all the elements of a city scenario. The platform also simulates inputs from car sensors, simulates more weather conditions and borderline cases, such as, for example, a pedestrian running in front of the car. The Cognata platform is used by AID (Autonomous Intelligent Driving), an Audi section which deals with the study of autonomous cars.

NVIDIA company believes that it is necessary to carefully test the autonomous vehicle in a variety of conditions and high scenarios before placing an autonomous car on the road. To this end, the NVIDIA Drive Constellation AV simulator has been created, which allow the generation of self-driving tests in virtual reality. This simulator makes use of photorealistic simulations able to recreate various scenarios and test environments. Using the Drive Sim software, which on a dedicated server, has the role of simulating the sensors present in an autonomous car, such as, for instance, radar, video cameras, LIDAR, it is possible to test the car in rare conditions such as snowfall, thunderstorms and other. A second server contains the Drive AGX Pegasus car computer, which runs the whole autonomous driving software. The system then processes the simulated data, as if they came from the sensors actually present in the real cars [27].

Other companies are working on the driving simulator production, like Toyota, with an important economic investment through the TRI (Toyota Reasearch Institute), is supporting the development of an open source driving simulator named CARLA (Car Learning to Act) developed by the CVC (Computer Vision Center) of the Autonomous University of Barcelona [28].

It is also necessary to mention the Genivi Vehicle Simulator, it is an open source driving simulator produced by GENIVI Alliance, entirely developed in Unity3D. The simulator includes three scenarios: a coastal scenario, inspired by the Pacific Coast Highway, an urban scenario, inspired by the city of San Francisco, and, finally, a scenario inspired by Yosemite. All the scenarios contain the typical elements of the scenario to which they belong, as well as a traffic system. Genivi Vehicle Simulator gives the possibility to choose between two cars, a Land Rover L405 and a Jaguar XJ and, in general, to modify a number of settings regarding the physics of the vehicle, such as, for example, the traction system, the friction index of the tires, the distribution of weights, the gear shifting regime, and so on [29].

2.3.1 Driving Simulator Validation

In order to have a reliable driver simulator, it is necessary to ensure that data collected during the driving simulations reflect the results that would be obtained in the real world, therefore to ensure the *Ecological Validity* of the experiment. Two types of ecological validity are generally evaluated for driving simulators: physical validity and behavioural validity [30].

Physical validity examines the degree to which there is an explicit correspondence of elements, dynamics and layout between a simulator and its real-world analogue. Instead, behavioural validity concerns to the extent to which a driver behaves the same in the real world and the simulator.

Therefore, behavioural validity assesses how the subject behaves during the experiment. Two kinds of behavioural validity are defined: absolute and relative validity. Absolute validity is obtained when numerical correspondence in the measurements are obtained in simulated and on-road drive test. Relative validity is achieved when data between real-life test and simulated test are correlated but not necessarily the same.

Over the years, a lot of research has been done in order to understand how to validate the driving simulator. The first studies regarding simulators validity were conducted by Mudd (1968) and McCormick (1970). The first focus his attention on human behaviour compared between real and simulated in-flight experience [31], the latter concerns the physical correspondence between the two different experiences [32]. Therefore, the concepts of behavioural and physical validity have been considered by the authors.

In 1982, a step forward was done, a distinction between driver behaviour and driver performance was made. Driver behaviour refers to how a subject chooses to drive, instead driver performance refers to the driver motor skills to drive [30]. For the following years, many authors which aimed to validate the driving simulator, based their validation approach on the concept to compare the driving performances in term of speed, lateral position and many others between simulated and on-road experience.

2.3.2 Ecological Validation - Driving Performance

As already mentioned, in literature the main approach used to validate a driving simulator is to measure the driving performance and compare these measurements between the on-drive test and the simulated test. Usually, driving performance measurement collected are speed, accelerator pedal, clutch pedal position, steering-wheel angle, lateral acceleration, longitudinal acceleration, yaw angle and brake pedal position. Thanks to these measures it is possible to compare them with the same measurements collected in simulated experience.

An example of these driving performance measurements was given by the RENAULT group, which studied the different behaviour between the simulator and a test track using an instrumented vehicle. They distributed 166 subjects into three groups for three different cars of the same model, individually conformed to oversteer or understeer on an emergency manoeuvre, (at different mean velocity: 60, 70 and 80 km/h) which performed the driving task in the test track for three laps. The track was composed by separators to create a specific path. For the last lap, the separators were changed in order to create a sudden narrow curve. For the simulator instead, 66 subjects were divided into three groups, one group for each car type. A questionnaire was used to asses the reality of the simulator. Type of driving performance measurement was previously reported, in addition, they also measured skin temperature, heart frequency, skin potential, blood rate breathing frequency for all subjects. The focus of their work was to assess if it is suitable to compare results obtained from a fixed-base simulator with results collected from the real road experience. Result collected showed no lane leaving in the first two laps, and the relative validity has been achieved between the two systems. In particular, their results show that understeer car is less suitable for sudden manoeuvre, and the simulator is poor in accuracy on narrow curves[33].

As predictable, one of the most common measurements of behavioural validity is the driver speed. Validation studies like [34], [30], [35] have established the relative validity, and in few cases also the absolute validity, in driver speed performance. While on a straight path the driver speed in the simulator is higher than the simulated drive test [36], on a curve the driver speed shows behavioural differences between the two systems. For narrow curves (radius of curve 146 m) the simulator driver speed is lower, while for large curves (radius of 582 m) the driver speed is higher in real drive conditions [37].

In a paper published by Bella et al. 2008 [38], the aim was to record the speed at eleven measurement places with different arrangements on a two-lane country road near Rome. The on-road experiment was recorded and recreated in a virtual reality, so the experience between the two systems is more similar. Results obtained shows that the relative validity has been achieved, also the absolute relativity has been established for 9 measurement sites, Figure 2.7.

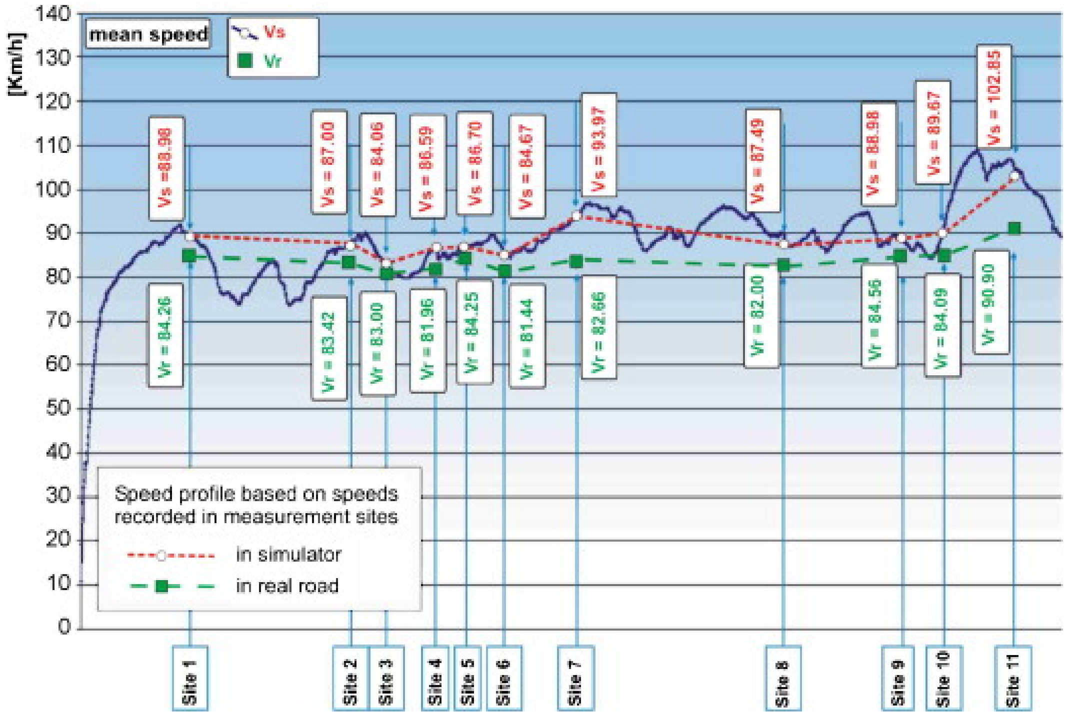


Figure 2.7: Image represents the mean speed from the field (Vr) and simulator (Vs) in the measurement positions [38].

.

The same approach was used by the authors of [39], where the focus of their work was to validate the driving simulator for speed. 16 subjects conducted the on-road drive test in a desert route, and the same road geometry was replicated in the simulator scenario, the environment was approximately reproduced. Half of the drivers performed the on-road test first and a half drove in the simulator first. Also for this study the simulator quite accurately reproduces speed sensations, therefore validity has been obtained.

Many other studies have analysed the behavioural validity of the driver simulator following similar approach previously described. Table 2.1 summaries papers published which aims to validate the driver simulator, it also represents some different validation approach.

2.3.3 Other Validation Approach

However, to compare driver performance measures between real and simulated driving experience is time-consuming and requires many efforts. Indeed, it is

Table 2.1: Table which summirizes some driving simulator validation studies [40].

Validation Studies	Driving Simulator	Data Collection Technique	n/Age	Statistical Analysis	Comments/Key Findings	
					Absolute Validity Established	**Relative Validity Established**
Aksan et al,2016 [41]	DriveSafety simulator DS600	Sim. and on-road	n younger=42 - 32 (M age= 35) n older = 37-30 (M age=77)	ANOVA, Pearson correlation		Ages differences in driver performances
Bella,2005 [42]	CRISS	Sim. and on-road	N=35 Age=24-45 yrs	Comparison of means Bilateral Z-test		Speed
Bella,2008 [43]	CRISS	Sim. and on-road	N=40 Age=23-60 yrs	Comparison of means Bilateral Z-test	Speed for non-demanding road configurations	Speed and complexity of maneuver
Bittner et al,2002 [37]	Uni. Whashington Real Drive	Sim. and on-road	N=12 Age=21-34 yrs	Generalized Youden plots		Curve entry speed
Charlton et al, 2008	MUARC	Sim. and on-road	N=100 M age=36 yrs	ANOVA correlation		N=Speed and gaze direction
Godley et al, 2002 [44]	MUARC	Sim. and on-road Speed countermeasures	N=24 M age=22-52 yrs	ANOVA; correlations and canonical correlations		Speed countermeasures
Helman et al, 2015 [45]	TRL's DigiCar driving sim., SCANeR	Sim. and on-road DBQ validation	N=32-31 M age=32 yrs	Cronbach's alpha, ANOVA		DBQ for speed
Hoffman et al, 2002 [46]	IDS	Sim. and on-road	N=16 M age=25-55 yrs	Descriptive statistics: Comparison of means		Mean braking onset
J.Johnson et al, 2011 [47]	STISIM	Sim. and on-road Physiological responses	Study1: n=24 Age1=21-57 yrs Study2: n2=9 Age2=20-47	Paired t-test Means comparison two-way ANOVA	Mean Heart rate and metabolic consumption in some physiological responses	Heart rate responses metabolic consumption
Lee et al, 2007 [48]	STISIM	Sim. and on-road	n=50 PD patients n=150 controls 60-80 yrs	Pearson correlation, Stemwise linear regression		Control participants' overall driving performance
Slick et al, 2006	DS-600c motion-based simulator	Sim. and on-road; Physiological responses	n=22 M age= 16 yrs	ANOVA	Physiological responses during turns at intersections	Physiological responses during turns during entire drive
Wang et al, 2010 [49]	STISIM	Sim. and on-road Eye-tracking	n=28-30 Age=22-28 yrs	Pearson correlation Means comparison	Total glance duration % eyes forward SD forward speed	Visual attention
Watts & Quimby, 1979 [50]	Transport and Road Research Laboratory	Sim. and on-road; Physiological responses perceived hazard risk ratings	N=60 Age=unknown	Correlation coefficient	Perceived risk hazard on the road and in sim.	

necessary for the utilization of an instrumented vehicle, which is a car equipped with video cameras sensors and more.

Some studies try to approach the validation problem differently. An example is given by the paper published by Reimer et al, 2006 [51]. In the article, the authors underline the problem that no real 'gold standard' method exists for driving validation, neither a unique definition of validity is given(face validity, external validity, internal validity concurrent validity, predictive validity, behavioural validity, physical validity etc.). The approach used is to assess the validity of driving simulator data through a self-report. 48 active drivers were selected, of which 25 subjects were affected by attention deficit hyperactivity disorder (ADHD).Participants filled out four written questionnaires pre and post simulation testing: a U.S. variation on the U.K. Driver Behavior Questionnaire [52], a survey which asked information about participant driving history, pre-simulator and post-simulator sickness survey and a health information questionnaire. A training phase of the virtual driving environment was included (about 10 min). Later, two testing segments were given: firstly more extended high-stimulus urban testing segment (Segment 1), and secondly a more extended low-stimulus rural testing segment (Segment 2). Incentives for participation were given to the participants, encouraging them to complete the task rapidly, keep their attention and guide safety. A full 2001 Volkswagen Beetle cab and the STISIM software constituted the driving simulator. Significant relationships were found across six measurements: velocity,passing, behavior at stop signs, weaving between traffic, speeding and accidents. They have evaluated concurrent and discriminant validity. Results obtained suggest that the considered measures were valid indicators of the behaviours of interest. Therefore a self-reported data regarding subject driving history, health status, pre-post simulator condition were recognised as a reasonable parameter to validate the driver simulator.

Further, in other studies, driving simulator validation is carried out through the concept of 'presence'. The general concept of 'presence', in simulator context, is the sense of being in a place which is not real, "*being there*" [53], inside the virtual reality. A reasonable assumption has been proposed; subjects which experience strong level of presence in a virtual environment will behave as they would have done in real life. The paper [54] gives an example, the 'presence' is measured assessing the degree of 'attention' in the primary driving task. The level of 'attention' was modified by adding oncoming traffic and a second task; they analysed subjective 'presence' and behavioral measures of driving performance. The second task provided the addition of double clicking of laptop mouse positioned so that the driver had to deflect their gaze from their main view. They analysed the driving performances; mean, standard deviation of speed and lateral position. Furthermore, they administered a reviewed version of the MEC Spatial Presence Questionnaire (MEC–SPQ) after each session [55]. Results obtained suggest that subjective measurements, based on a questionnaire survey, no significant behavioural effects were noticed. It has been ascertained that driving is highly related to automation in the cognitive process [56], therefore use of both subjective (e.g. questionnaire) and objective (such as physiological) measures is the best way to evaluate 'presence' in driving simulators, following the idea that greater is the 'presence', and closer behaviour to the reality can be obtained [54].

2.3.4 Physiological Measures to Assess Driving Simulators

The main issue related to the validation of a driving simulator is the necessity to compare the simulated with the real outcomes. An instrumented vehicle is expensive and time-consuming, furthermore the conducted on-road experiments will never be exactly the same for each subject (e.g. weather conditions, traffic conditions and others) and reproduce the real-test in a simulated environment is time-consuming.

A proposed solution to limit the complexity and necessity of records the on-road experiments are using physiologic signals. As represented in table 2.1, a few numbers of studies aim to validate the both traditional and virtual driving simulators using objective measurements which are not (at least not the only ones) the driving performances. In 1979, Watts and Quimby validated their traditional driveing simulator using the skin conductance, considering that valid simulator should induce the same physiological responses as in the real. Their participants assessed risk during on-road driving and while they viewed a film representing the same condition of the on-road experience. Skin conductance data were collected during the film. Result obtained showed that risk ratings were similar between the two systems and a high value of hazard was highly correlated with changing in skin conductance values [40].

Bédard in 2008, during of the STISIM users group meeting, presented a paper: "*Driving Simulator: are they useful tool?*" which examined physiological responses from the subjects which were involved in the simulator. This paper was reviewed and published [47]. Two studies have been conducted, number and age of participants are written in Table 2.1. In the first study (Study 1), participants conducted a non-immersive driving simulation while the HR signal was continuously measured via 3 ECG lead. In the driving simulation, three unexpected events were pre-programmed to happen. Such situations provide a car that suddenly pulled out in front of the driver, requiring a fast manoeuvre to avoid the crash, while into the additional two situations a fast-braking action was required to respond a green lights changing to amber. Instead, Study 2, 9 participants were involved in on-road driving, and also the same non.immersive driving simulation was done (same scenario), but unexpected events were removed. The on-road session is the real version of the simulated route. Data measurement collected were: HR, oxygen consumption, and ventilation. Figure 2.8 shows the result of the Study 1, Figure 2.9 depicts the results of the second study. The first result shows that in Study 1 significant

HR elevations to surprising simulated events can be observed, therefore the sensation of "being in the simulation" is felt by the subjects (i.e. 'presence'). Study 2 result shows that cardiopulmonary responses increase from the baseline for simulated an on-road task, an especially strong correlation for MV_e has been observed. Relative validity for mean heart rate and mean oxygen consumption was achieved comparing simulated and on road experiments, while absolute validity was achieved for mean ventilation rate.

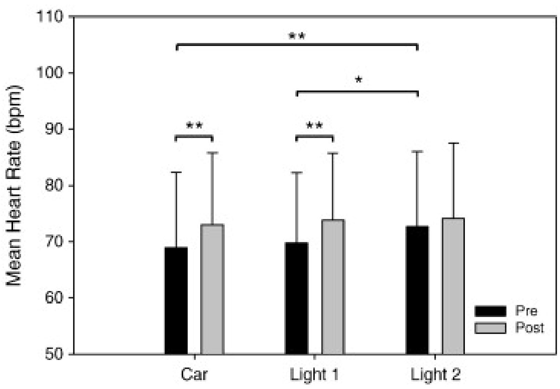


Figure 2.8: Mean heart rate during the 15s pre and post unexpected simulated events. BPM: beats per minute. *Significant difference, $p < 0.05$. **Significant difference, $p < 0.01$. Values are means ± SD. Results obtained in paper[47].

.

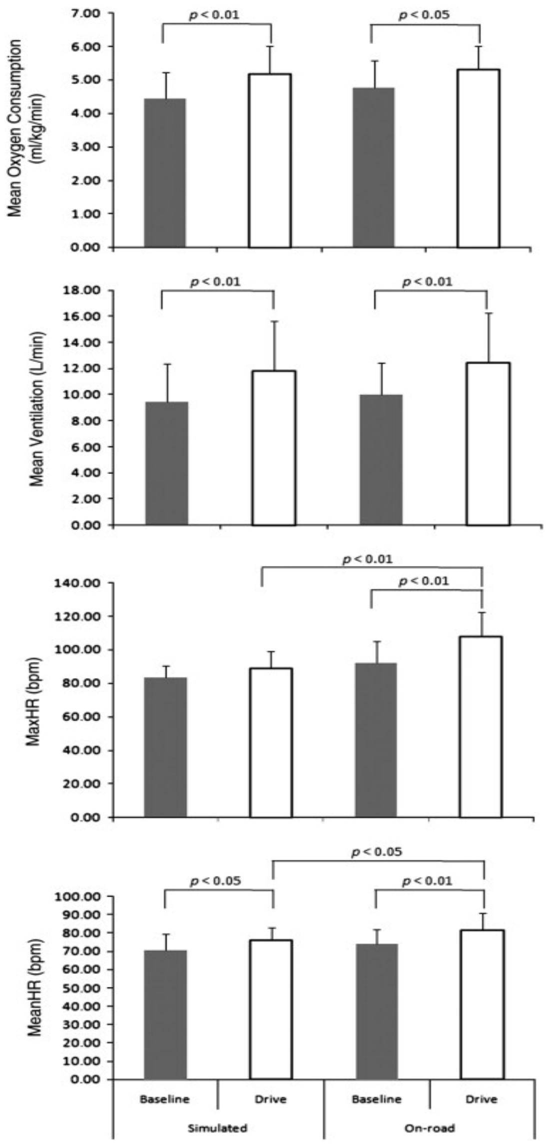


Figure 2.9: The bar images represent the mean and standard deviation value of the above mentioned physiological responses to on-road driving and simulated conditions. BPM: beats per minute. Results obtained in paper [47].

.

A paper published by Eudave et al. 2017 [57] compares the immersive virtual environment by using a modern head-mounted display (HMD) driving simulator, named IDS with the standard driving simulation, named SDS, using tv monitor. The comparison is carried out evaluating the sense of 'presence' through physiologic signals: heart rate (HR), electrodermal activity (EDA) from the right palm, electromyographic (EMG) signal from the tibialis anterior muscle. 5 young male subjects were recruited for the experiment. Later the respective training phase, two driving sessions SDS and IDS were performed. During the test sessions, 20 unexpected events are programmed to occur. Results report an increased level of mean heart rate during the session between SDS and IDS, depicted in figure 2.10. Authors have also analysed the mean value for the subjects of the HR and EDA signal between the 5 s prior to an unexpected event to 10 s later, named event-related signals (ER). The ER-HR activity shows an increase of the cardiac activity when emergency manoeuvre was performed for the IDS, instead, the ER-EDA shows the same increasing trend for the SDS and IDS for the kind of situations, figure 2.10 c) and d). However, for this study only a few numbers of participants have been enrolled, for more consistent results an higher number of participants would be required.

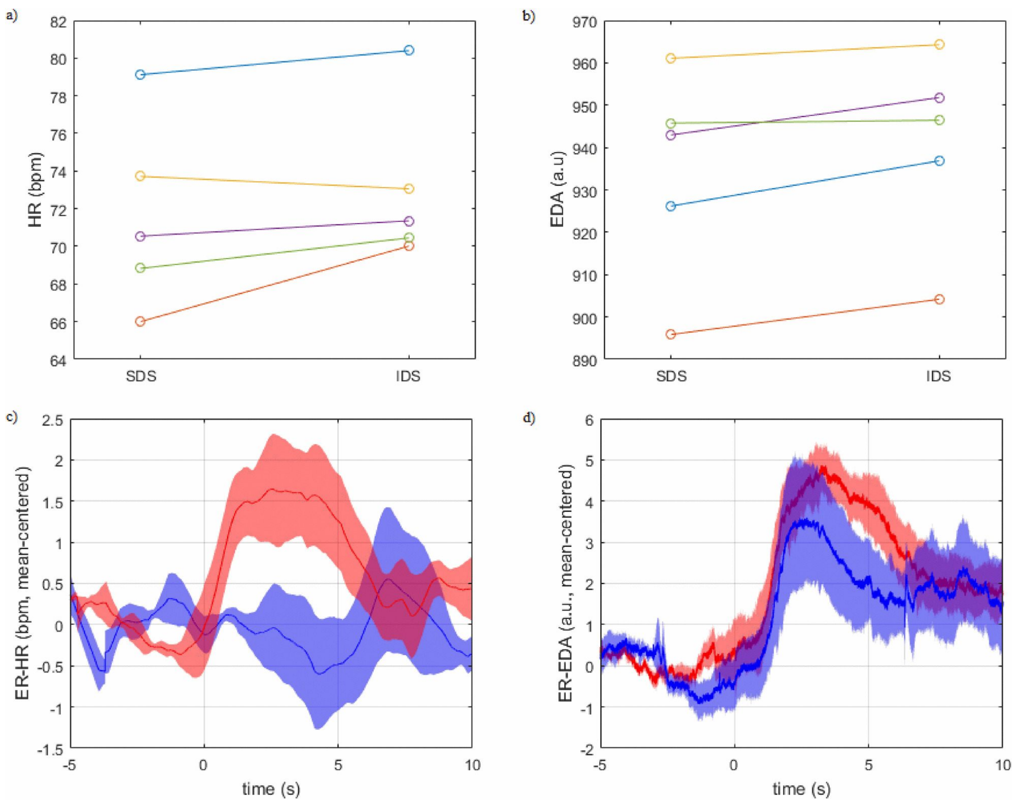


Figure 2.10: Mean session HR(a) and EDA (b) for each of the 5 subjects. Event related (ER) changes measured as mean and SEM (upper and lower limits) in HR (c) and EDA (d) during emergency maneuvers. Blue = standard display simulation (SDS); red = immersive display simulation (IDS)[57].

.

Other studies like [58] have used the surface facial temperature and EDA signal to asses the mental workload during a driving simulation experience. Their findings report a nose temperature changes, decrease of the skin potential and increase of skin conductance during driving stress conditions (for their work was to drive at 180 km/h).

Further, other authors have analysed biosignal during driving simulation experience. Zuzewicz et al, 2013 [59] have studied the HRV and EMG signals under the condition of simulation of crash threat. The experiment is carried out by 22 participants which were passengers of the vehicle during the simulation. Signals recorded were the ECG and the EMG of flexor digitorum superficialis (FDS) and m. trapezius (TR), which were recorded on the right and left sides. From the ECG signal, the HRV was computed, from this last signal many features were extracted in time and frequency domain. In frequency space the power spectrum of the HRV has been computed, the power of the low frequency (LF, frequency range of 0.04–0.15 Hz) reflects mainly the sympathetic activity, while high-frequency power (HF, a frequency range of 0.15–0.40 Hz) reflects the parasympathetic activity. The test ended with a crash, the EMG and ECG have been continuously recorded also later the crash. Results reported in figure 2.11 shows an increase in HR during the crash and a strong increase of the myoelectrical activity of the FDS muscle, figure 2.12. In figure 2.13 is presented the HRV results in the frequency domain. It shows a slight decrease in the sympathetic activity (LF power) and a consistent decrease in the parasympathetic activity (HF power) during a crash.

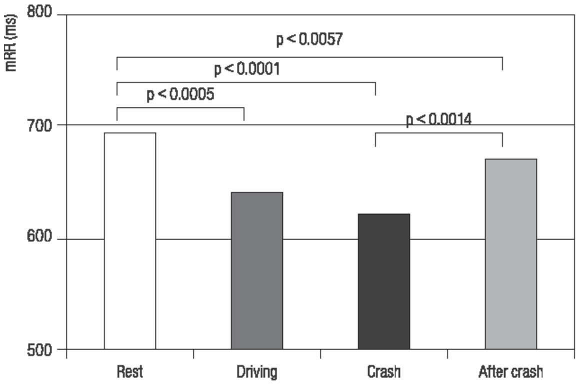


Figure 2.11: Histogram representing the arithmetic mean of all RR intervals of the heart [59].

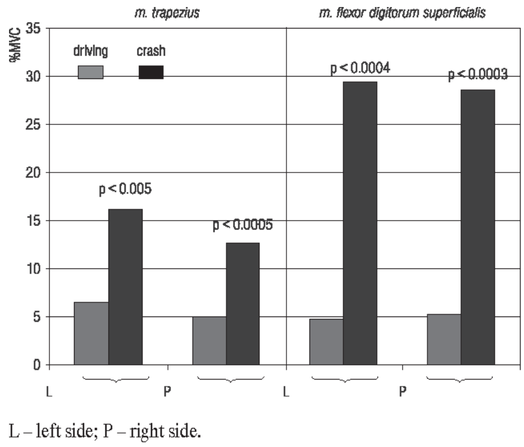


Figure 2.12: Values of TR and FDS muscle contraction median expressed by RMS (root mean square) and %MVC (maximal voluntary contraction [59].

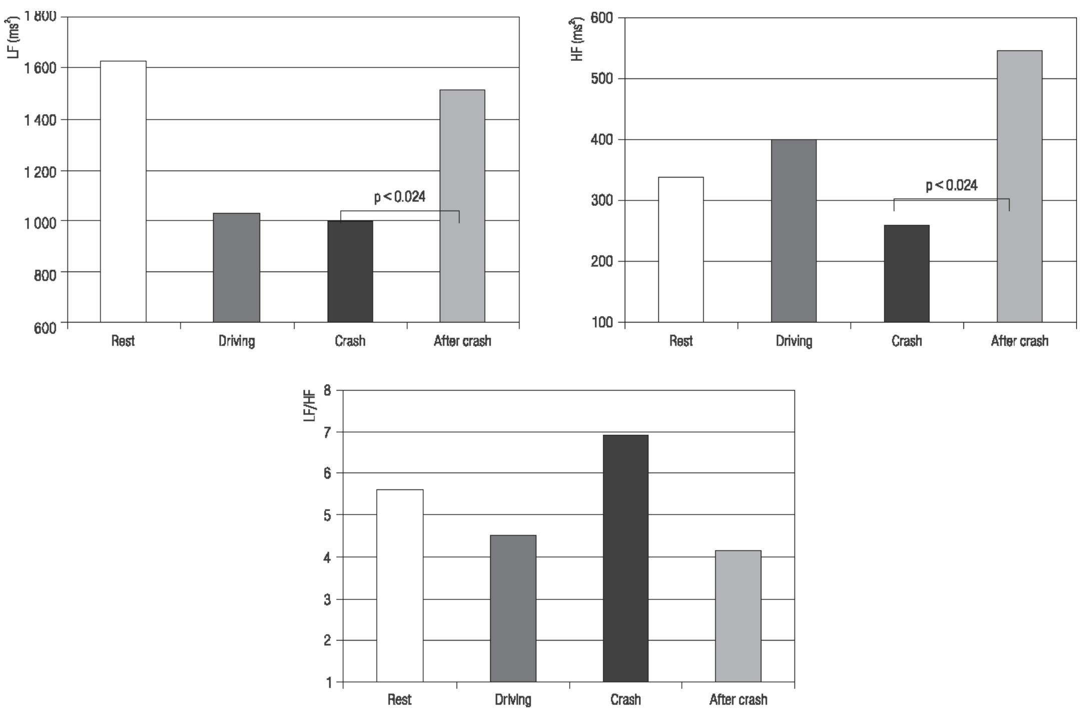


Figure 2.13: LF (low-frequency)power of RR intervals, HF (high-frequency) power of RR intervals, LF/HF is the ratio of low-frequency to high-frequency power. [59].

A paper recently published [60] used physiologic measurement to increase the performances of real-time crash prediction already developed by the same group [61]. The focus of their work was to design and validate a prediction method based on machine learning. Data from 130 driver simulator have been collected. Discriminant Analysis (DA) have been used to combine the new features from behavioural and physiological assessments with the original features, similarly extracted in the previous study, from vehicle dynamics and distance metrics. The experimental session is presented in Figure 2.14. Physiological measures collected were Skin Conductance (SC) and Heart Rate (HR). Results show that an improvement of classification performance has been obtained, adding the physiologic measures lead to an 88.09% sensitivity and 90.15% specificity of the studied crash prediction model.

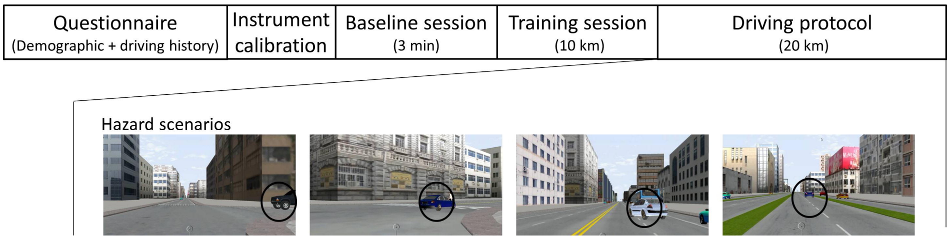


Figure 2.14: Image representing the experimental procedure, the approaching danger is indicated with the circle [60].

In Milan, a study has been done [62] to validate the driving simulator. In this study, driving performance such as mean speed and duration of the driving task was analysed. Electrodermal activity signal was also collected. Real drive experience was replicated in the simulator. Result obtained shows a strong correlation between the simulator and real drive comparison, instead, relative validity for the skin conductance signal was not properly achieved (non-parametric Mann–Whitney analysis showed significant differences in the variance of Electrodermal Activity recorded during on-road test and EDA recorded in simulated test, Figure 2.15. The main limit of this study is that only one participants was analysed.

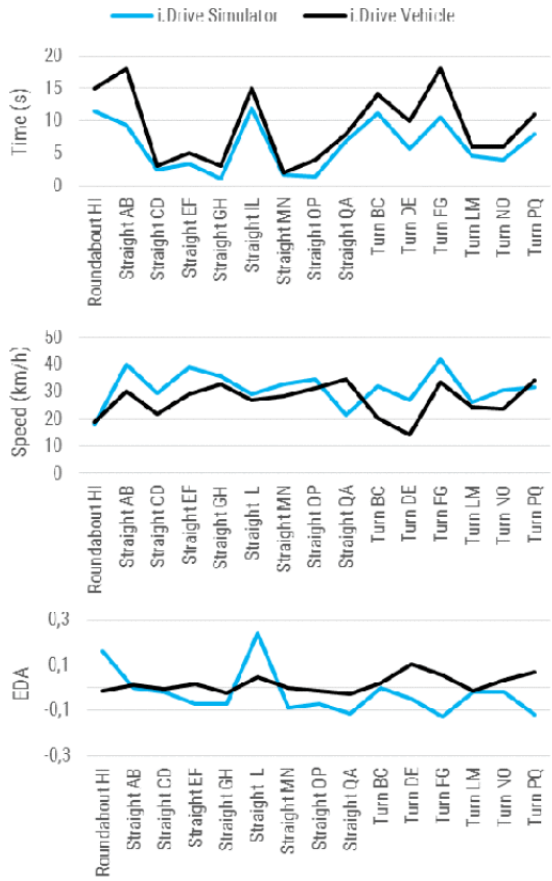


Figure 2.15: The image shows the collected variables estimated over the experimental environment. Significant differences were found for the electrodermal activity (EDA). No significant differences in the distribution of time and speed between real and virtual driving. [62].

Another study which takes into observation the skin conductance response related in an immersive virtual reality experience is the paper published by Slater et al. 2006 [63]. The authors use the concept of 'break in presence'(BIP). Such BIP is a stimulus given to the tester under the simulation experience. The core idea is to evaluate the event-related galvanic skin response (GSR) and heart rate signals. They analysed the GSR response 10 seconds pre and post each BIP, four BIPs events were programmed to occur. The min-max normalized GSR was averaged for the 4 BIPs and for all the subjects. The same procedure was performed 100 times, for each of them averaging 4 random generated times. The result is presented in figure 2.16, in which the goodness of the information provided by the GSR signals is highlighted.

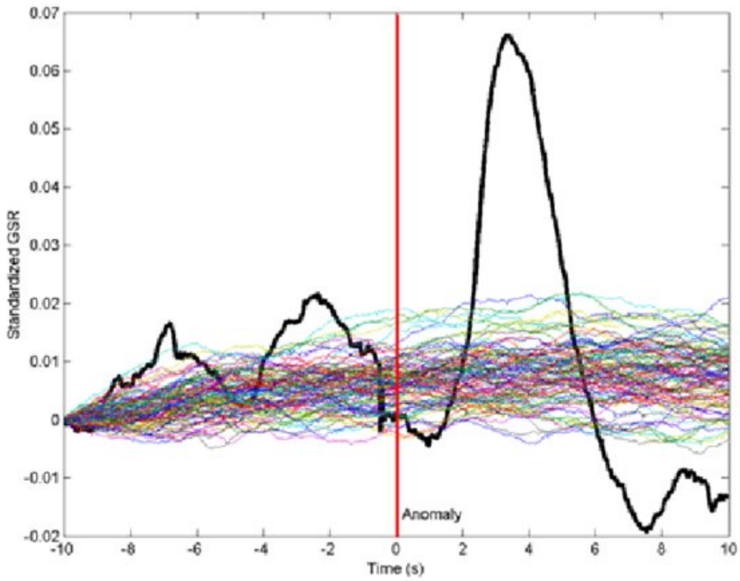


Figure 2.16: GSR averaged plot. the black line represents the mean shape of the true anomalies, the others are randomized mean shape generated from random time sequences [63].

To sum up, HR is established to be an excellent physiological marker. Guger et al. 2004 [64] proved that the heart rate variability could be used as a parameter that reflects the physiological state of the participant. Additionally, the paper published by [65] provides a long table which reports several studies based on HR signal related to the driving in real or simulated experience, figure 2.17 summarise the 17 papers reviewed, details are reported in the article.

Furthermore, other works have shown that skin conductance signal is an interesting physiological variable which can assess the 'presence' in the virtual environment [66],[40].

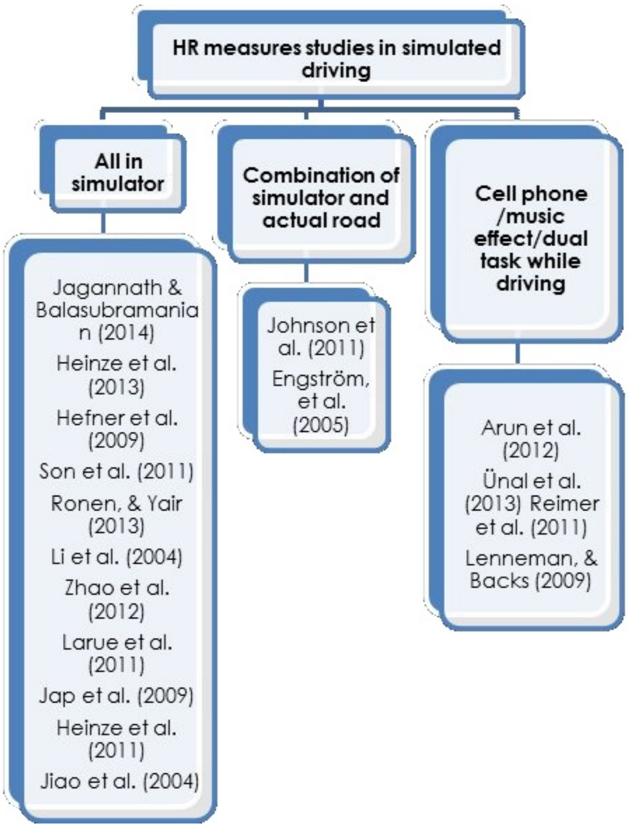


Figure 2.17: Categorization of studies analysed from Faizul Rizal Ismail et al. 2016 [65].

2.4 Considerations

The state of the art of self-driving the survey cars evidenced that regardless of the particular degree of vehicle automation, to promote the full approval of this technology it is a priority evaluate human factors, in particular, the level of trust on the autonomous vehicle. Examining the problem of trust is crucial because a lack of trust between the consumer and the car would prevent the use of the latter and therefore its adoption and consequent diffusion.

Based on studies on trust between human and Autonomous cars, it is evident that to permit the consumer to be able to trust the new technology, feedbacks given by the system are required. This kind of information can be displayed thanks to human-machine-interface, which is a communication system between the consumer and the machine. By displaying such information to the person on board, it increases its situational awareness, and he can build a mental model based needed to understand the actions and intentions of the car. Therefore, once understood that the autonomous vehicle must inform the user about what is going on, the HMI serves as the point of communication between the human and the autonomous system. Hence, it is helpful to evaluate how to design the user interface.

Concerning the design HMI for autonomous vehicles, much of the literature has focused the attention on user interfaces for semi-automatic cars, which are those nowadays available on the market. An optimal user interface must consider priority the issues of cognitive load, awareness of the situation, trust and usability.

For a higher level of automation vehicles, the research on the design of user interface is less active but very appetible for the future market. For a high level of automated car, it is conceived with big spaces for the passengers inside the vehicle to engage other activities. However, there is the main thought to maintain visualisation of the driving information situation via a display of HUD which project information augmented (AR-HUD). AR-HUD offers several advantages to the driver; the principals are to reduce the reaction times and keeping an eye on the road context. Providing road context information should allow the user to acquire confidence and build trust. However, such information is necessary for level 2-3 automated cars in order to maintain the driver attention, but it is interesting, therefore, to ask oneself if in the project of an interface for autonomous cars level 4 or 5 the inclusion of road context information is something useful or not for the passenger's travel experience, to what extent and of what type.

Such a question could be evaluated by comparing a more informative interface with an interface which is deficient in road context information, providing a selection of the displayed information.

Therefore, scientific literature presents many studies regarding the HMI in semi-automated vehicles, while the research is less active toward HMI for level 4 and 5 autonomous vehicles. On the other hand, physiologic signal related to the assessment of virtual driving simulator has been established as informative indicators; Johnson at al. 2011 [47] utilised mean heart rate and mean ventilation outcomes to validate the driver simulator, while Eudave et al. 2017 [57] have used the electrodermal activity to evaluate differences between a traditional driving simulator with an immersive driver simulator.

In conclusion, physiological measurements can be used as a real-time marker of the affective states related to a condition, or in our case, to evaluate the HMI effectiveness.

Chapter 3

Physiological Signals

The relationship between the body and the mind and their influences from each other are explored by the psychophysiology. Physiological signals reflect spontaneous responses of the autonomic nervous system (ANS). A lot of physiologists have adopted these measurements as objective quantifiers of human emotions such as depression, sadness, anger [67], while other researchers in human studies have utilised them to ascertain stress and mental workload [68].

Thanks to the literature studies described in chapter 2, the heart rate and the galvanic skin response are the physiological signals which have been selected for this thesis project due to their intrinsic information and correlation with the ANS. In addition, these are physiological measures are easily recorded in a non-invasive way.

Therefore, the following pages briefly describe the nature of the *heart rate* and the *galvanic skin response* signals.

3.1 Heart Rate

The main energy and power source of our body is the heart. The heart acts as a pump, it gives the required amount of energy to the blood in order to beat the resistance provided by the arteries or veins and gravity force to have a continuous flow throughout the body. From the anatomy perspective, the human heart is composed by of four chambers, right and left atrium and right and left ventriculus. Each of the upper chambers is the atrium and they represent the collecting chamber. In addition, their role is also to contract in order to push blood toward the lower chambers, the left ventricle and the right ventricle. Indeed, the main pumping chambers of the heart are the ventricles which drive the blood to the rest of the body or to the lungs.

In particular, there are two different and connected paths in the human circulation system named the systemic and the pulmonary circuits. Even if both circuits carries the blood and its micro-component, they differ in the oxygen concentration. On one side the systemic circuit transports oxygenated blood to the tissues of the body releasing oxygen to the system and accumulated carbon dioxide. This deoxygenated blood is conducted to the pulmonary circulation. On the other side, the pulmonary circuit transports deoxygenated blood to the lungs in order to oxygenate the blood and releasing the carbon dioxide [69].

Inside the heart, the blood flows from the right atrium to the right ventricle, once the pressure inside the Rx atrium is greater than the pressure inside the Rx ventricle the tricuspid valve, which separates these two compartments, opens and the blood goes inside the ventricle. In this last chamber, the blood is pumped into the pulmonary circuit. In the pulmonary capillaries, the gas exchange occurs and blood oxygenated blood goes towards the left atrium. From the left atrium, when the pressure inside the Lx atrium is greater than the pressure in the left ventricle the mitral valve opens and the blood enters the left ventricle, which pumps it into the systemic circuit once the aortic semilunar valve opens, figure 3.1.

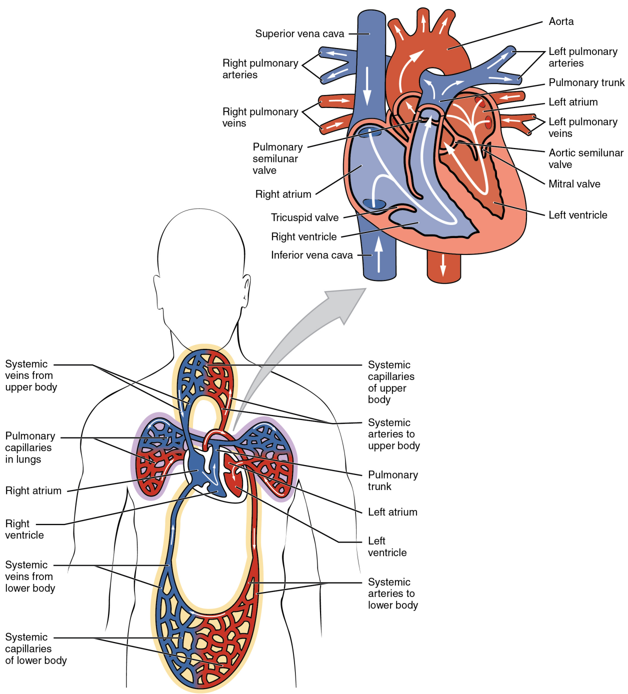


Figure 3.1: Picture representing the dual system of the human blood circulation on the bottom left and the heart components in the top right [69].

Cardiac muscles cells lead the heart to contract and consequently to pump the blood. In rest condition, the heart muscle cells are polarized, it indicates that slightly unbalanced concentration of ions across their cells is present. A major concentration of sodium ions on the outside of the cell membrane induces a positive charge related to the inside cell charge, this potential difference is about 90 mV and is named resting potential. The cell membrane is approximately impermeable to the entrance of the sodium. Though, this permeability is modified by the stimulation of the muscle cells. Then, some sodium ions migrate into the cell causing depolarization, a change of the sign of the charge inside the cell. Consequently, the permeability of ions is again modified and the repolarization occurs. The aforementioned electrical activity is named action potential which leads to muscle contraction. The registration operation of the depolarization and polarization potential performed on the skin surface is named electrocardiography (ECG). The ECG signal, or also called electrocardiogram, is firstly characterized by a flat line indicating no electrical cardiac activity ongoing, also knows as isoelectric line. The first changing on the potential is represented by the P-wave which is represented by the depolarization and contraction of the Rx atrium. Once back at the isoelectric line, a rapid sequence of waves are present: Q-wave, which is a descending oscillation, followed by the R-wave which is a sharp upswing pulse and lastly the S-wave which is a strong descending pulse returning to the isoelectric line. This succession of waves is also known as the QRS-complex and represent the ventricular depolarization linked with the ventricular contraction. Next, the ions concentration back to their original state and this ions movement cause the T-wave which corresponds to the repolarization of ventricles [70].

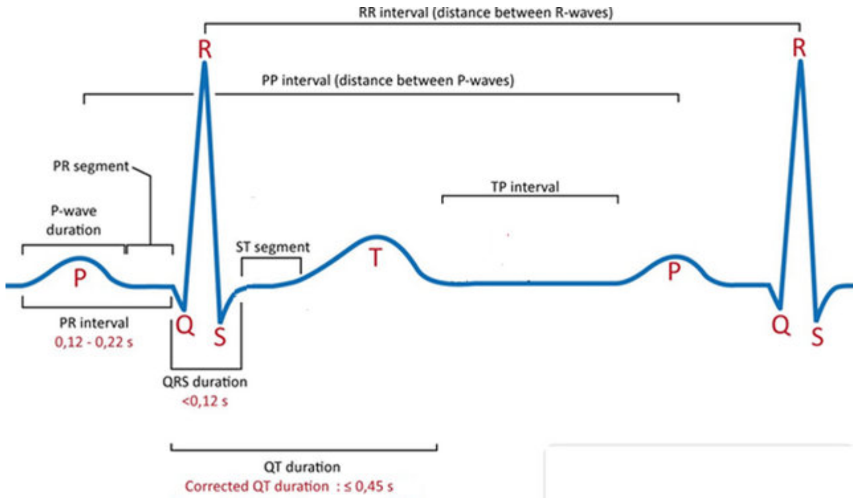


Figure 3.2: Picture representing typical shape of the electrocardiogram [71].

The shape of the QRS complex and the other waves in the represented figure 3.2 is an idealization. Indeed, the shape turns depending on which registration system procedure and electrodes are being utilised. Figure 3.2 reviews the terminology used to define the typical components of the ECG signal.

An interesting and very informative feature related to the ECG is the Heart Rate and represents the frequency of cardiac contractions, measured in beats per minute (bpm). Typical values for resting condition are 60-100 bpm. It has been studied that HR is influenced by stress, physical exercise, ingestion of drugs, illness, anxiety and sleep [72][73][74].

In addition, related to this biologic signal, is the heart rate variability

(HRV), an example is depicted in figure 3.5. This indicator extent the time between two consecutive R-waves called RR-interval. The HRV is seen to be a very informative source of ANS state. Many studies support that ANS activity can be measured from such a signal [75]. Parasympathetic and sympathetic nerves transmit efferent signals to the heart and afferent signals to the brain in order to actuate the reflex functions. Sympathetic nerves increase the heart rate, instead parasympathetic nerves slow heart rate. There is a multitude of factors that can affect the HRV signal [76]:

- Drugs assumption
- Cardiovascular conditions
- Respiration system
- Age
- Renin-angiotensin System
- Physical exercise
- Mental or physical stress/state

Further information which can be extracted by the HRV signal are the features in the frequency domain. The spectral analysis of the HRV can be performed with parametric or non-parametric methods. Four frequency range are denoted from the spectral analysis of the HRV: Ultra Low Frequency (ULF, frequency lower than 0.003 Hz), Very Low Frequency (VLF, power in frequencies range 0.0033-0.04 Hz), Low Frequency (LF, power in frequency range 0.04-0.15 Hz) and High Frequency (HF, power in frequency range 0.15-0.4 Hz) [77]. The LF reflects the sympathetic activity while the HF reflects the parasympathetic activity [78]. Ultra-short-term frequency analysis research done by Salahuddin et al, 2007 [79] showed that the frequency analysis of 50s segment length of the RR data statistically contains the same information of longer registration (2-5 minutes).

Therefore, to sum up, the HR and HRV signals are uniformly considered to be accurate measures in order to objectively assess the ANS state which is going to be further discussed in this document.

3.1.1 HR-Biosensor

In order to record the electric cardiac activity an heart rate sensor is used. The biosensor available and used is a wireless chest strap sensor. The sensor is then composed by a transmitter and a receiver.

The **transmitter** is the Polar T34 heart rate sensor which is composed of three electrodes, one act as a reference electrode. The electrodes do not need the utilization of a conductive gel, however, moistening the electrode is always recommended to enhance the conductivity. An adjustable elastic strap is further used in order to ergonomically support the electrodes for secure sensing. It electrically detects the heartbeat and wirelessly sends the heartbeat pulse to the receiver through a low-frequency electromagnetic field. Figure 3.3 depicts the transmitter component with the chest strap.

Figure 3.3: Transmitter of the heartbeat sensor [80].

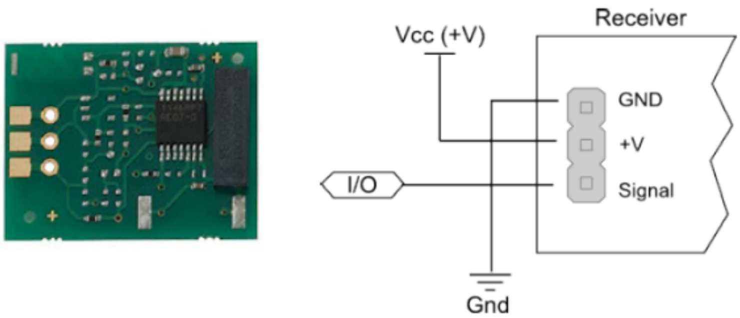


Figure 3.4: Heart Rate heartbeat receiver.

The **receiver** instead, Figure 3.4, is the Polar heart rate receiver component, which is a stand-alone module equipped with power input (3.3 V to 5 V), ground, and signal output. The system wirelessly receives data from the Polar transmitter component. The receiver, that is fixed on a breadboard base, receives the pulses sent by the transmitter and generates a corresponding digital output pulse. Together, the receiver, the sensor system and the transmitter provide a low-cost heart rate monitoring system which can be easily interfaced to the most microcontroller.

Main characteristics are:

- Power voltage: 3.3-5.5 VDC , 200 μA (at 5 VDC)

- Communication: 5 kHz uncoded or coded

- Length and width: 0.75 in x 1 in
- Wireless receives data
- HIGH digital output for heartbeat received, 15 ms TTL high level pulse on received heart beat (+V)
- 1.5 meter of working distances.

An example of recording, and computed distance between two consecutive heartbeats detected is provided by Figure 3.5.

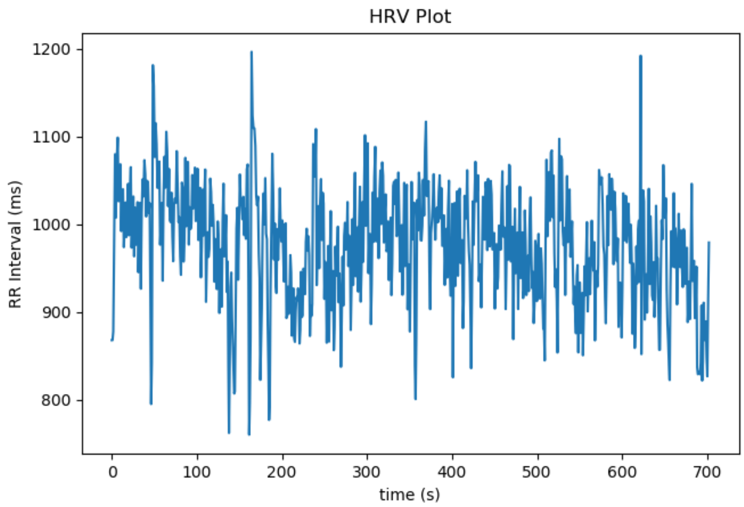


Figure 3.5: Picture represents an HRV signal, on x axis the time is reported and on y axis the RR-interval in milliseconds.

3.2 Galvanic Skin Response

The sphere of interest, in the current section, is the skin response itself or also known as the electrodermal activity (EDA) or galvanic skin activity. With term EDA or galvanic skin response (GSR) we are actually referring to the same concept. Interest in the skin conductance among electrodes arose because of the engagement of the sweat glands in such a measure. The sympathetic nerve activity influences the sweat gland activity, therefore this measure has been recognised as an ideal method to monitor the ANS. However, the foundation for the galvanic skin response is not fully understood and much is yet to be discovered to completely define the physiological phenomena.

From the anatomic point of view, Figure 3.6 shows the main peculiarities of the skin structure. The most external layer is named the epidermis and is composed by the basal layer, the prickle cell layer, the granular layer, the stratum lucidum and the most external the stratum corneum. This last layer, which corresponds to the skin surface is constituted of dead cells and living cells. Instead, blood vessels are located in the dermis area whereas the eccrine sweat gland secretory cells are located at the edge between the hypodermis and the dermis. The excretory canal of the eccrine sweat glands is a tube composed by of a one or two layer of epithelial cells; reaching the skin surface. The epidermis, ordinarily from top to bottom, has a high electrical resistance, the thickness of the layer of dead cells increases with the electric resistance. Such a feature is predictable since the main purpose of the skin is to represent a barrier against the external environment. Nevertheless, the sweat tubes from underlying cells penetrate the stratum corneum; these tend to decrease the electrical resistance considering that sweat has a good electrical conducibility property.

The actual skin conductance can considerably change depending on the past and present eccrine activity. This behaviour is especially prominent in the plantar and palmar regions because of the eccrine glands are remarkably dense while the thickness of epidermis is high: Generally, the conductibility range between tens and few hundreds of *microSiemens*.

Therefore, the principal aim of the skin is to preserve the body from the environment. On one side the skin tends to inhibit the loss of water by the body. While, on the other side, the vaporisation of water in order to enable homeostasis in terms of body temperature must be promoted. These demands appear to be managed by the most superficial barrier layer (i.e. stratum corneum) which limits the loss of water to the external environment except by the sweat glands. The activity of sweat glands is controlled by the autonomic nervous system, in particular, the sympathetic activity. The electrodermal response measures the output of the sweat glands activity, therefore representing an indirect measure of the sympathetic system [81].

Hence, it is well established that galvanic skin response is directly linked to the sweat gland activity. Pieces of evidence are represented Fowles et al, 1986 [82] studies in which a direct correlation between stimulation of sweat gland activity and GSR has been seen, in addition, the absence of GSR signal when the sweat gland activity is silenced.

There are two common measurements to quantify the EDA, the endosomatic and the exosomatic methods. The *exosomatic* method manages to apply from an external source a small current (limited to 10-15 $\mu A/cm^2$ in order to not damage sweat ducts [83]), which can be direct current (DC) or alternating current (AC), undergoes the skin. Then, the resistance opposed by the skin surface to the current passage is measured. The *endosomatic* method instead, do not use an external current and it measures the potential differences skin surface, named skin potential. Nowadays, the exosomatic method is mostly used in the EDA literature in which the reciprocal of resistance, the skin conductance in *Siemens*, is measured. An explanation is given by the more difficulties related to the interpretation of the recorded skin potential signals[84]. Skin conductance is preferred to the skin resistance because the sweat ducts represent small variable resistors in parallel, thus the whole conductance of a parallel circuit is easily measurable as the sum of every conductance.

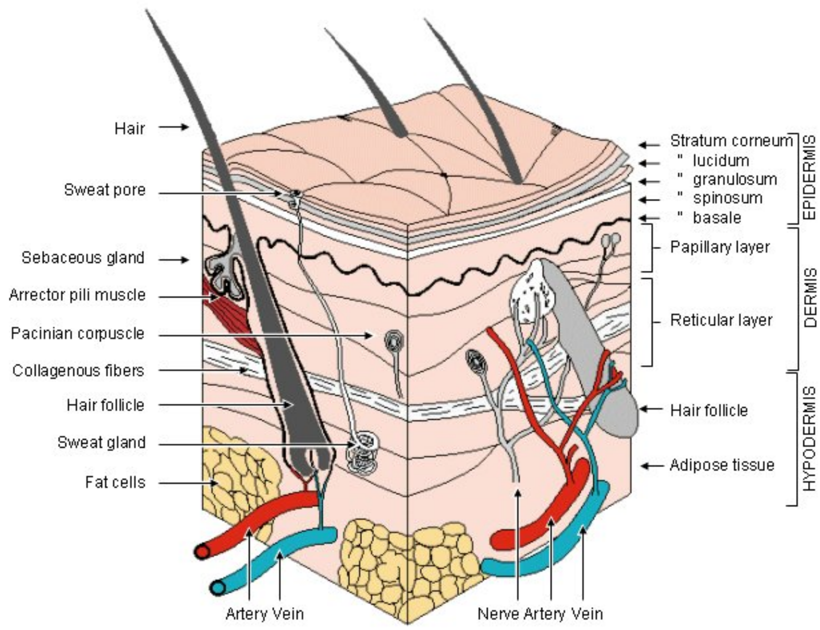


Figure 3.6: The skin structure is depicted in the figure, three main component are highlighted:epidermis,dermis and hypodermis [81].

One of the most primary features of the skin conductance signal is that it can be separated in two principal components: the *Skin Conduction Level* (SCL) and the *Skin Conduction Response* (SCR). These properties respectively represent the tonic and the phasic phenomena. The SCL extent the overall conductivity of the skin over large time intervals, usually varying from ten seconds to ten minutes. It is a measure of the general level of arousal or habituation to the environment and general affective states [85]. On the other side, the SCR is the galvanic response to the arousal in a time scale completely different from the SCL. More specifically, the skin conductance response can be noted when the sudomotor nerve is activated. Therefore, such a measure extent the sympathetic activation to respond to a stimulus. Figure 3.7 shows an example of raw EDA signal in which both SCL and SCR are visible, the signal is recorded from a subject during a high-stress session during a virtual driving experience.

Given the slow behaviour of the SCL component, the helpful information carried by the signal ranges from 0 to 0.05 Hz. Likewise, the SCR component has faster behaviour than the SCL, information ranges from 0.05 to 1.5 Hz [86]. Figure 3.8 represent the typical shape of the raw SCR and filtered SCR.

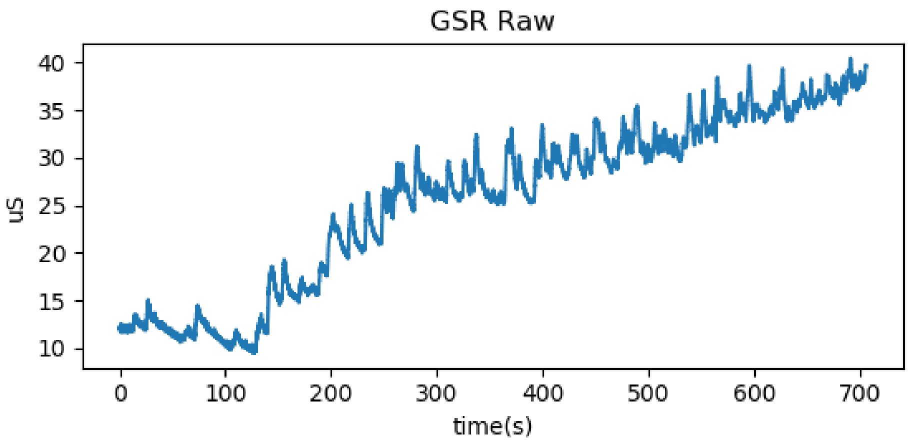


Figure 3.7: Raw EDA signal recorded in high-stress session during a motion sickness experience. The increasing trend of the signal is notable, representing the SCL component, while sharps peaks represents the SCR component. Increasing trend of SCL and high number of SCR peaks are stress indicators.

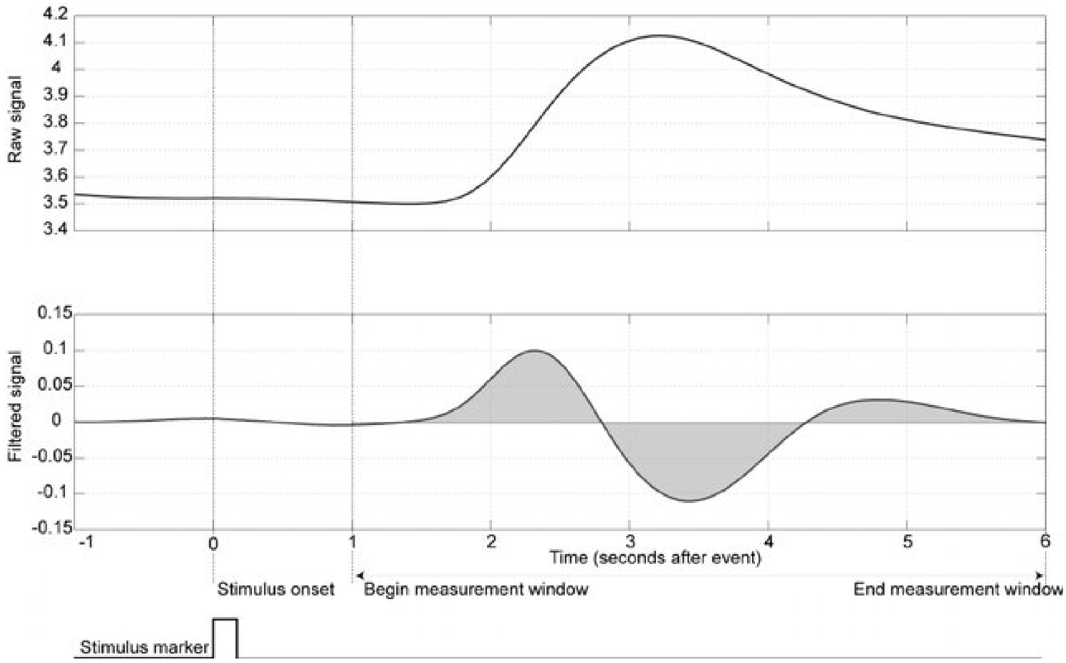


Figure 3.8: Raw SCR and filtered SCR, usually the peak is located few seconds after the stimulus input [87].

Hence, the tonic components do not respond to the stimulus but to the general emotion or states, the phasic component is the counterpart; it is predominant when a stimulus is applied to the subject. For the purpose of this thesis project, both SCL and SCR can be taken into consideration. More interest in scientific literature is toward the SCR thanks to the direct relationship with the sympathetic system. SCR peaks are not present only when a stimulus occurs, also spontaneous (also named non-specific) SCR appears on raw EDA signal. Such responses are typically present one-three times in rest period [88]. A practical rule whether an SCR is non-specific or specific (linked to a stimulus) suggests that SCRs that start after 5 sec from the stimulus onset should be classified as non-specific [87]. Generally, because of the direct relationship between the electrodermal activity and the sympathetic activity is well-established, the GSR signal provides an optimal source of information related to the physiological state.

3.2.1 GSR-Biosensor

In order to record the electrodermal activity, the Groove GSR has been used [89]. It measures the resistance of the skin to the passage of the current from the finger. Figure 3.9 depicts the biosensor.

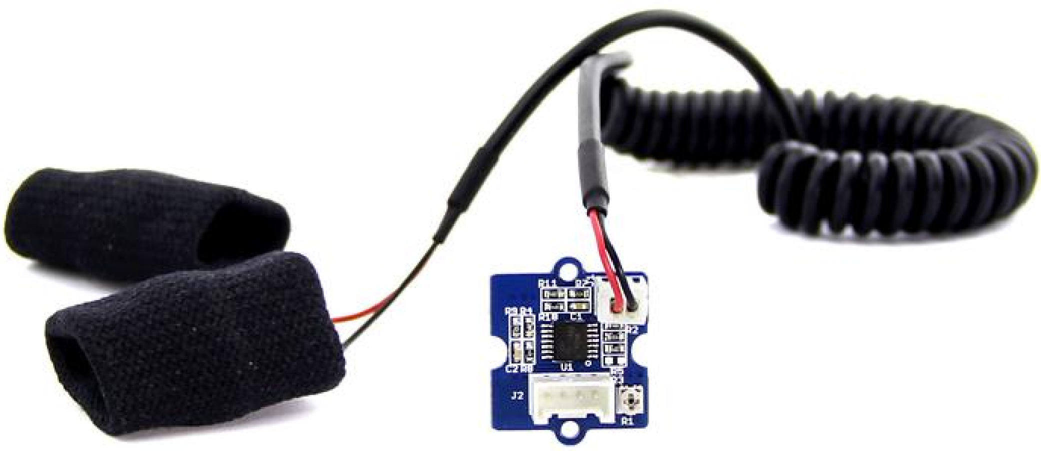


Figure 3.9: The groove biosensor [89].

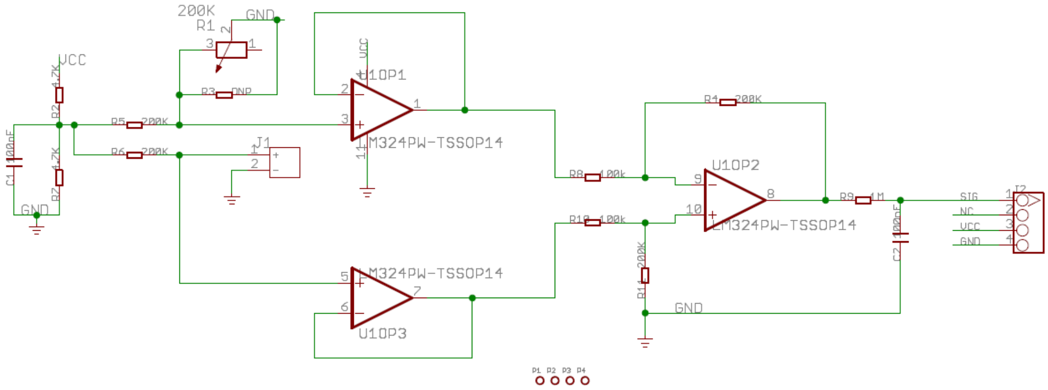


Figure 3.10: Electric circuit of the Groove GSR sensor [89].

The electric circuit is represented in Figure 3.10. The resistance value is translated into voltage value and outputted.

Specifications are:

- Working voltage ranging 3.3V-5V
- Adjustable sensitivity through a potentiometer
- Resistance as the signal input
- Signal voltage is the output, analog reading
- Electrodes made up of Nickel

The formula to calculate the skin conductivity is the following:

$$R = \frac{(1024 + 2 * V_{out}) * 10^3}{(512 - V_{out})}\,(\Omega)$$

$$G = \frac{1}{R}\,(S)$$

Chapter 4

Simulator and Technologies

Form the consideration reported in the State of Art chapter; it has been established the utility of HUD technology to build trust on the autonomous machine [90]. Thus, the presence of such techniques inside a semi-automated car is fundamental. Within this document, it is studied the utility of the HUD system for a fully automated vehicle, with particular attention toward the amount of information displayed by the HUD system.

For this aim, the following chapter presents an overview of the technology implied in this master thesis project. Two main components need further observation, an *immersive virtual driving simulator* which concerns the equipment required for virtual reality experience, description of the simulated driving experiences and differences between user interfaces. Last but not least, the *data acquisition hardware* needed to store the physiological values.

The immersive virtual driving simulator has been recently developed by two master student from the Polytechnic of Turin [91][25].

4.1 The Driving Simulator

The driving simulator is the expression adopted to indicate those kinds of video games in which the user must drive a virtual vehicle. Usually, the software uses a physical engine that can simulate the real behaviour of a vehicle in every situation.

Most video games are arcade-type; it means that physics is greatly simplified to make the game user-friendly. Suspensions or stresses from the tires are often not replicated. Other video games, instead, are simulation-type. The software attempts to reproduce as closely as possible the physic laws of the real car. For instance, advanced simulators are used for training professional pilots of the Formula 1 tournament, to faithfully reproduce the driving experience to pilots, saving time, money and risks.

With the term immersive virtual driving simulator, we refer to such simulators in which the user drives within a virtual environment thanks to equipment like the Head Mounted Display (HMD) in an environment that elicit the sense of immersion to the user.

For this thesis, an immersive driving simulator recently developed at the Polytechnic of Turin, from another thesis project, has been used. Among the various simulators, reported in the Chapter of the State of the Art, the open source driver simulator produced by GENIVI Alliance was utilised and equipped with built-in modalities. The simulator presents the possibility to choose the autonomous driving modality thanks to an algorithm developed by such a thesis project.

As already mentioned, the GENIVI Vehicle Simulator project is an Open Source project for Unity [29]. The project and the initial software code were developed by Elements Design Group of San Francisco and the Jaguar Land Rover Open Software Technology Center in Portland, Oregon. The intention was to provide an open source extensible driving simulator for the development community. The primary objective was to create an application to assist in the development and verification of In-Vehicle Infotainment (IVI) systems. The simulator provides three different driving scenes: Yosemite, Pacific Coast Highway and San Francisco, depicted in Figure 4.1.

A discrete level of accuracy characterises all the virtual environments according to the purpose of this thesis work. The coastal environment represents the Pacific Coast Highway, a quiet path along the ocean, characterised by bridges, mountains and slope changes. Instead, the urban environment is vast and represents a portion of San Francisco city. Many streets composed the virtual town, buildings, skyscrapers and various types of buildings, intersections, signs, traffic lights, the famed Golden Gate Bridge and generally all the objects which can be observed in a city.

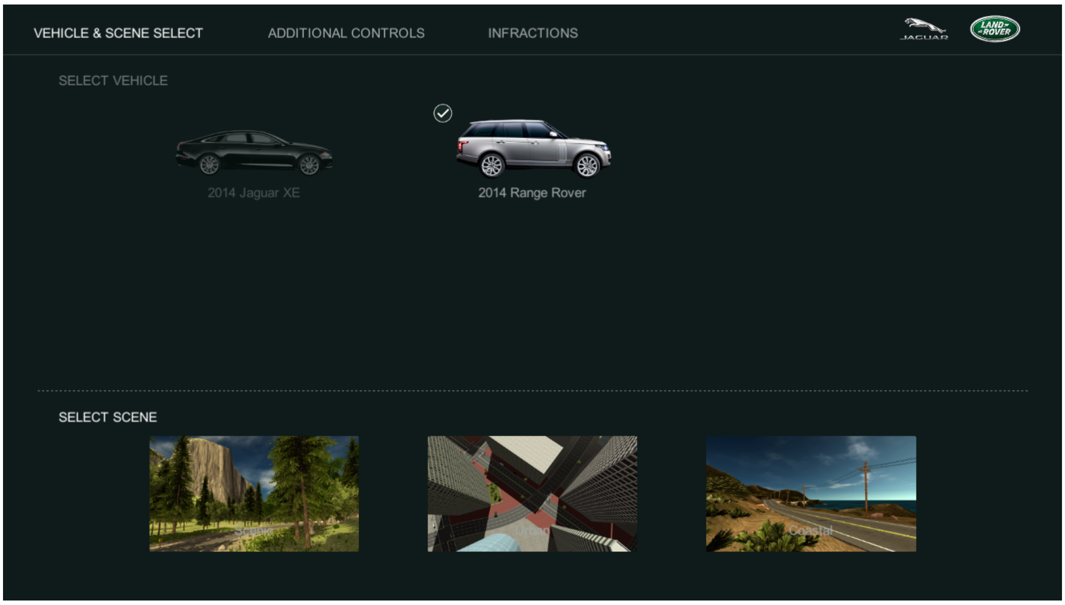


Figure 4.1: The vehicle and scenes possibility selection provided by GENIVI.

.

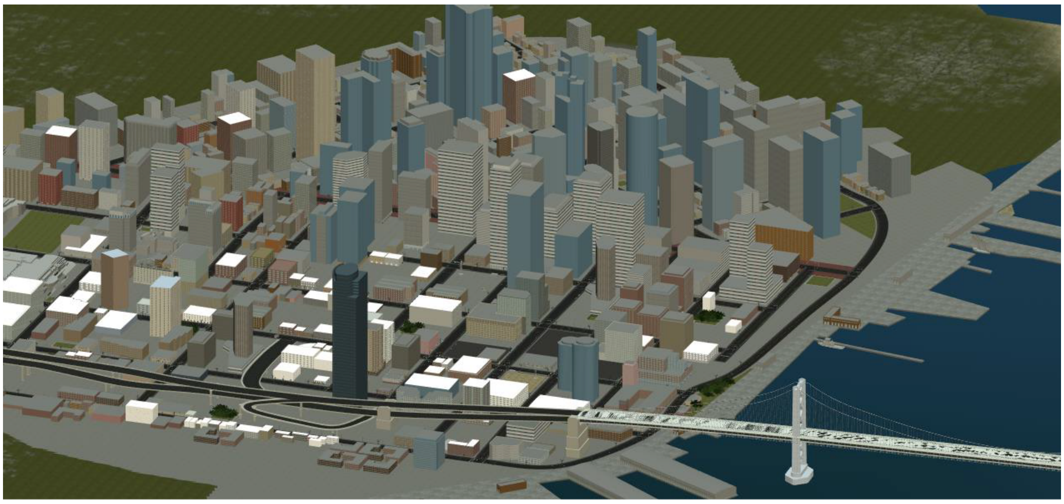


Figure 4.2: The reproduction of San Francisco city in Unity.

.

4.1.1 VR Equipment

Before proceeding towards a detailed explanation of the virtual driving experience, a brief overview of the virtual reality equipment. Using a driving simulation software, without adequate hardware, does not allow an accurate and truthful simulation. Thus, it is necessary to have the proper hardware to elicit the sense of presence to the passenger within the virtual autonomous driving experience.

HTC-VIVE

Virtual reality devices utilised for the realisation of the immersive experience are the HTC Vive system, designed by Valve in collaboration with HTC, which through an Optical viewer allows you to see the simulated environment. The system is composed by Vive Headset, Vive Controllers, Vive Base Stations and Vive Tracker.

Vive Headset has two AMOLED panels with 3.6 "diagonal, each with a resolution of 1080x1200 pixels per eye, 2160x1200 total resolution. The display is updated at a frequency of 90 Hz and has a viewing angle of 110°. Within the headset, IR sensors are fixed which detect the infrared pulses from two base stations to determine the current position in the space [92].

Vive Base Stations are two black boxes which produce a 360-degree virtual area up to 15x15 foot range. Timed infrared pulses at 60 Hz are emitted from base stations and then detected by controllers and the headset and with high precision (sub-millimetre sensitivity) [93].

Vive Controllers have many input systems like trackpad, buttons and other. The controller is equipped with a battery, six working hours for a charge.[26] At the top, 24 infrared sensors are fixed to detect the location (detectable fraction of a millimetre) of the controller by pulses emitted from the base stations [94].

Vive Tracker acts as motion tracking accessory, it is therefore designed to be fixed in some particular location of interest to be tracked thanks to the IR base stations pulses.

Figure 4.3: HTC Vive system [95].

.

Simulatore Atomic A3 Racing

Motion Platform

Another fundamental step to produce a highly realistic driving simulation is to use an inertial platform. Motion platforms can support movement up to 6 degrees of freedom different. These represent the three degrees of translation (i.e. surge, heave and sway) and the three degrees of rotation (i.e. roll, pitch and yaw), Figure 4.4.

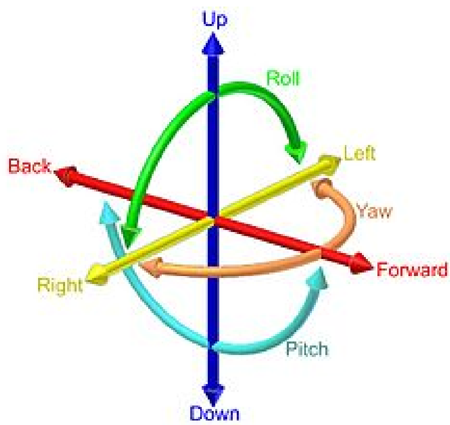


Figure 4.4: 6 degree of freedom [96].

.

The motion platform adopted for the virtual driving experience is the Atomic A3 Racing, designed by Atomic Motion Systems, a company that builds ultra-compact high-performance driving simulators for home and commercial use [97]. It is a motorcycle platform which provides a high fidelity movement and powerful accelerations to obtain an immersive driving experience.

Figure 4.5: 6 degree of freedom [98].

.

Mainly, during a real driving session, the most perceived forces are the lateral and longitudinal ones. For such a reason, the Atomic A3 Racing motion platform has only 2 degrees of freedom (i.e. yaw and pitch) supported by two actuators with high speed and high precision.

Leap Motion

Another method utilised to enhance the immersion of the user whether as a driver or as a passenger of the autonomous system is the possibility to visualise their hands and movements in the virtual environment. It has been established that the presence of the user's hand increases the sense of the presence of the subject [99]. For this purpose, a device named Leap Motion Controller was used. It can be attached directly on the Vive headset, and it can identify the hands and track their movements thanks to the sensors inside the device.

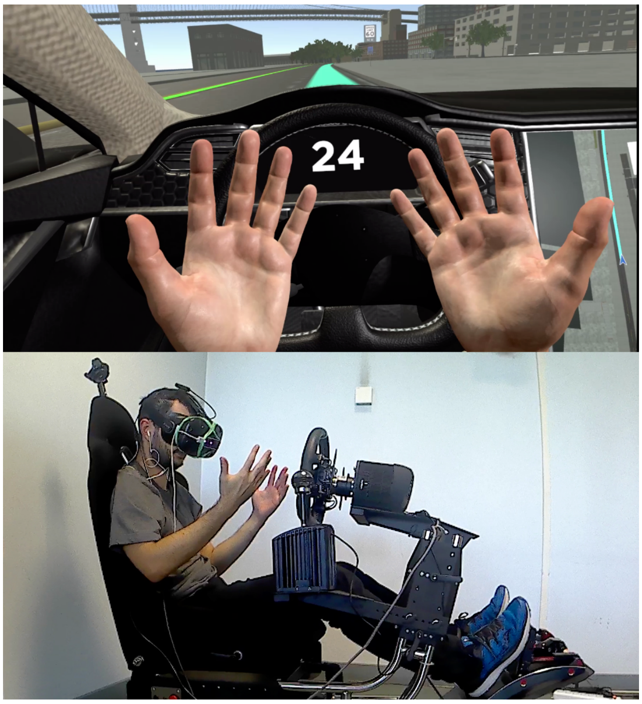


Figure 4.6: Leap mostion system. At the top the view in VR, at the bottom the outside view [91].

4.2 VR Driving Experience

After the unavoidable introduction on the virtual driving simulator and its related technologies to ensure an adequate immersion to the user, in the following pages the detailed description of the driving experience simulation is provided.

In the previous section, it has been already mentioned that GENIVI software offers the possibility to select between three different driving scenes: Yosemite, Pacific Coast Highway and San Francisco. For the aim of this thesis project, the most interesting scenario is the urban scenario, i.e. San Fransisco. This decision has been made not only for the greater possibility to test the autonomous driving performances but also to promote the engagement of the subject to the driving context increasing the cognitive load related to the urban scenario. Additionally, the urban contest supplies also the chance to exhibit the system ability to face a traffic light, road signs or driving rules. Instead, the Range Rover has been selected for the travel.

It was thought to insert into the simulated experience not only everyday driving events but also some specific situation. Such situations have been set up to allow the user to understand whether or not to trust in the autonomous car. Mainly, the urban scenario provides more flexibility to add particular situations into VR experience.

Concerning the everyday driving situations, it was decided to include pedestrian crossings at a traffic light and in correspondence to a stop sign, and other cases in which the ability of the car to take over the traffic lights and the vehicles preceding it is shown.

The global duration of the scenario was set at about 12 minutes. Such a duration length is justified by the evidence that at an initial phase, the user takes in confidence with the autonomous vehicle. For this reason, the experience begins ina little trafficked area and in which amount of elements within the environment is limited enabling the user can to evaluate the behaviour of the car and determine whether to build a relationship of trust with this.

Over time, an increasing amount of virtual objects appear into the VR, enabling the user to take confidence towards the HUD system. Following, dangerous situations are programmed to occur in which the car shows its decision-making skills and communicates them to the user through information. The simulation proceeds, alternating risky situations with some ordinary ones, to show to the user how the car face a typical urban context.

To sum up, the autonomous car can manifest to the user his decision-making skills through the driving actions undertaken and through the user interface (HUD) that displays the interpretation of the current driving situation. More specifically, it has been thought to insert specific dangerous situation into the driving experience, giving the possibility to the car to prove its capability to assess risks and take appropriate countermeasures to protect the passenger.

To be clarified, to recreate comparable user-experiences, pre-programmed events and path followed into the virtual driving experience were the same for all the subjects which took part in this thesis project. It was added one random factor between different driving simulation: the colour of the traffic car may change.

Furthermore, one other factor is different into the driving simulation; the HMI selected, omni-comprehensive or selective HUD on which the this thesis' focus besides.

4.2.1 Simulation User Interface

Once described the driving simulator with relative devices, the scenario and the general description of the driving experience, it is necessary to examine the autonomous car user interface deeply. Thanks to the literature studies previously performed, two kinds of user interfaces have been developed [91][25]. Generally, information displayed in both user interface are:

- Dynamic and static elements within the environment;
- Speed-related information, distance and dangerousness level of elements using color scale;
- Traffic lights and road signs;

- Own navigation line and other vehicles as information about paths;
- Route of the road;
- Center-line.

Therefore two simulation user interfaces have been developed, the omni-comprehensive and selective.

Omni-comprehensive

Firstly, the characteristics of the omni-comprehensive user interfaces are illustrated. Globally, such a user interface provides a considerable amount of information related to the virtual environment.

More specifically, the HMI system shows both static information like traffic lights, parked cars, etc. and dynamic information like pedestrians, animals or other vehicles. HMI system also provides information about the distance, expressed in meters, between the object and the autonomous car and the indication of the absolute speed of that object, in km/h.

Additionally, the system also provides identification information for each element, so that the user realises that it recognises what surrounds the car. Moreover, the HUD system shows information about the road signs and relative distance and information about the traffic lights which directly influence the current autonomous driving actions. As for the other objects, the car identifies the type of signal or the status of the traffic light and shows it to the user. Subsequently, the vehicle takes the appropriate decisions, such as stopping in the event of a stop signal or a red light.

Furthermore, the car shows to the user that it can understand its road route, through information in the form of a navigation line and also shows the trajectory acquired by the cars circulating on the road at that moment, through the same type of representation. Moreover, the information regarding the identification of the direction-of-travel separation line offers the user a further understanding of the car's ability to understand the road section it is being driven on, identifying the lane's travel direction and its position on the lane.

Concerning dynamic elements, the omni-comprehensive HMI displays information about all the dynamic elements which are inside a detection diameter (150 meters). However, a certain level of selection is present, static elements, like a parked car, are not highlighted from the HUD. Moreover, the HUD also ignores the dynamic element which cannot involve in any way the current driving task, for instance, a traffic car which is moving away from our car.

The HUD system aims to show up itself to the user as a competent system since it demonstrates to be able to monitor the environment in its entirety and demonstrates to be able to manage various driving situations.

In general, the elected information characterise the omni-comprehensive user interface, since it provides a considerable amount of information relating to the driving context.

Selective HUD

Concerning the selective user interfaces, the general idea is to provide a determined subset of information of the omni-comprehensive HUD system. In particular, the main selection criteria adopted is to consider if an element is about to influence the actual driving. For instance, regarding the road signs within the 150 m range, two signs are present: one informs to give the precedence to pedestrians for crossing the street, the other is a speed limit. The latter modifies the actual current behaviour of the car to set the speed limit to the machine; instead the second would influence the driving behaviour only in the case that a pedestrian crosses the street. Therefore, the HUD displays just the speed limit sign, the one which modifies the current driving behaviour.

Therefore, the information that has been decided to show is related to elements of the road context that affect driving at the current moment. Immediate dangers in the environment, a traffic vehicle that intersects the car's trajectory, road signs as mentioned earlier or the car that precedes the vehicle are the selected information which are displayed by the selective user interface. Instead, information presented for the traffic lights does not change between the two user interfaces.

In general, the selective HUD system displays all the information which modify the current driving behaviour; therefore potential risks are signalled once they became real risks for the passenger.

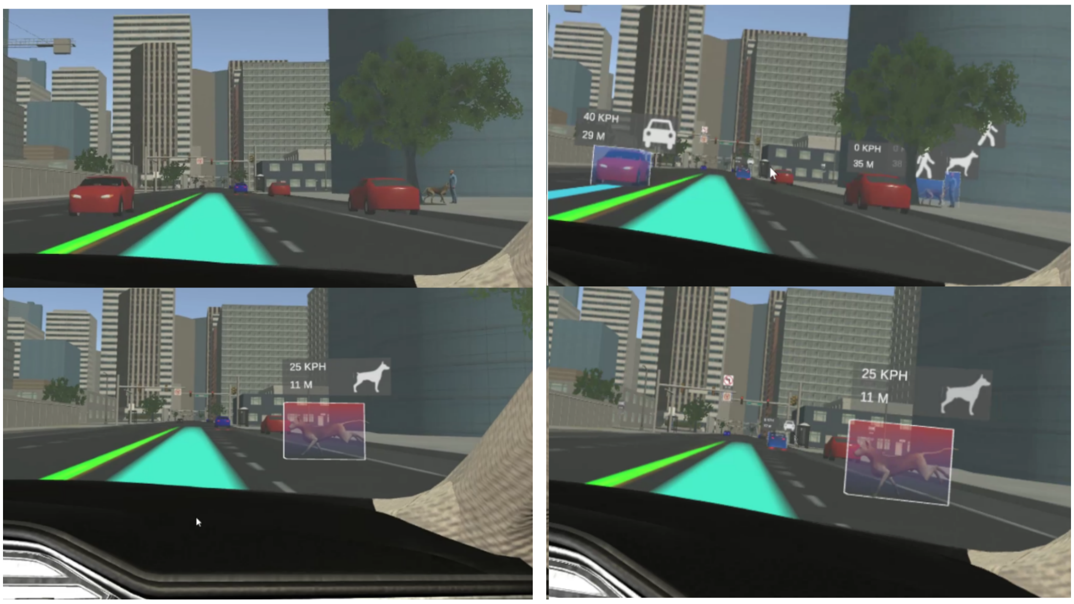
4.2.2 Test Events

To compare data between different subjects, the user experience has to be the same; one of the main reason why scientific literature vastly uses driving simulator is for the capability to program the test experience within a completely controlled environment. For this reason, it has been set up the complete driving experience to be the same.

In the previous sections, it has been overmentioned the fact that unexpected situations, namely test events, are programmed to occur. Such events represent typicals hazardous driving situation which can unfortunately happen. More specifically, seven test event are pre-programmed to occur which differ between each other in the degree of related risk.

In the first stage of the scenario, the car manages the driving on a non-straight road and various traffic lights. Traffic lights are also controlled via the event system, meaning that these are not free parameters either. The full control of the situation is a priority.

Once the subject has acquired a discrete knowledge about the virtual experience (at about 2 min from the start), the first event is programmed: a dog suddenly crossing the street, Figure 4.7. Such an event is a high-hazardous situation; the car not only is forced to fastly brake but also to steer in the opposite direction to avoid the rear-end collision. An adult and a child were also added to improve the likelihood of the situation, even the owners of the dog walk towards the road, without entering the lane, to show the intention of trying to stop the dog during his run. The dog barks when it begins its run. After every dangerous situation, like the one just mentioned, the car stops for a few seconds, so that the user has time to understand what happened and reflect on how the vehicle managed this situation.



(a) Selective user interface. **(b)** Omni-comprehensive user interface.

Figure 4.7: Dog test event.

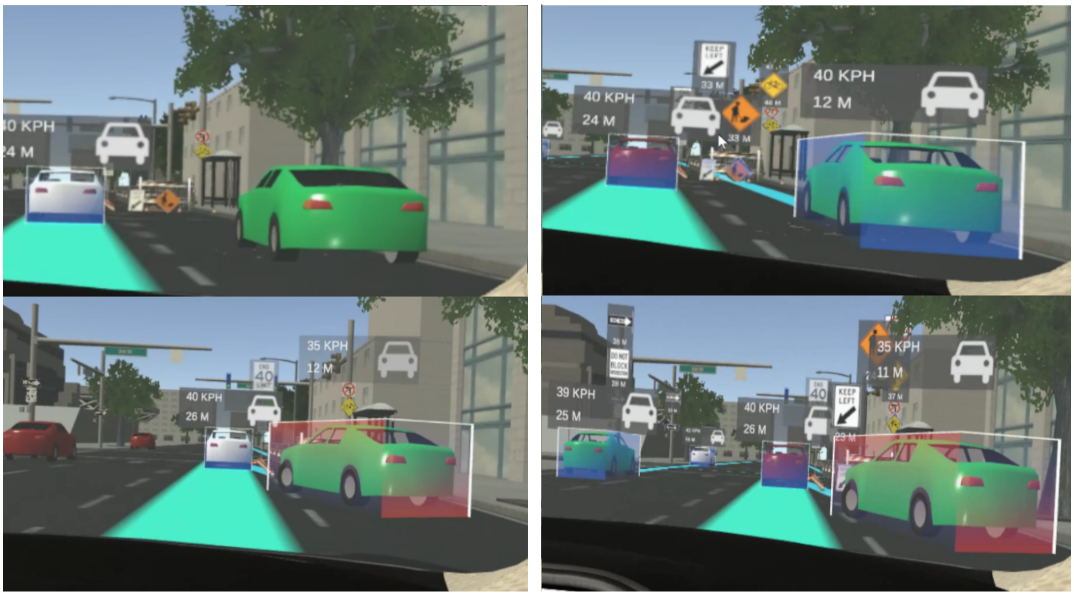
Proceeding, into the simulation a few situations of applications of precedence rules are programmed. Generally speaking, throughout the entire journey, the management of traffic lights, the interaction with traffic cars and the application of precedence rules are shown to the user.

Following an orange light, which does not allow the car to stop in time, thus the vehicle operates as a green light, the second test event occurs; a ball, thrown by a child, that crosses the road and forces to sharply brake and counter-steer. To improve the likelihood, in this situation have been inserted other children and adults who monitor them, as well as two sources of sound: noise in the background of kids playing in the park, and a scream of a man when the ball was thrown, Figure 4.8. Such a situation is discussed amply in the literature since the autonomous system could be faced to choose between rear-end collision with two different objects, the algorithm to give weight on the related object is applied [25].

(a) Selective user interface. (b) Omni-comprehensive user interface.

Figure 4.8: Ball test event.

Later, in the path, there is a building site for work in progress which blocks the right lane of the car. The road signs warn the vehicles about the required lane changing left and to decrease the speed. A traffic car which was on in the lane of the building site cut off the road. The autonomous vehicle is forced to reduce the speed to avoid the collision. Such a situation represents the third test event, Figure 4.9.



(a) Selective user interface. (b) Omni-comprehensive user interface.

Figure 4.9: Car1 test event.

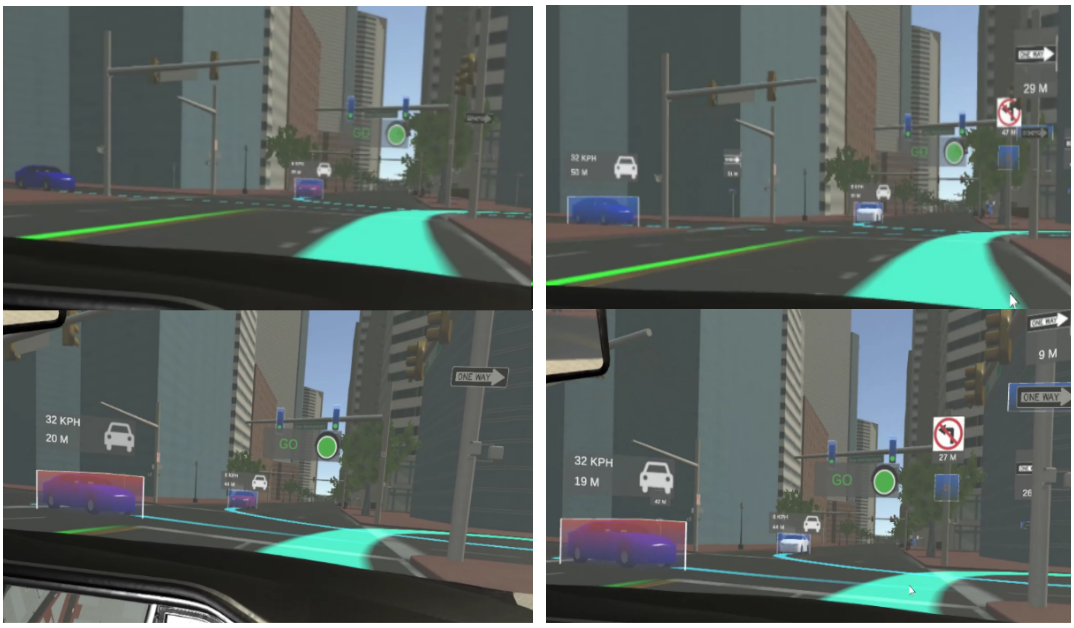
Beyond the area of work in the courses, the traffic car that precedes the vehicle is still the one which cut off the road. Such a traffic car shows some driving difficulties again; therefore, safely, the vehicle decides to overtake the traffic vehicle. This situation highlights an adequate decision-making capacity of the car.

Meanwhile, the scenario becomes increasingly busy. The path that follows does not present situations that are particularly relevant. During this phase, the most exciting situation is represented by a traffic car that precedes the vehicle which have to turn right at the next intersection. Therefore, it slows down and forces the autonomous vehicle to slow down; during the braking, it is possible to hear the noise of a scooter on the left of the car, which cuts the road and enters the lane where the driver-less vehicle was continuing. This situation, which corresponds to the fourth test event, is not especially threatening; the displayed colour indication of the scooter, which tells about the degree of danger, is yellow/orange, red instead is adopted to indicated more hazardous situations, Figure 4.10.

(a) Selective user interface. **(b)** Omni-comprehensive user interface.

Figure 4.10: Scooter test event, no much difference in the user interfaces is present.

The fifth test event it should be perceived as extremely dangerous, at the moment when the autonomous car is approaching an intersection, another traffic car, at full speed, does not respect the traffic light and suddenly cuts the road, forcing the passenger-vehicle to brake abruptly and to steer to avoid impact. Therefore the situation should be perceived as very threatening because the traffic car, ignoring the traffic light, appears in the field of view of the user suddenly and unexpectedly, Figure 4.11.



(a) Selective user interface. **(b)** Omni-comprehensive user interface.

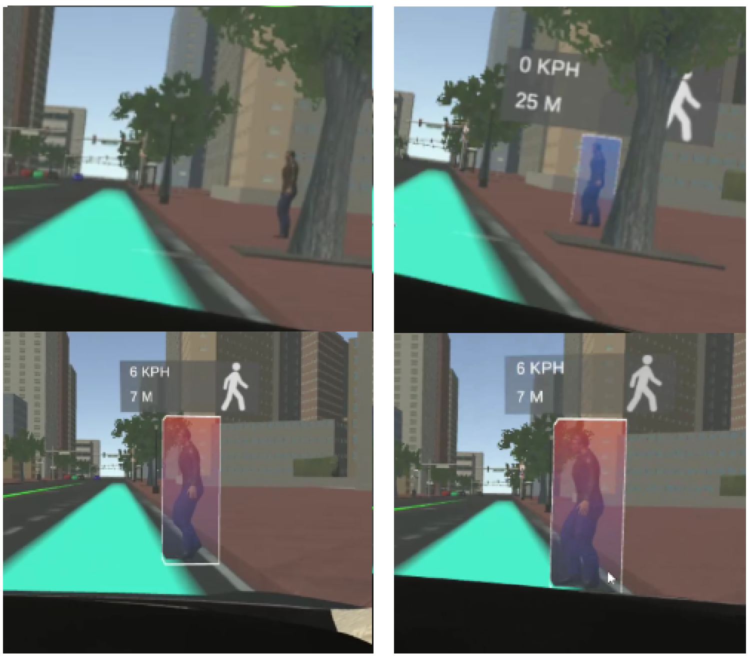
Figure 4.11: Car2 test event.

Instead, the sixth test event concerns a simple crossing of a pedestrian. The car is about to stop at a stop sign and the pedestrian, even before the car stops, begins to cross. This situation, like the fourth test event, is not particularly dangerous since the automated vehicle is already braking and the pedestrian passes over the stop line, Figure 4.12.

Finally, the last situation should provide the most hazardous situation, also suggested by the literature. The case concerns the crossing of a pedestrian in an not allowed area. The walker, behind a tree, unexpectedly crosses the street when the car was at full speed. An abrupt fast-braking action is performed by the vehicle, Figure 4.11.

(a) Selective user interface. **(b)** Omni-comprehensive user interface.

Figure 4.12: Man1 test event.



(a) Selective user interface. **(b)** Omni-comprehensive user interface.

Figure 4.13: Man2 test event.

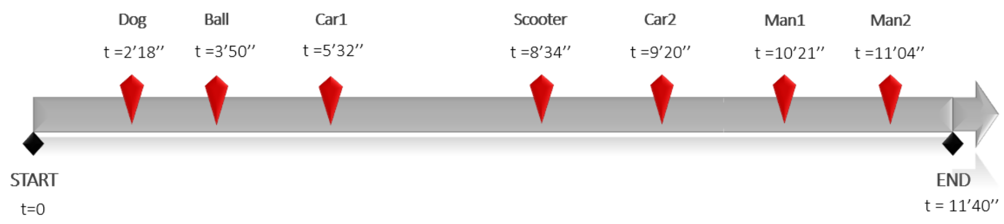


Figure 4.14: Test event timeline.

4.3 Data Acquisition

The following section aims to produce an overview of the data acquisition hardware. Not only the driver simulator and its equipment are necessary, but it has also been seen that physiological measures provide an affective states indicator, for instance, to assess the degree of stress related to a specific situation. The driver simulator, and explicitly test events, as the source of stress. A quantitative measure of such measures under test events represents the primary purpose of this master thesis to help HUD designer.

For this purpose, it is necessary to record and store data between different user-experiences. Therefore the necessity of a memory where to store the recorded data is a priority. The storing device has to be capable of communicating with sensors described in Chapter 3. Additionally, the storing device should also be portable to transfer the data logger in a real autonomous car eventually.

Such consideration drove to a choice between two very different available platforms: Arduino YUN or a Raspberry Pi. Arduino is a low-powered platform initially based on Atmel ATmega microcontrollers. Instead, Raspberry (also named RPi) is a low-cost, single-board computer (SBC) produced by the Raspberry Pi Foundation. The SBC Raspberry Pi is a miniaturised Linux computer based on an ARM core working at clock speeds in the order of 1GHz and upwards.

Pro and contra for both of them, from the Arduino side the main advantage is represented by the already integrated ADC (analogue to digital converter) which allow the device to read the analogue input from a sensor (the GSR) and convert the value. The hard-realtime capability represents another interesting feature, considering that the microcontroller has only to run the uploaded code.

On the other side, Raspberry does not allow the hard-realtime and ADC either. However, being an SBC, it is empowered of a significant computational power compared than the Arduino; therefore such a platform is particularly suitable when data elaboration is required.

For this thesis work, initially, it has been considered to work with both Arduino YUN [100] and Raspberry Pi 3 Model [101]. Therefore, the Arduino was used as ADC, and data was sent via USB to the Raspberry for storing and elaborating if requested.

However, the timing was critical considering that the heart rate sent only when a heartbeat is detected, the time has a fundamental role in computing the heartbeat distance between two consecutive beats. The serial communication does not allow to have such good control of timing operation; therefore the combination of the two platforms was re-considered. The possibility to have a portable miniaturised computer has motivated the choice: Raspberry Pi was then selected.

The absence of an ADC natively integrated in RPi represented an issue which had to be solved. The solution was to integrate an external ADC (MCP3008 [102]) with already developed API provided for Raspberry which makes use of SPI communication. In Figure 4.15 it is shown the whole acquisition system with the sensors.

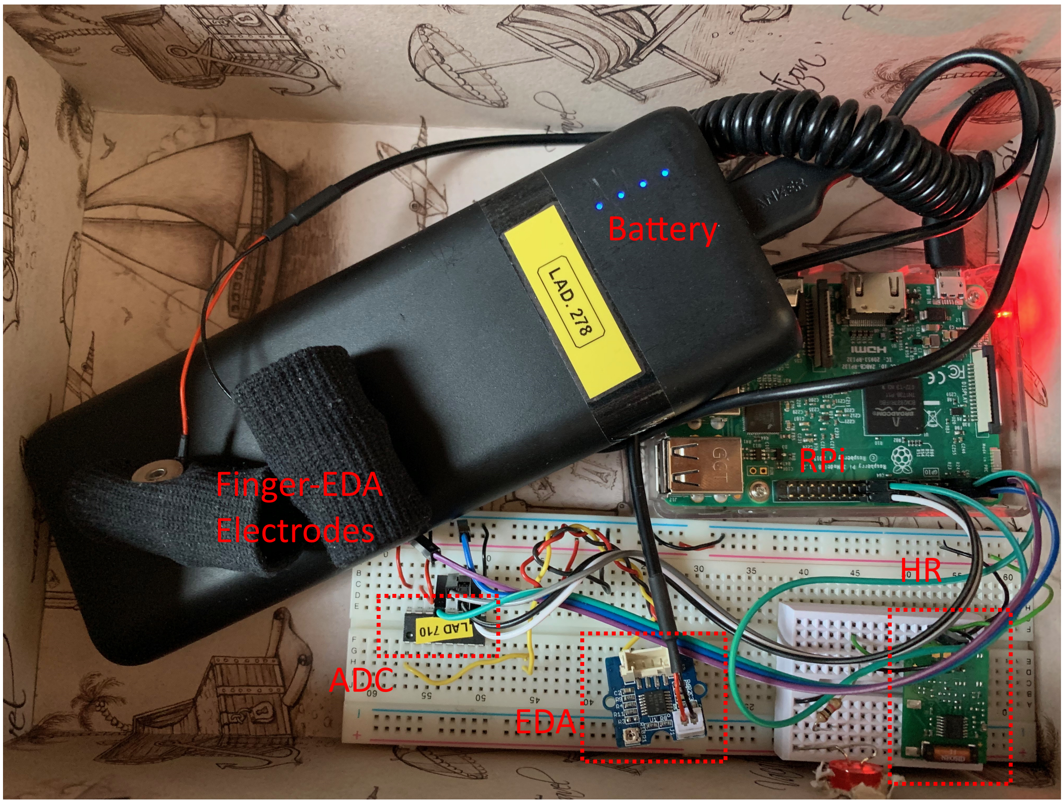


Figure 4.15: Complete data logger with bio-sensors.

.

Chapter 5

Methods

In order to properly compare user-responses between the two user interfaces samples, a straightforward approach has to be followed. In the current chapter, it is presented the experimental and technical procedure followed to record, collect and analyse the data.

Therefore, firstly an overview is provided of the participants who took part in the test, secondly the experimental procedure is illustrated, and lastly, the data analysis is pointed out.

5.1 Study1 and Study2

For the purpose of this thesis project, a preliminary study, named *Study1*, has been conducted. Such work was found particularly useful to lead the final *Study2* results. Therefore, Study1 includes all the analysis performed in Study2 with a sample of 36 subjects, 13 of them experienced the selective HUD, while the other the omni-comprehensive HUD. Results obtained from such a study reported significant differences in term of GSR data between the HUD while the questionnaire delivered in such a study showed mean values that generally tell the user preference toward the omni-comprehensive HUD in term of trust. *Study1* results were utilised to set the *Study2* experimental protocol properly. Two main points were modified:

- Standard Operating Procedure (SOP)
- Questionnaire

However, a qualitative comparison between the GSR data has been reported in Result Chapter. Therefore, in this thesis only the Study2 is discussed.

5.2 Participants

A number of 39 healthy individuals volunteered to participate in the virtual driving experience. One exclusion criteria was adopted which concerns the presence of motion sickness. In literature, subjects who report motion sickness are always excluded from the dataset, mainly if the focus is to evaluate the physiological response quantitatively; it would be biased from the sickness state [103]. One subject was excluded from the data set because of the motion sickness experienced; he expressed a high nausea sensation and stomach ache later completed the test. In addition, because of recording issues related to the registration bot HR or EDA data, eight subjects were rejected from the dataset. Participants randomly experienced the selective or omni-comprehensive HUD.

Moreover, personal data were also recorded: age, gender, user-experience with the Virtual Reality and traditional driving simulator and general information about their attitude towards a travel experience with a real level 4 or 5 automated vehicle. The adoption of an ad-hoc modified questionnaire was provided to evaluate the subjective user response and other related questions. Questionnaires still represent a commonly used method to assess the user-experience, even if the adoption of physiological measures can provide support, or ideally substitute questionnaires outcomes [54].

The questionnaire which has to be filled by the testers generally tries to assess:

- Motion sickness
- Evaluation of autonomous driving
- Test event evaluation
- Quality and quantity of Human-Machine Interaction information
- Trust
- Virtual reality experience

More specifically, eleven sections compose the questionnaire. The first section aims to collect personal information like gender, age, prior experience in virtual reality and driving simulator.

The second and third sections assess if the subject has experienced or not ”motion sickness” and to what extent. Such evaluation is done making use of the Simulator Sickness Questionnaire (SSQ) [104].

The third section evaluates autonomous driving; for this purpose questions taken from a questionnaire used to assess trust relationship between human and machine have been adopted [106].

Additionally, the fourth section aims to evaluate test event situations, Figure 4.14. Such a section aims to evaluate if the user perceived the danger/fear from the test event and to what extent. Such a section is particularly useful to check the correlation between subjective and objective responses and to explain, if the case, why there are differences between the two HUD.

Continuing, other sections aim to evaluate the user interface. As suggested by Ekman et al. [12], factors which directly influence the effectiveness of an

HMI are trust, the cognitive load and the situational awareness. Trust is, in turn, controlled by the mental model, *why* information and system competence. Therefore, questions aim to understand how clear, through the interface, the intentions and decisions of the car were (mental model), if it is always clear why the vehicle carried out a particular operation (*why information*) and about the perception that the autonomous system had everything under control (*system competence*). Later, the impact of the interface on the user's cognitive load was evaluated, based on the tool "NASA Task Load Index NASA-TLX" evaluation (NASA-TLX) [107].

Moreover, section eight evaluates the situational awareness, taking inspiration from the self-assessment questionnaire of the situation "Situational Awareness Rating Technique" (SART) [108]. More specifically, questions related to *quality* and *quantity* were ad-hoc modified. Quality refers to the usefulness of the information or how much the person has appreciated that such information existed to help for the situational awareness, for example, to make the journey comfortable or for safety concern. On the other hand, quantity refers to the number of information that the interface has provided to the passenger. From the comparison between the HMI, these latter questions enable us to understand if the user would prefer few or many information in a level 4 or 5 autonomous car.

Following the evaluation of the HMI aspects, the questionnaire was concluded with a direct question about trust and general questions about the whole user-experience, to have a confirmation of the results obtained from the previous.

The last two sections concern questions related to the simulation in virtual reality and aim to evaluate the quality of the simulation in term of immersion and fidelity. Such considerations are done because a negative subjects' response to the 'faithful to reality', would have lead to obtain insignificant results on the evaluation of the user interfaces. Thus, some sections of VRUSE [105] were adopted, which is a questionnaire used to evaluate the usability of a virtual reality system based on the user's attitude and perception. In detail, it was posed a greater attention on VRUSE sections which consider the sense of immersion/presence, which is the perception that the user has of being physically in a non-real world and the fidelity of the simulation, to evaluates how much the experience within the virtual reality has been comparable to a similar one in the real world.

The questionnaire was administered to users in the Italian language to avoid misunderstanding. Ratings range in 1-5, Likert-type scale; Appendix A provides the full version of the questionnaire.

5.3 Experimental design

The test phase began for each user with a brief explanation of the experience. Additionally, the subject was informed to wear a virtual reality headset and, as a passenger, he or she would have experienced a fully automated driving in an urban scenario.

The Standard Operating Procedure (SOP) is below summarized:

1. Preparation of the virtual driving simulator software.
2. Subject arrived in the laboratory, some minutes of familiarization with the environment (about 5 min) are programmed; it is provided an explation of the experiment.
3. Filling out the pre-test questionnaire.
4. Electrodes preparation.
5. HR sensor placement
   (a) Moisten the three electrode areas.
   (b) Clip the heart rate sensor around the chest and adjust the strap to fit snugly.
   (c) Check that the electrodes are firmly against the skin, and that the sensor is at the middle of the chest.
   (d) Check if the system is working.
6. GSR sensor placement
   (a) Clean and wash the hands.
   (b) Apply the electrodes.
   (c) Check that the electrodes are firmly against the skin.
   (d) Check if the system is working.
   (e) Check that the serial output from the sensor is between 200-512, otherwise adjust resistor with a screw driver until the serial output falls in such a range.
7. Collecting the baseline outside the virtual environment (1 min).
8. Collecting the baseline wearing the virtual reality headset (1 min).
9. Starting of the test (12 min about).
10. At the end of the test, remove the electrodes.
11. Filling out the post-test questionnaire.

5.4 Data Collecting

In order to start the acquisition and store the data, a VNC software has been used to navigate inside the Raspberry Pi. Once set a static IP address, the Raspberry autonomously connect to a server, and therefore it is possible to log in.

For the programming side, the Raspberry has been programmed in Python language. Signals which are received coming from the HR sensor and GSR sensor. HR sensor is critical because the bio-sensor sent high voltage when a heartbeat is detected. The digital value has not to be lost, therefore an interrupt routine has been used. Interrupt coming from the GPIO was implemented; the interrupt request was sent when a rising voltage value was read from the GPIO. The interrupt routine aims to store the time when the heartbeat was received and compute the beat-beat distance.

For the GSR side, a multithreading approach has been used; multithreading is the capability of a CPU to execute multiple threads or processes simultaneously, supported by the operating system.

The main program is a GUI implementation to facilitate the start of the acquisition, pause, stop and plot if requested. PySimpleGUI package was used to create the GUI [109].

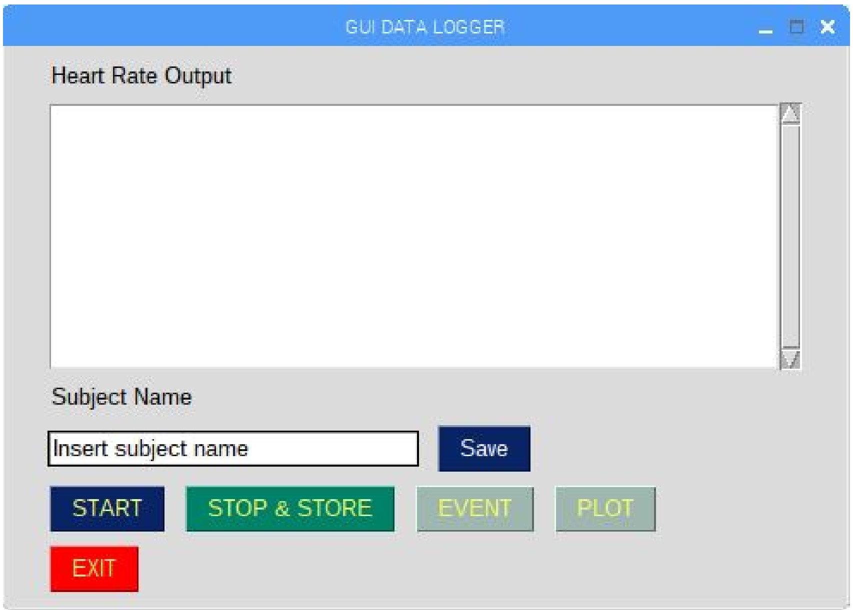


Figure 5.1: GUI data acquisistion.

5.5 Signal Processing

Post-Processing of the data has been performed with programming language Python 3.6.5 version, thanks to API provided by Numpy to process the signal

Algorithm 1 Data Acquisition

Ensure: HR & GSR write file
  → GUI design
  → Insert test name
  → Folder creation where store the data
  while EXIT is not pressed **do**
    if START is pressed **then**
      → Start GSR acquisition
      → Enable GPIO Interrupt for HR
      → Plot Instantaneous HR
    end if
    if STOP&STORE is pressed **then**
      → Pause GSR acquisition
      → Disable GPIO Interrupt for HR
      → Store GSR and HR
    end if
    if EVENT is pressed **then**
      → Save current time (to digitally store some particular information, as test event time)
    end if
    if PLOT is pressed **then**
      → Plot GSR signal
      → Plot HRV
    end if
  end while

as an array, Scipy providing the filters and Matplotlib package to visualize the data [110][110][111]. Computations were performed on an HP Pavilion, Intel Core i5-3230M CPU. As the literature suggests, the minimum sampling rate for the galvanic skin response signal has to be at least 200 Hz [87] to separate the two component of the GSR data. Therefore a sampling frequency of 256 Hz was used for the skin conductance signal. For the HR side, all heartbeat detected previously than 200 milliseconds (time order measure of the refractory heart period) of the prior heartbeat were rejected.

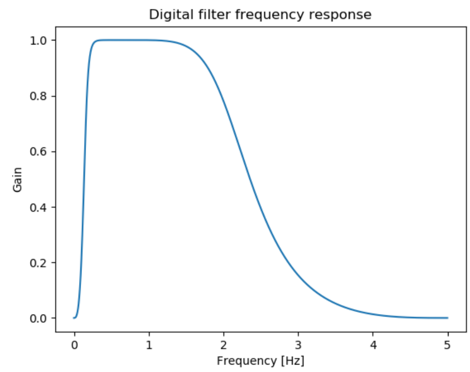
5.5.1 GSR Preprocessing

As already discussed in the chapter 3.2, the GSR signal is composed by the sum of two components: the skin conductance level (SCL) and the skin conductance response (SCR). The first indicates the tonic activity, reflecting the habituation to the situation. Instead, the second indicates the phasic activity, reflecting the response to arousal. To discriminate these two components, a filtering of the signal has been applied. The SCR data can be extracted by filtering the skin conductance signal, thus it has been filtered using a 3$^{\text{rd}}$ order Butterworth band-pass filter of 0.16 Hz to 2.1 Hz [112]. The digital filter is showed in figure 5.2a. Instead, the filtering outcome is showed in figure 5.2b, from this figure, both skin conductance level and the skin conductance response can be noted. Another approach to extract the SCR were also tried, like the proposed by Kim et al. [113], which provides to downsample the data to 10 Hz, then a data differentiation and subsequent convolution with a 20-point Bartlett window. However, the filtering technique previously described gives better results.

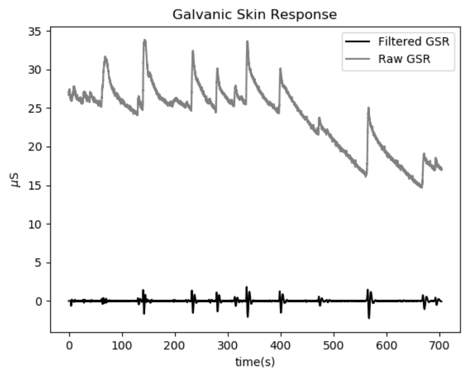
GSR data needs to be corrected to account for intrinsic inter-individual differences in skin conductance. Therefore, to correct the data two different approaches have been used, equation 5.1 [114] and 5.2 [115]. The first refers to the z-score standardisations leading to transform data with mean equal to zero and unit variance. The latter instead provide the data normalisation in a 0-1 range which is fundamental to allow the inter-subjects averaging. The GSR min-max scaled signal has been used for trend analysis and questionnaire correlation analysis.

$$\hat{GSR_{z-score}} = \frac{GSR - GSR_{mean}}{GSR_{std}} \tag{5.1}$$

$$\hat{GSR_{min-max}} = \frac{GSR - GSR_{min}}{GSR_{max} - GSR_{min}} \tag{5.2}$$

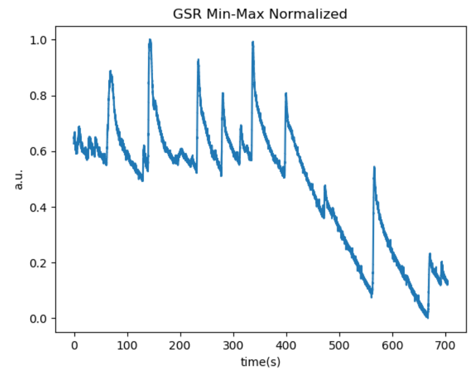


(a) Digital filter, the low cut and high cut frequencies are notable, moreover, in the range of interest 0.16-2.1 Hz the maximum gain is present.

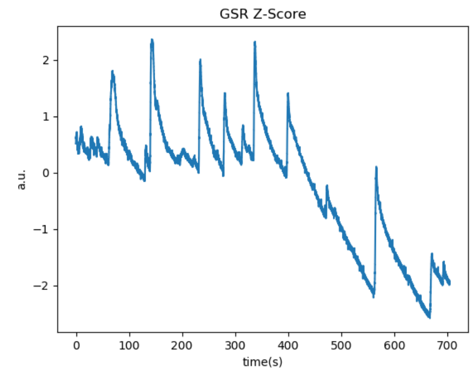


(b) Raw and filtered GSR data, in x-axis the time and the y-axis the μS for the raw GSR value, GSR data of 15thsubject.

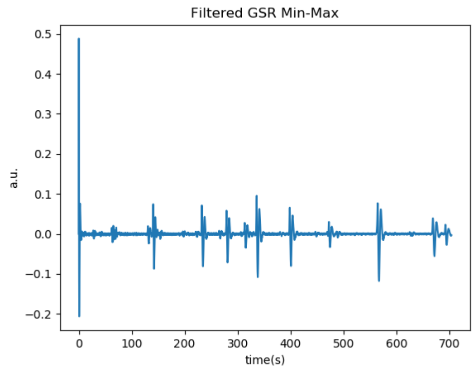
Figure 5.2: Digital filter and filtering result.



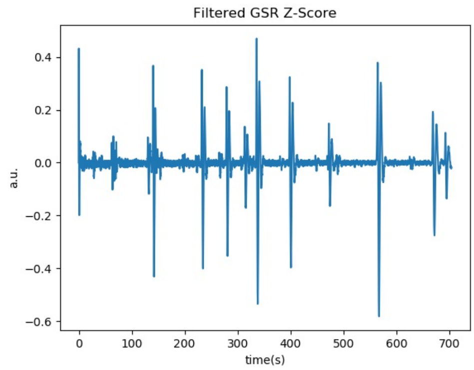
(a) Normalized GSR data.



(b) Standardised GSR data.



(c) Filtered normalized GSR data.



(d) Filtered standardised GSR data.

Figure 5.3: Normalization and standardisation with filtering outcomes. The initial filtering results of few seconds signal contains untrue values due to digital filtering limitations. Z-score standardisation faces better with such a limitation.

5.6 Feature Extraction

5.6.1 Heart Rate Features

From the HR sensor, the heart rate variability (HRV) has been derived. In Figure 5.4 the HRV plot is presented. As can be notable an outlier is present, it is not clear if such value is a real beat or due to a loss beat, this is a drawback of the HR sensor used. The mean cardiac frequency is also expressed as reported below in equation 5.3.

$$BPM = \frac{60}{\overline{HRV}} * 10^3 \tag{5.3}$$

Where BPM are the *beats per minute* and $\overline{HRV}$ is the mean of the HRV data in the time epoch.

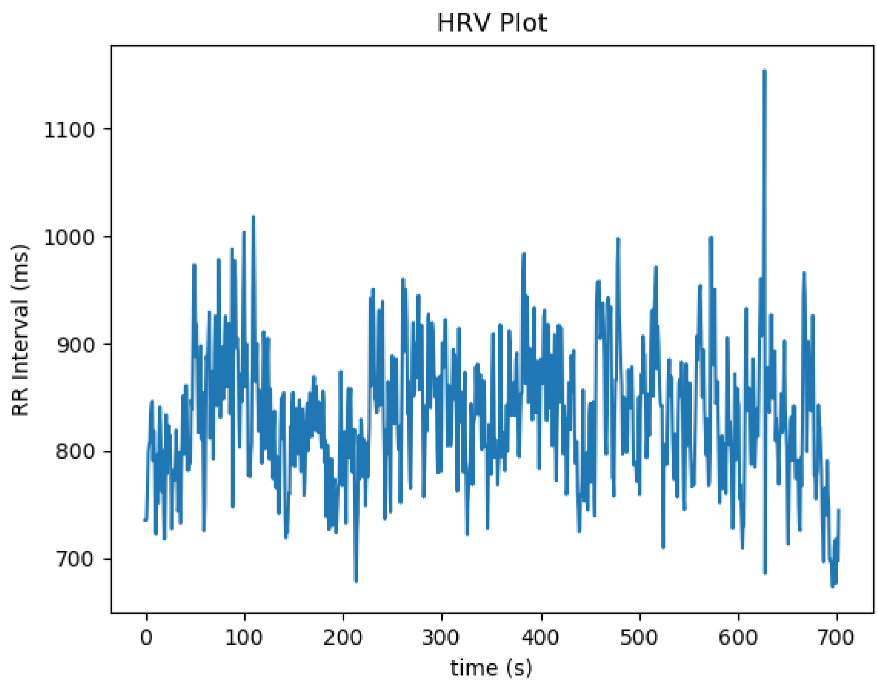


Figure 5.4: HRV plot of subject 17.

As reported in section 2.3.4, Johnson et al. 2011 [47] studies tries to validate the simulator using the HR signal. A similar approach has been used, particularly the one showed in Figure 2.8. To perform a similar analysis, epochs of 15 seconds prior and post an expected were extracted. From these epochs, the mean HR was estimated.

The previously described features are computed in the time-domain, but scientific literature reports important outcomes also in the frequency domain analysis. Thus, the spectrum analysis of the HRV signal has been computed.

Considering the unequally sampled nature of the HRV signal, and also considering that the interest would be to analyse the user-response to the test event, a frequency spectrum analysis of the 50 seconds epoch of the HRV data before and after a test event is performed. Frequency analysis of the HRV data in a time epoch of 50 seconds is investigated by Salahuddin et al. [79]. Moreover, the RR time series were resampled with a frequency of 2 Hz, then the power spectrum of the resampled time series was estimated with the Burg method of order 15 [63]. The parametric power spectral density estimation (PSD) Burg method has been chosen for its reliability for unequally spaced and short time epoch data. Parametric methods tries to build a linear prediction model which is able to whitening the input signal. More specifically, hypothesising that signal derives from an autoregressive (AR) model filter of a white noise, the result of a linear prediction model which withening the signal represent an approximation of the autoregressive model. Thus, the PSD can be computed by the transfer function of the coefficient of linear prediction model, as defined by equation 5.4. Outliers were removed interpolating the data.

$$P_{ARMA}(f) = T\rho_w \left|\frac{B(f)}{A(f)}\right|^2 \tag{5.4}$$

$$A(f) = 1 + \sum_{k=1}^{q} a(k)e^{-j2\pi fk}$$

$$B(f) = 1 + \sum_{k=1}^{p} b(k)e^{-j2\pi fkT}$$

Where: p and q are the order of the model, T is the time and ρ is the variance, while *a(k)* and *b(k)* are the coefficient of the filter, the Burg technique for computing the prediction error power and prediction error filter coefficients [116]. For the AR model *b(k)* coefficients are equal to 0. An example of spectrum analysis is provided by Figure 5.5. The ratio between the power in the LF frequency band (0.04-0.15 Hz) and power in the HF band (0.15-0.4 Hz) denotes the sympathetic system activation [78].

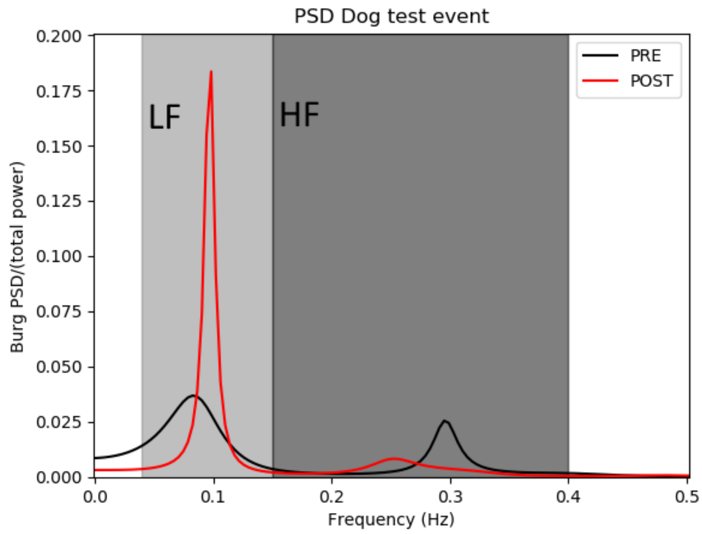


Figure 5.5: Spectrum analysis of the HRV data of 16^{th} tester during the 'dog' test event. Figure highlight an increase of the LF band frequency power and a decrease of the HF band frequency power meaning a high activation of the sympathetic activity. Values are normalised for the total power density.

5.6.2 GSR Features

Following the same approach used by Slater et al. [63], depicted in figure 2.16, from the filtered SCR data, 10s epochs are extracted prior and after a test event, considering that generally skin conductance response is present 1-5 seconds later a stimulus and has a time-span about 10 seconds. The samples are further divided by the first value of each extracted epoch, in order to observe changes rather the absolute value. The seven test epochs event are mediated for each subject, and for all the subjects, this allows to obtain one final mean shape curve, as defined by equation 5.5. Moreover, the same algorithm has been used 30 times, using seven random epochs extraction of the signal each time (which act as a control), normalised and averaged obtaining in this case 30 mean control curves. Such control curves serve as a proof that the SCR is not a random factor and is then evoked by the stimulus during the simulation.

Additionally, it has been decided to operate the same approach as previously described to extract the mean shape separately for each test event.

$$SCR_{shape}(i) = \frac{\sum_{k=1}^{N} \sum_{j=1}^{M} SCR_(k, j, i)}{N * M} \tag{5.5}$$

Other features extracted are:

$$\overline{GSR}_{mean}(k, j) = \frac{\sum_{i=0}^{L} G\hat{S}R(k, j, i)}{L} \tag{5.6}$$

$$GSR_{Acc}(k,j) = \sum_{i=0}^{L} G\hat{S}R(k,j,i) \tag{5.7}$$

$$Max(k,j) = \max(G\hat{S}R(k,j,i)) \tag{5.8}$$

$$PeaktoPeak(k,j) = \max((SCR)(k,j,i)) - \min((SCR)(k,j,i)) \tag{5.9}$$

Where:

- $G\hat{S}R$ is the raw and Z score standardized electrodermal activity, as defined by equation 5.1
- SCR is the filtered signal of the standardized GSR representing the skin conductance response
- $L = f_{sampling} * T_{time-interval} = 256Hz * 10s = 2560$
- $G\hat{S}R(k,j,i)$ is the standardized data value of subject k, at the event epoch j and the data point i, M is equal to the number of true or random event and N is equal to the number of subjects.

Above mentioned features are extracted from both the time interval preceding the test event and the following one, in order to obtain 'pre' and 'post' features. A delta difference between pre and post was also computed:

$$\Delta_{feature} = Post_{feature} - Pre_{feature} \tag{5.10}$$

These features are extracted considering the main characteristics of the galvanic skin response biosignal. Moreover, such features are also used for many researches. [87][117].

5.7 Statistical Analysis

On the following pages the statistical analysis procedure is described. A number of 30 subjects were recruited for the autonomous driving simulation experience. For half of the participants the Selective HUD was set, for the others the Omnicomprensive HUD. All the statistical analysis have been performed with software IBM SPSS Statistics 20.0 [118].

5.7.1 Main Analysis

The primary analysis aims to compare the mean values obtained from the subset of subjects who performed the test with the less informative user interface (i.e. selective HUD) and the other subset of subjects with the more informative user interface named all-embracing or omni-comprehensive HUD.

Procedure

During the test, the heart rate and skin conductance (GSR) signals were recorded for both user interfaces. Four features were extracted from the band-pass filtered and normalised GSR signal: the peak-to-peak value and the maximum value ten seconds before and after a test event, in total 7 test events are present. The difference between pre-post features (i.e. Delta) is computed both for peak-to-peak value and maximum value, lastly obtaining the two final features: Delta peak-to-peak value and Delta max value. In addition, other seven random fake events are considered which act as a 'control' group. Therefore to summarise, for each subject seven Delta values are obtained from the true test event and other seven Delta values are obtained from the fake test event. From Delta peak-to-peak value and delta max value a separate and joined statistical analysis is performed. The joined statistical analysis considers both the Delta peak-to-peak and Delta max features.

Design

For a single feature, a two-way mixed ANOVA with one within-subjects factor and one between-groups factor is applied. Delta value is the dependent variable, true and fake events are the within-subjects factor; the user interface (Selective or omni-comprehensive) is the between-subjects factor. Therefore, a 2 x 14 ANOVA with the user interface as an independent factor and true/-fake Delta values as a within-subjects factor was run, table 5.1 represent the dataset format. Bonferroni correction is applied.

Between-Subjects effects hypothesis of the model are:

H_0 The user interface has no significant effect on the EDA response

H_1 The user interface has a significant effect on the EDA response.

Obtaining a statistical significance lead to reject the null hypothesis, therefore considering the EDA response statistically significant in the two HUD.

Within-Subjects effects hypothesis of the model are:

H_0 Events have no significant effect on the EDA response

H_1 Events have a significant effect on the EDA response.

Obtaining a statistical significance lead to reject the null hypothesis, therefore considering EDA change significantly in the presence of an event.

5.7.2 Secondary Analysis

Further data analysis has been performed. These analyses are focused on the GSR trend, GSR data correlation with questionnaires and data analysis within the user interface between pre and post test event.

Table 5.1: Table which summirizes the dataset format required for the Mixed ANOVA [119].

2x14 MIXED ANOVA	**True1**	**True2**	**True3**	**True4**	**True5**	**True6**	**True7**	**Control1**	**Control2**	**Control3**	**Control4**	**Control5**	**Control6**	**Control7**
Selective HUD	*subj. R1,1 measure 1*	*subj. R1,1 measure 2*	*subj. R1,1 measure 3*	*subj. R1,1 measure 4*	*subj. R1,1 measure 5*	*subj. R1,1 measure 6*	*subj. R1,1 measure 7*	*subj. R1,1 measure 8*	*subj. R1,1 measure 9*	*subj. R1,1 measure 10*	*subj. R1,1 measure 11*	*subj. R1,1 measure 12*	*subj. R1,1 measure 13*	*subj. R1,1 measure 14*
	subj. R1,2 measure 1	*subj. R1,2 measure 2*	*subj. R1,2 measure 3*	*subj. R1,2 measure 4*	*subj. R1,2 measure 5*	*subj. R1,2 measure 6*	*subj. R1,2 measure 7*	*subj. R1,2 measure 8*	*subj. R1,2 measure 9*	*subj. R1,2 measure 10*	*subj. R1,2 measure 11*	*subj. R1,2 measure 12*	*subj. R1,2 measure 13*	*subj. R1,2 measure 14*
	. *etc* .	. *etc* .	. *etc* .	. *etc* .	. *etc* .	. *etc* .	. *etc* .	. *etc* .	. *etc* .	. *etc* .	. *etc* .	. *etc* .	. *etc* .	. *etc* .
	subj. R1,15 measure 1	*subj. R1,15 measure 2*	*subj. R1,15 measure 3*	*subj. R1,15 measure 4*	*subj. R1,15 measure 5*	*subj. R1,15 measure 6*	*subj. R1,15 measure 7*	*subj. R1,15 measure 8*	*subj. R1,15 measure 9*	*subj. R1,15 measure 10*	*subj. R1,15 measure 11*	*subj. R1,15 measure 12*	*subj. R1,15 measure 13*	*subj. R1,15 measure 14*
Omni-comprehensive HUD	*subj. R2,1 measure 1*	*subj. R2,1 measure 2*	*subj. R2,1 measure 3*	*subj. R2,1 measure 4*	*subj. R2,1 measure 5*	*subj. R2,1 measure 6*	*subj. R2,1 measure 7*	*subj. R2,1 measure 8*	*subj. R2,1 measure 9*	*subj. R2,1 measure 10*	*subj. R2,1 measure 11*	*subj. R2,1 measure 12*	*subj. R2,1 measure 13*	*subj. R2,1 measure 14*
	subj. R2,2 measure 1	*subj. R2,2 measure 2*	*subj. R2,2 measure 3*	*subj. R2,2 measure 4*	*subj. R2,2 measure 5*	*subj. R2,2 measure 6*	*subj. R2,2 measure 7*	*subj. R2,2 measure 8*	*subj. R2,2 measure 9*	*subj. R2,2 measure 10*	*subj. R2,2 measure 11*	*subj. R2,2 measure 12*	*subj. R2,2 measure 13*	*subj. R2,2 measure 14*
	. *etc* .	. *etc* .	. *etc* .	. *etc* .	. *etc* .	. *etc* .	. *etc* .	. *etc* .	. *etc* .	. *etc* .	. *etc* .	. *etc* .	. *etc* .	. *etc* .
	subj. R2,15 measure 1	*subj. R2,15 measure 2*	*subj. R2,15 measure 3*	*subj. R2,15 measure 4*	*subj. R2,15 measure 5*	*subj. R2,15 measure 6*	*subj. R2,15 measure 7*	*subj. R2,15 measure 8*	*subj. R2,15 measure 9*	*subj. R2,15 measure 10*	*subj. R2,15 measure 11*	*subj.R2,15 measure 12*	*subj. R2,15 measure 13*	*subj. R2,15 measure 14*

GSR trend

For both user interfaces, the GSR trend is examined. In order to accomplish this inspection, separately for both user interfaces, the mean and the standard deviation values of normalized skin conductance data are estimated.

Correlation with Questionnaires

The fourth section of the post-exposure test questionnaire concerns the subject perception of risk under a true test event condition. Subjects rated with 1-5 scale the level of perceived hazard of a driving situation experienced in the simulation and depicted in the questionnaire in order to avoid ambiguity. The main idea is to examine the correlation between such ratings and the GSR response in the following 10 seconds of a specific test event.

To evaluate the dependence between questionnaire and GSR, a multinomial regression model has been created for each sample of the user interface and each test event. The model considers as dependent variable a categorical value which has to be predicted (the questionnaire response for the particular test event) and as predictor variables the continuous measurement (GSR response, feature Peak to Peak) throughout all the test event.

In order to avoid the multicollinearity problem, which affirms that linear regression models present issues when the predictor variables are highly correlated, pearson correlation analysis between these variables has been performed in order to evaluate the correlation. Such an analysis found that SCR $\Delta_{peak-to-peak}$ feature, extracted from the z-score standardised GSR, presents a relatively small correlation.

The hypothesis of the model are:

H_0 The predictors variables (GSR Response) have no significant effect on the model fitting.

H_1 The predictor variables (GSR Response) have significant effect on the model fitting.

Obtaining a statistical significance lead to reject the null hypothesis, therefore considering the GSR response statistically significant to predict the questionnaire response for such test event and user interface.

Within User Interface Analysis

In this section is performed the statistical analysis inside the user interface set. Therefore two separate analyses were performed for each HUD used. In order to check the statistical significance, for each feature previously described (i.e. peak to peak and maximum value), the pre and post mean values were compared. Therefore, an Independent T-Test within the user interface and between pre and post test event was performed. The hypothesis of the model are:

H_0 The event has no significant effect on pre and post EDA.

H_1 The event has a significant effect on pre and post EDA.

Questionnaire Analysis

As above mentioned, a questionnaire was provided to users to evaluate subjective responses. Two user interfaces have been used; therefore subjective responses between the two HUD have to be compared. In order to compare categorical data such data, an alternative to the independent t-test has been used; the Mann-Whitney U Test which is adopted when the data is ordinal and not normally distributed, and it uses non-parametric statistics. The Mann-Whitney U Test is employed to investigate whether any differences observed by the researcher are there just due to chance or whether two data samples are significantly different from one another.

Chapter 6

Study2 - Results

6.1 Questionnaire Outcomes

The final dataset was composed of 19 individuals, with an average age of 22.53±2.97 years, who experienced the selective HUD, and other 19 individuals with an average age of 25.27 ± 6.40 years experienced the omni-comprehensive HUD. Such individuals did not experience motion sickness, Figure 6.1.

HUD	Gender		Age
SEL	Male	Female	22.53 ± 2.97
	12	7	
OMN	Male	Female	25.27 ± 6.40
	13	6	

Table 6.1: Table reporting the general characteristics of the dataset. Mean age and standard deviation reported.

Figure 6.1: On the left the selective HUD, on the right omni-comprehensive HUD. Pre and post motion sickness questionnaire for each user interface. Scale ranges from 1 to 4.

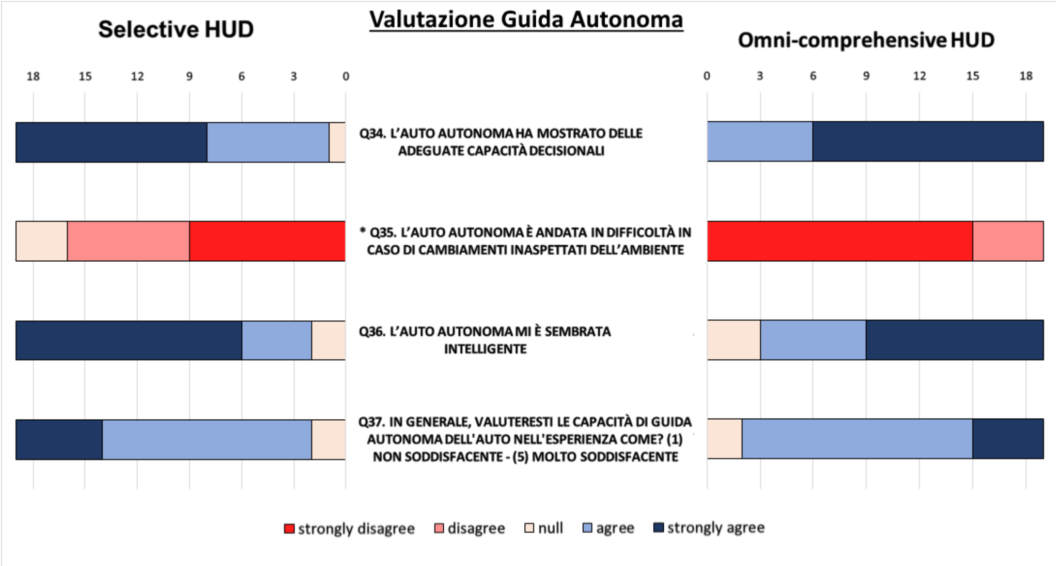


Figure 6.2: On the left the selective HUD, on the right omni-comprehensive HUD. Assessing the autonomous driving simulation. Scale starting from 1 to 5, (1) I do not agree with the statement - (5) I fully agree with the statement. Mann-Whitney Test, $^{*}p-value < 0.05$, $^{**}p-value < 0.01$, $^{***}p-value < 0.001$.

Autonomous Driving Assessment

HUD		Mean	Std. Dev.	Std. Error Mean	Mann-Withney Test p-value (two-tailed)
Q34	SEL	4.5263	.61178	.14035	p=.278
	OMN	4.6842	.47757	.10956	
Q35	SEL	1.6842	.74927	.17189	p=.041*
	OMN	1.2105	.41885	.09609	
Q36	SEL	4.5789	.69248	.15887	p=.415
	OMN	4.3684	.76089	.17456	
Q37	SEL	4.1579	.60214	.13814	p=.863
	OMN	4.1053	.56713	.13011	

* statistical significance, p<.05

Table 6.2: Descriptive statistics and Mann-Whitney U Test.

Interesting results notable from the Figure 6.2 and Table 6.2, which reports the results of the fourth section of the questionnaire, is that HUD reports similar results, but a significant difference is obtained on the evaluation if the autonomous vehicle faced some difficulties when the external environment unexpectedly changed. More specifically, selective subjects reported that the car had more difficulties to respond to such situations than omni-comprehensive HUD. The autonomous driving performances between selective and omni-comprehensive HUD are the same; therefore such differences can only be due to the HUD.

Two-way RM ANOVA Questionnaire Danger Assessment	F statistic	p-value	
Within-Subjects Effects (Event)	F(6,216)=54.05	p=.000*	Q38-Q43-Q48
Within-Subjects Effects (Event*HUD)	F(6,216)=2.05	p=.060	Q53-Q58-Q63
Between-Subjects Effects (HUD)	F(1,36)=15.91	p=.000*	Q68

Table 6.3: Two-way ANOVA with repeated measures run on Q38-Q43-Q48 Q53-Q58-Q63-Q68 questions which are related to the evaluation of the risk perception. Bonferroni correction applied. Significant results are marked.

Some statements present in the questionnaire are correlated each other, in particular the test event section, therefore to take into consideration their relationship, a two-way ANOVA with repeated measures was conducted to compare the main effects of HUD and the interaction effect between HUD on the questions regarding the perceived danger/fear under test events. HUD included two levels (i.e. selective and omni-comprehensive user interface) and each question included five levels (not dangerous at all, slightly dangerous, medium dangerous, discretely dangerous, completely dangerous), seven repeated measures of them was evaluated. Levene's Test of Equality of Error Variances was conducted to check the ANOVA assumptions which were satisfied, exception represented by Q53. The HUD had a significant main effect (p=.000) meaning a significant difference between selective (M=2.61, SD=.09) and omni-comprehensive user interface (M=3.13, SD=.09). Ratings to questions also had a significant main effect (p=.000), meaning significant difference between Q38 (M=3.21, SD=.87), Q43 (M=3.02, SD=.99), Q48 (M=3.00, SD=1.21), Q53 (M=1.63, SD=1.02), Q58 (M=3.82, SD=.98), Q63 (M=1.29, SD=.57) and Q68 (M=4.10, SD=.09). The interaction between HUD and questionnaire was not significant (p=.060). Results are reported in Table 6.3.

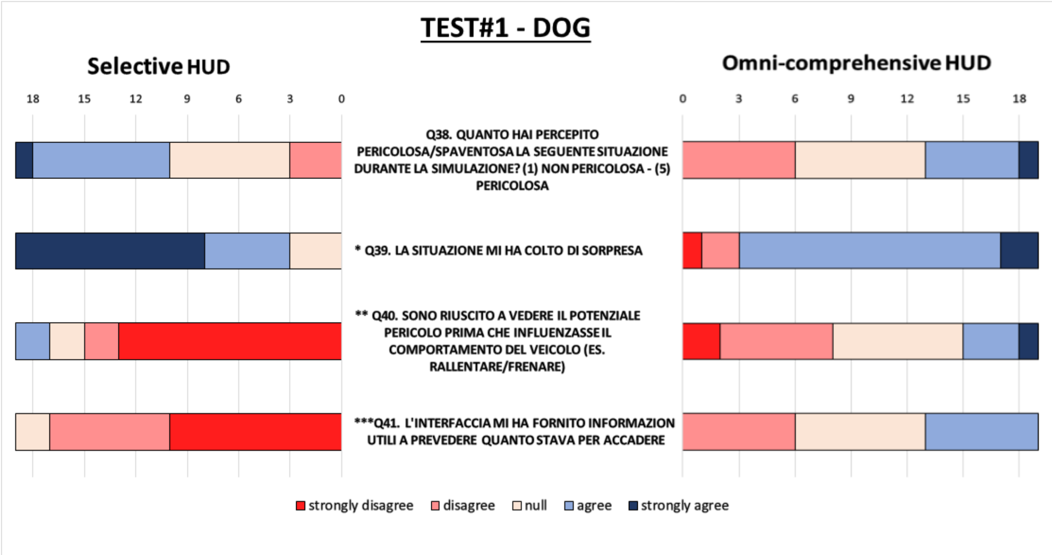


Figure 6.3: On the left the selective HUD, on the right omni-comprehensive HUD. Assessing the experienced dog test event. Scale starting from 1 to 5, (1) I do not agree with the statement - (5) I fully agree with the statement. Mann-Whitney Test, $*p - value < 0.05$, $**p - value < 0.01$, $***p - value < 0.001$.

Test#1 - Dog Event Assessment					
Question	HUD	Mean	Std. Dev.	Std. Error Mean	Mann-Withney Test p-value (two-tailed)
Q38	SEL	3.3684	.83070	.19058	p=.134
	OMN	3.0526	.91127	.20906	
Q39	SEL	4.4211	.76853	.17631	p=.016*
	OMN	3.7368	.99119	.22739	
Q40	SEL	1.6316	1.06513	.24436	p=.001*
	OMN	2.7368	1.04574	.23991	
Q41	SEL	1.5789	.69248	.15887	p=.000*
	OMN	3.0000	.81650	.18732	

* statistical significance, p<.05

Table 6.4: Descriptive statistics and Mann-Whitney U Test.

From Figure 6.3 and Table 6.4 generally selective sample reported that the *dog* event has been perceived more dangerous or frightful in mean value. Significant differences are reported in question related to how much the situation was surprising, to evaluate if the subjects noticed the potential danger previously to influence the driving actions and lastly to assess if the HUD provides useful information to put on alert regarding that particular element.

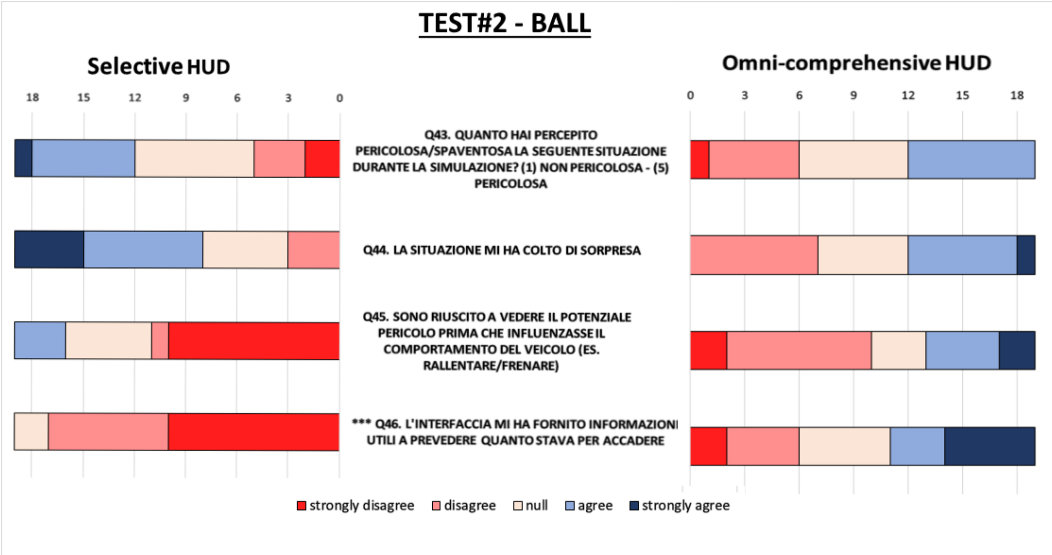


Figure 6.4: On the left the selective HUD, on the right omni-comprehensive HUD. Assessing the experienced ball test event. Scale starting from 1 to 5, (1) I do not agree with the statement - (5) I fully agree with the statement. Mann-Whitney Test, $^{*}p-value < 0.05$, $^{**}p-value < 0.01$, $^{***}p-value < 0.001$.

Test#2- Ball Event Assessment					
Question	HUD	Mean	Std. Dev.	Std. Error Mean	Mann-Withney Test p-value (two-tailed)
Q43	SEL	3.0526	1.07877	.24749	p=.872
	OMN	3.0000	.94281	.21630	
Q44	SEL	3.6316	1.01163	.23208	p=.102
	OMN	3.0526	.97032	.22261	
Q45	SEL	2.0526	1.22355	.28070	p=.064
	OMN	2.7895	1.22832	.28180	
Q46	SEL	1.5789	.69248	.15887	p=.000*
	OMN	3.2632	1.36797	.31383	

* statistical significance, p<.05

Table 6.5: Descriptive statistics and Mann-Whitney U Test.

Continuing, Figure 6.4 and Table 6.5 reports similar outcomes on the evaluation of perceived dangerous or fear in *ball* event. No significant differences are reported in question related to how much the situation was surprising, to evaluate if the subjects noticed the potential danger previously to influence the driving actions, however their one-tailed p-value is smaller than .05. Lastly, significant differences are present in the question which evaluates if the HUD provides useful information to put on alert regarding that particular element.

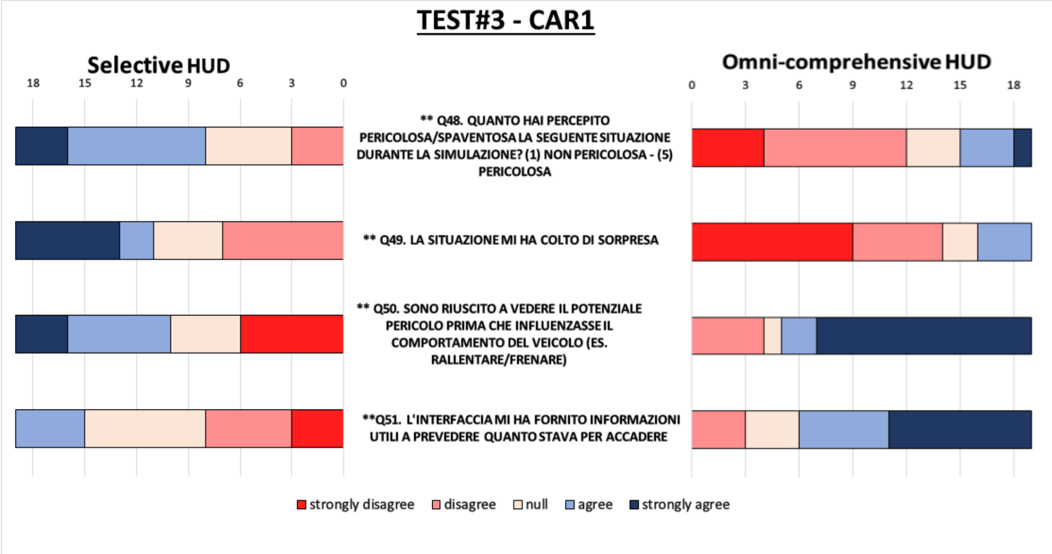


Figure 6.5: On the left the selective HUD, on the right omni-comprehensive HUD. Assessing the experienced car1 test event. Scale starting from 1 to 5, (1) I do not agree with the statement - (5) I fully agree with the statement. Mann-Whitney Test, $^{*}p-value < 0.05$, $^{**}p-value < 0.01$, $^{***}p-value < 0.001$.

Test#3 - Car1 Event Assessment					
Question	HUD	Mean	Std. Dev.	Std. Error Mean	Mann-Withney Test p-value (two-tailed)
Q48	SEL	3.5789	.96124	.22052	p=.003*
	OMN	2.4211	1.16980	.26837	
Q49	SEL	3.3684	1.30002	.29825	p=.001*
	OMN	1.9474	1.12909	.25903	
Q50	SEL	3.0000	1.52753	.35044	p=.007*
	OMN	4.1579	1.25889	.28881	
Q51	SEL	2.6316	1.01163	.23208	p=.001*
	OMN	3.9474	1.12909	.25903	

* statistical significance, p<.05

Table 6.6: Descriptive statistics and Mann-Whitney U Test.

From Figure 6.5 and Table 6.6 selective sample reported that significantly the *car1* event has been perceived more dangerous or frightful than omni-comprehensive sample. Additionally, significant differences are also obtained in question related to how much the situation was surprising, to evaluate if the subjects noticed the potential danger previously to influence the driving actions and lastly to assess if the HUD provides useful information to put on alert regarding that particular element.

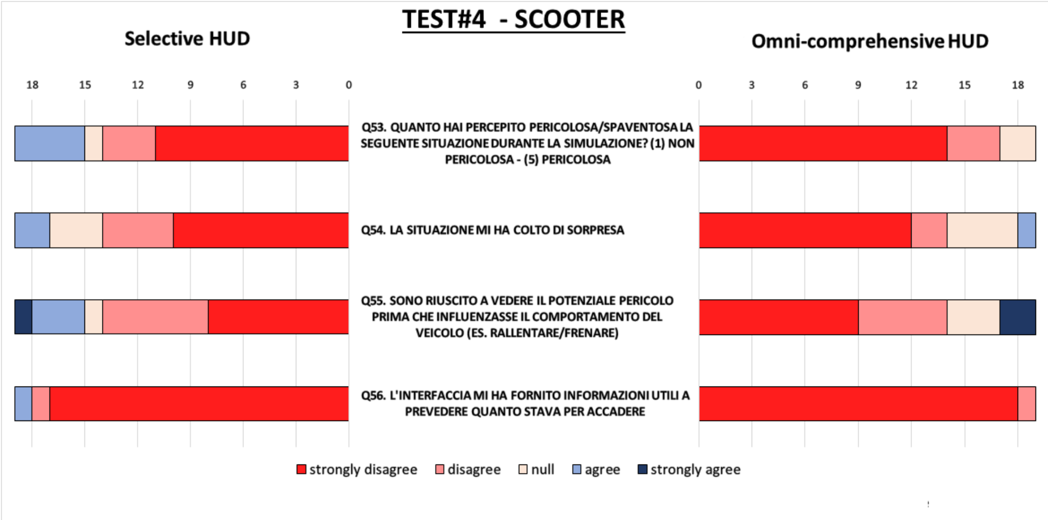


Figure 6.6: On the left the selective HUD, on the right omni-comprehensive HUD. Assessing the experienced scooter test event. Scale starting from 1 to 5, (1) I do not agree with the statement - (5) I fully agree with the statement. Mann-Whitney Test, $^{*}p-value < 0.05$, $^{**}p-value < 0.01$, $^{***}p-value < 0.001$.

Test#4 - Scooter Event Assessment					
Question	HUD	Mean	Std. Dev.	Std. Error Mean	Mann-Withney p-value (two-tailed)
Q53	SEL	1.8947	1.24252	.28505	p=.208
	OMN	1.3684	.68399	.15692	
Q54	SEL	1.8421	1.06787	.24499	p=.628
	OMN	1.6842	1.00292	.23009	
Q55	SEL	2.1053	1.28646	.29513	p=.773
	OMN	2.0000	1.29099	.29617	
Q56	SEL	1.2105	.71328	.16364	p=.743
	OMN	1.0526	.22942	.05263	

* statistical significance, p<.05

Table 6.7: Descriptive statistics and Mann-Whitney U Test.

From results reported in Figure 6.6 and Table 6.7, the *scooter* event has not been perceived dangerous or frightful by users. No main differences are obtained.

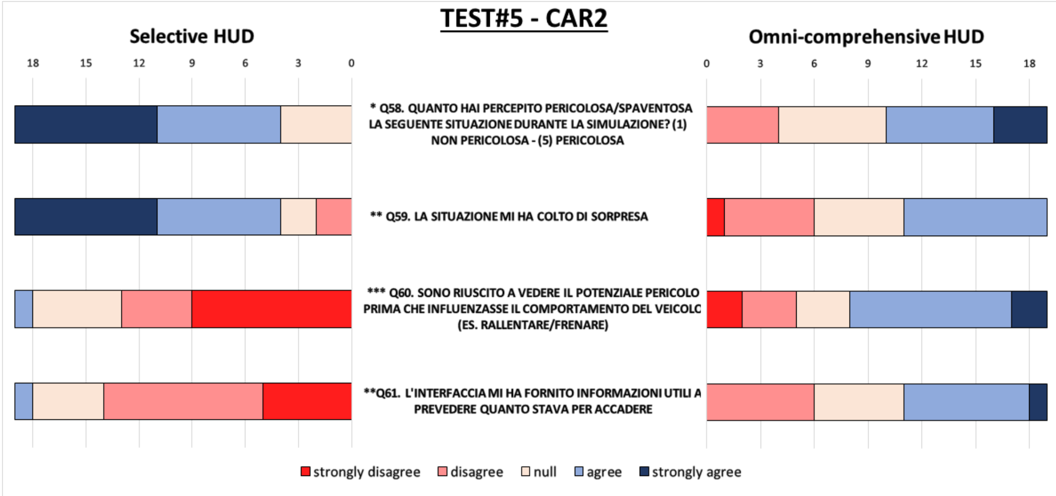


Figure 6.7: On the left the selective HUD, on the right omni-comprehensive HUD. Assessing the experienced car2 test event. Scale starting from 1 to 5, (1) I do not agree with the statement - (5) I fully agree with the statement. Mann-Whitney Test, $^{*}p - value < 0.05$, $^{**}p - value < 0.01$, $^{***}p - value < 0.001$.

Test#5 - Car2 Event Assessment					
Question	HUD	Mean	Std. Dev.	Std. Error Mean	Mann-Withney Test p-value (two-tailed)
Q58	SEL	4.2105	.78733	.18063	p=.017*
	OMN	3.4211	1.01739	.23341	
Q59	SEL	4.1053	.99413	.22807	p=.002*
	OMN	3.0526	.97032	.22261	
Q60	SEL	1.8947	.99413	.22807	p=.000*
	OMN	3.3158	1.20428	.27628	
Q61	SEL	2.0526	.84811	.19457	p=.001*
	OMN	3.1579	.95819	.21982	

* statistical significance, p<.05

Table 6.8: Descriptive statistics and Mann-Whitney U Test.

From Figure 6.7 and Table 6.8 selective sample reported that significantly the *car2* event has been perceived more dangerous or frightful than omni-comprehensive sample. In addition, significant differences are even obtained in question related to evaluate the situation was surprising and to what extent, to evaluate if the subjects noticed the potential danger previously to influence the driving actions and lastly to assess if the HUD provides useful information to put on alert regarding that particular element.

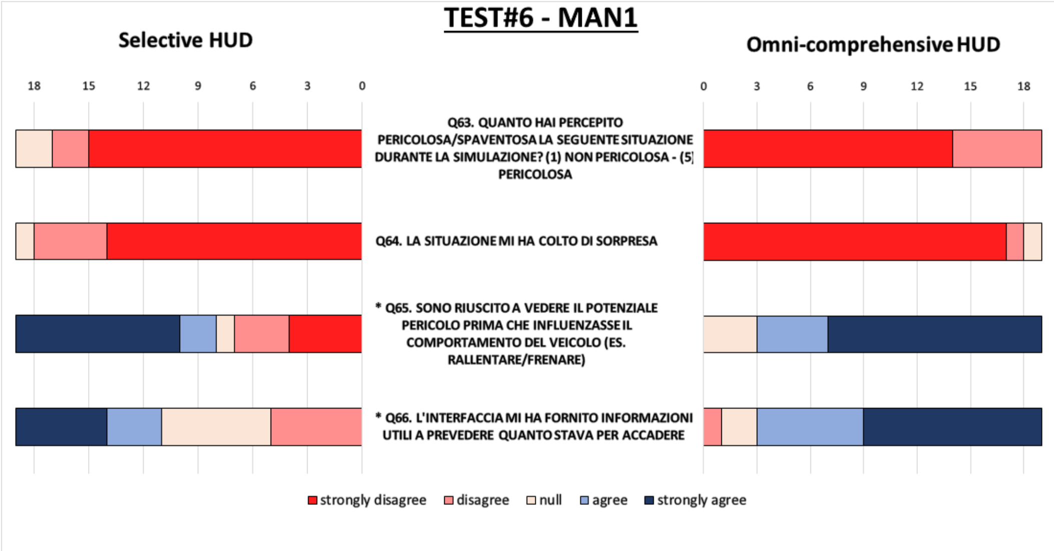


Figure 6.8: On the left the selective HUD, on the right omni-comprehensive HUD. Assessing the experienced man1 test event. Scale starting from 1 to 5, (1) I do not agree with the statement - (5) I fully agree with the statement. Mann-Whitney Test, $^{*}p-value < 0.05$, $^{**}p-value < 0.01$, $^{***}p-value < 0.001$.

Test#6 - Man1 Event Assessment					
Question	HUD	Mean	Std. Dev.	Std. Error Mean	Mann-Withney p-value (two-tailed)
Q63	SEL	1.3158	.67104	.15395	p=1.000
	OMN	1.2632	.45241	.10379	
Q64	SEL	1.3158	.58239	.13361	p=.390
	OMN	1.1579	.50146	.11504	
Q65	SEL	3.4737	1.71167	.39268	p=.093
	OMN	4.4737	.77233	.17718	
Q66	SEL	3.4211	1.16980	.26837	p=.018*
	OMN	4.3158	.88523	.20308	

* statistical significance, p<.05

Table 6.9: Descriptive statistics and Mann-Whitney U Test.

From results reported in Figure 6.8 and Table 6.9, the *scooter* event has not been perceived dangerous or frightful in mean value. No main differences are obtained. Significant difference is obtained in the question which assesses the usefulness of HUD information regarding the potential danger.

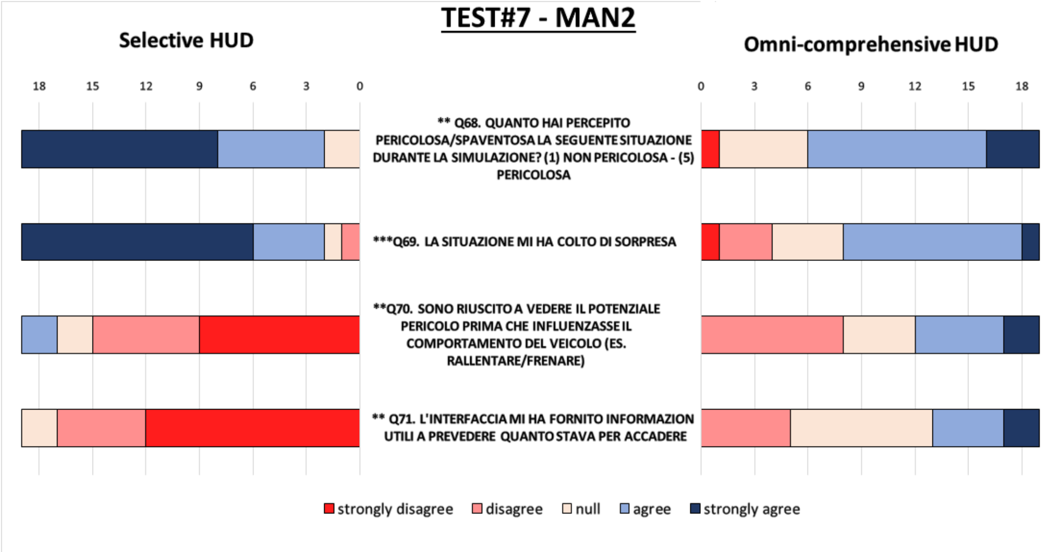


Figure 6.9: On the left the selective HUD, on the right omni-comprehensive HUD. Assessing the experienced man2 test event. Scale starting from 1 to 5, (1) I do not agree with the statement - (5) I fully agree with the statement. Mann-Whitney Test, $^{*}p - value < 0.05$, $^{**}p - value < 0.01$, $^{***}p - value < 0.001$.

Test#7- Man2 Event Assessment					
Question	HUD	Mean	Std. Dev.	Std. Error Mean	Mann-Withney Test p-value (two-tailed)
Q68	SEL	4.4737	.69669	.15983	p=.008*
	OMN	3.7368	.93346	.21415	
Q69	SEL	4.5263	.84119	.19298	p=.000*
	OMN	3.3684	1.01163	.23208	
Q70	SEL	1.8421	1.01451	.23275	p=.001*
	OMN	3.0526	1.07877	.24749	
Q71	SEL	1.4737	.69669	.15983	p=.000*
	OMN	3.1579	.95819	.21982	

* statistical significance, p<.05

Table 6.10: Descriptive statistics and Mann-Whitney U Test.

The last test event is *man2*, Figure 6.7 and Table 6.8 reports the outcomes. The selective sample reported that such a test event has been perceived significantly more dangerous or frightful than the omni-comprehensive sample. Moreover, significant differences are obtained in question related to the evaluation if the situation was surprising and to what extent, to evaluate if the subjects noticed the potential danger previously to influence the driving behaviour and lastly to assess if the HUD provides useful information to put on alert regarding that particular element.

Furthermore, proceeding with the questionnaire results, the following section evaluates the situational awareness. More specifically this section assesses if specific information displayed by HUD helped the user to understand that the autonomous system takes over virtual objects and how they have been managed. In particular, the first two questions concern the utility of bounding box and label displayed for the traffic lights; continuing with road signs, potential dangers, other vehicles and navigation lines.

Between HUD, the section regarding situational awareness do not present significant differences with one exception; HUD has a main effect on the evaluation of the usefulness of navigation line of the traffic vehicles(p=.046). The omni-comprehensive sample reported smaller satisfaction level on such a question. Results are reported in Figure 6.10 and Table 6.11.

Concerning the quantity of information, the following results section reports outcomes related to the mental workload, Figure 6.11 and in Table 6.12. In this section it would expect significant differences between user interfaces, considering that the amount of information consistently changes between the HUD. In this case the scale ranges from 1 to 5, 1 indicates that the amount of information is extremely insufficient; instead, 5 indicates that information is extremely excessive; therefore the adequate amount of information is 3. As expected, results show no significant differences in traffic lights assessments because the information displayed for traffic lights do not change between HUD. Many questions report significant differences. In particular, the selective sample expressed slightly insufficient information displayed for bounding box and labels of road signs, potential danger, traffic cars and acoustic alert for road signs. On the other hand, omni-comprehensive sample reported slightly redundant information for bounding box and label of traffic cars and the traffic cars' navigation lines.

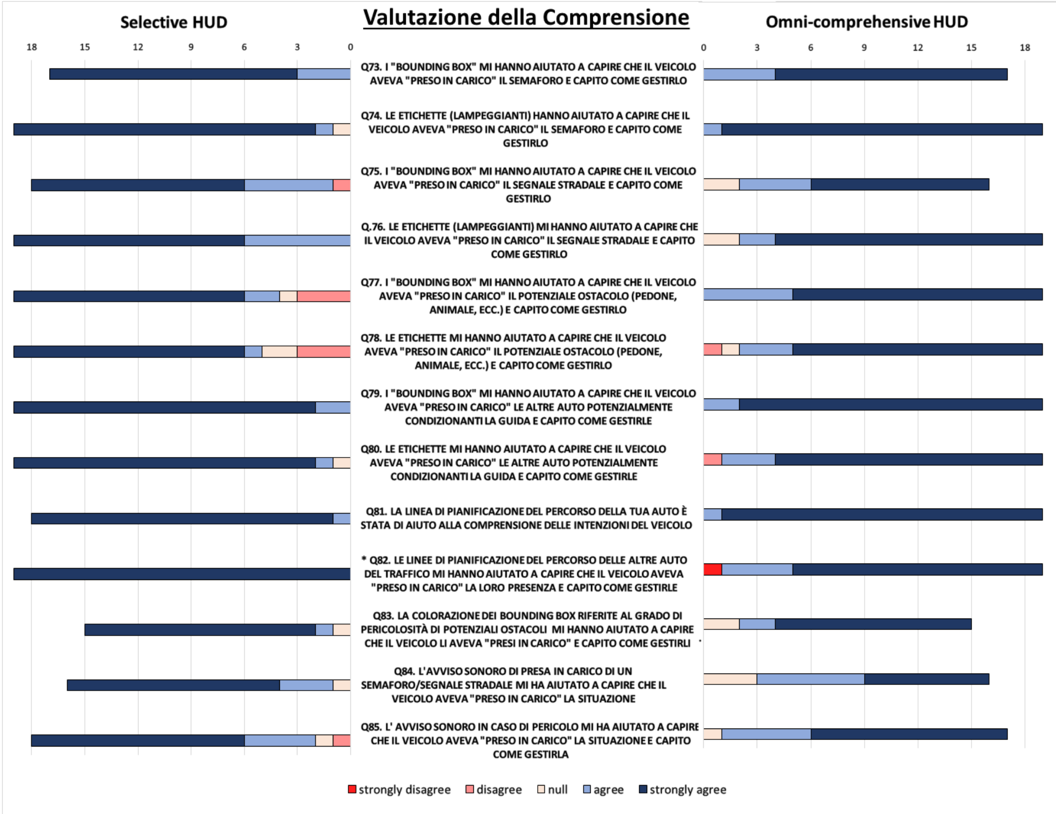


Figure 6.10: On the left the selective HUD, on the right omni-comprehensive HUD. Questions related to the situational awareness. Scale starting from 1 to 5, (1) I do not agree with the statement - (5) I fully agree with the statement. Mann-Whitney Test, $^{*}p - value < 0.05$. Subjects who could not answer to the question (i.e. he or she did not noticed the required information) were ignored.

Situational Awareness Assessment					
Question	HUD	Mean	Std. Dev.	Std. Error Mean	Mann-Withney p-value (two-tailed)
Q73	SEL	4.8235	.39295	.09531	p=1.000
	OMN	4.7647	.43724	.10605	
Q74	SEL	4.8421	.50146	.11504	p=.743
	OMN	4.9474	.22942	.05263	
Q75	SEL	4.5556	.78382	.18475	p=.772
	OMN	4.5000	.73030	.18257	
Q76	SEL	4.6842	.47757	.10956	p=.714
	OMN	4.6842	.67104	.15395	
Q77	SEL	4.3158	1.15723	.26549	p=.492
	OMN	4.7368	.45241	.10379	
Q78	SEL	4.2632	1.19453	.27404	p=.558
	OMN	4.5789	.83771	.19218	
Q79	SEL	4.8947	.31530	.07234	p=1.000
	OMN	4.8947	.31530	.07234	
Q80	SEL	4.8421	.50146	.11504	p=.500
	OMN	4.6842	.74927	.17189	
Q81	SEL	4.9444	.23570	.05556	p=1.000
	OMN	4.9474	.22942	.05263	
Q82	SEL	5.0000	0.00000	0.00000	p=.046*
	OMN	4.5789	.96124	.22052	
Q83	SEL	4.8000	.56061	.14475	p=.555
	OMN	4.6000	.73679	.19024	
Q84	SEL	4.6875	.60208	.15052	p=.106
	OMN	4.2500	.77460	.19365	
Q85	SEL	4.5000	.85749	.20211	p=1.000
	OMN	4.5882	.61835	.14997	

* statistical significance, p<.05

Table 6.11: Descriptive statistics and Mann-Whitney U Test.

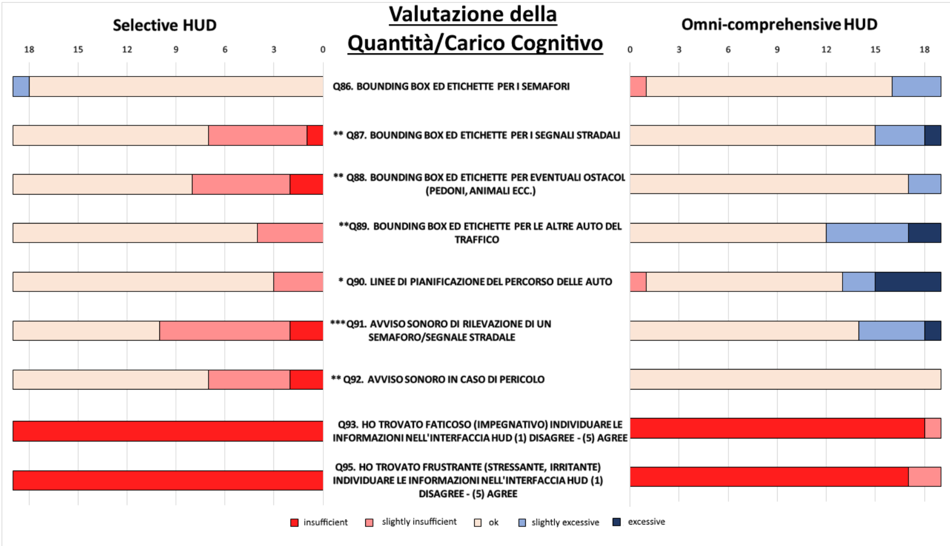


Figure 6.11: On the left the selective HUD, on the right omni-comprehensive HUD. Questions related to the evaluation of the amount of information and the mental workload. Scale starting from 1 to 5, (1) insufficient info - (5) excessive info. Mann-Whitney Test, $*p-value < 0.05$, $**p-value < 0.01$, $***p-value < 0.001$.

Quantity/Mental Workload Assessment					
Question	HUD	Mean	Std. Dev.	Std. Error Mean	Mann-Withney p-value (two-tailed)
Q86	SEL	3.0526	.22942	.05263	p=.604
	OMN	3.1053	.45883	.10526	
Q87	SEL	2.5789	.60698	.13925	p=.001*
	OMN	3.2632	.56195	.12892	
Q88	SEL	2.4737	.69669	.15983	p=.001*
	OMN	3.1053	.31530	.07234	
Q89	SEL	2.7895	.41885	.09609	p=.001*
	OMN	3.4737	.69669	.15983	
Q90	SEL	2.8421	.37463	.08595	p=.014*
	OMN	3.4737	.90483	.20758	
Q91	SEL	2.3684	.68399	.15692	p=.000*
	OMN	3.3158	.58239	.13361	
Q92	SEL	2.5263	.69669	.15983	p=.008*
	OMN	3.0000	0.00000	0.00000	
Q93	SEL	1.0000	0.00000	0.00000	p=1.000
	OMN	1.0526	.22942	.05263	
Q95	SEL	1.0000	0.00000	0.00000	p=.486
	OMN	1.1053	.31530	.07234	

* statistical significance, p<.05

Table 6.12: Descriptive statistics and Mann-Whitney U Test.

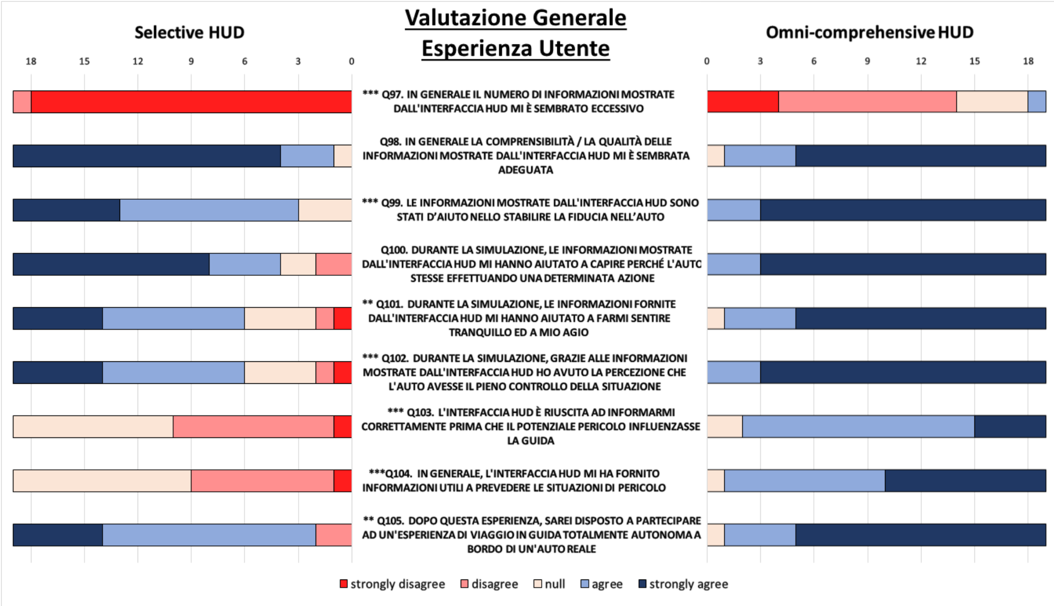


Figure 6.12: On the left the selective HUD, on the right omni-comprehensive HUD. Questions related to the evaluation of the whole user-experience. Scale starting from 1 to 5, (1) I do not agree with the statement - (5) I fully agree with the statement. Mann-Whitney Test, $*p-value < 0.05$, $**p-value < 0.01$, $***p-value < 0.001$.

User-Experience Assessment					
Question	HUD	Mean	Std. Dev.	Std. Error Mean	Mann-Withney p-value (two-tailed)
Q97	SEL	1.0526	.22942	.05263	p=.000*
	OMN	2.1053	.80930	.18567	
Q98	SEL	4.7368	.56195	.12892	p=.908
	OMN	4.6842	.58239	.13361	
Q99	SEL	4.1579	.68825	.15789	p=.001*
	OMN	4.8421	.37463	.08595	
Q100	SEL	4.2632	1.04574	.23991	p=.055
	OMN	4.8421	.37463	.08595	
Q101	SEL	3.7895	1.08418	.24873	p=.003*
	OMN	4.6842	.58239	.13361	
Q102	SEL	3.7895	1.08418	.24873	p=.000*
	OMN	4.8421	.37463	.08595	
Q103	SEL	2.4211	.60698	.13925	p=.000*
	OMN	4.1053	.56713	.13011	
Q104	SEL	2.4737	.61178	.14035	p=.000*
	OMN	4.4211	.60698	.13925	
Q105	SEL	4.0526	.84811	.19457	p=.005*
	OMN	4.6842	.58239	.13361	

* statistical significance, p<.05

Table 6.13: Descriptive statistics and Mann-Whitney U Test.

Other meaningful results are reported in 6.12 and Table 6.13. Such questions try to evaluate the whole user-experience including overall considerations of the previous sections. The first interesting result is represented by the overall evaluation of the quantity of information displayed by HUD; significant difference is obtained (p=.000). Such results tell that omni-comprehensive HUD displays slightly redundant information compared to the selective user interface. The capability of HMI to build trust on the autonomous system is significantly different between HUD (p=.001), while higher mean value for the omni-comprehensive sample (p=.055) are obtained for *why* information concerning the evaluation if HUD helps to understand why the autonomous vehicle performed a general or specific driving action. Information provided by the user interface in order to calm the subjects is significantly different between HUD (p=.003), better results are obtained for omni-comprehensive HUD. The selective sample also reported that the autonomous car had worse control of the situation compared to the other HUD (p=.000). The same behaviour is observed in question which evaluates the capability of HUD to alert the driver previously about the potential danger (p=000) and if the HUD information provided useful information to foresee the danger (p=.000). Finally, the last question of this section evaluates if the user would experience a real driving experience in a fully automated vehicle, Figure 6.13. A significant difference between pre and post in omni-comprehensive user interface (p=.002) and between HUD on the post section (p=.005).

Results reported in Figure 6.14 and Table 6.14 reports the answers of the section which aims to evaluate the user immersion and presence inside the virtual driving simulator. User interface did not have a main effect on this section; similar outcomes are obtained.

Furthermore, HUD did not have a main effect on the evaluation of the driving simulation fidelity to the real driving section. Outcomes are reported in Figure 6.15 and Table 6.15.

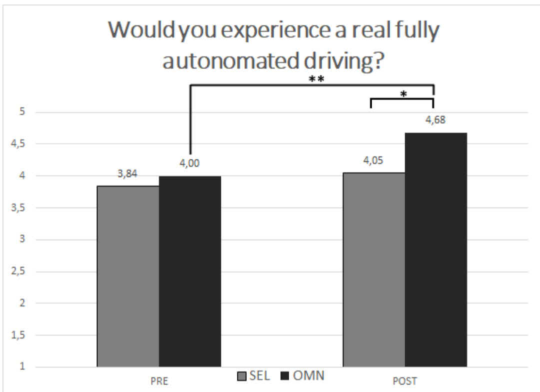


Figure 6.13: On the left the prior answer for selective HUD (gray) and omni-comprehensive HUD (black), on the right the post answer. Scale starting from 1 to 5, (1) Absolutely negative - (5) Absolutely positive. Mann-Whitney Test between HUD and related t-test between pre-post, $^{*}p-value<0.05$, $^{**}p-value<0.01$.

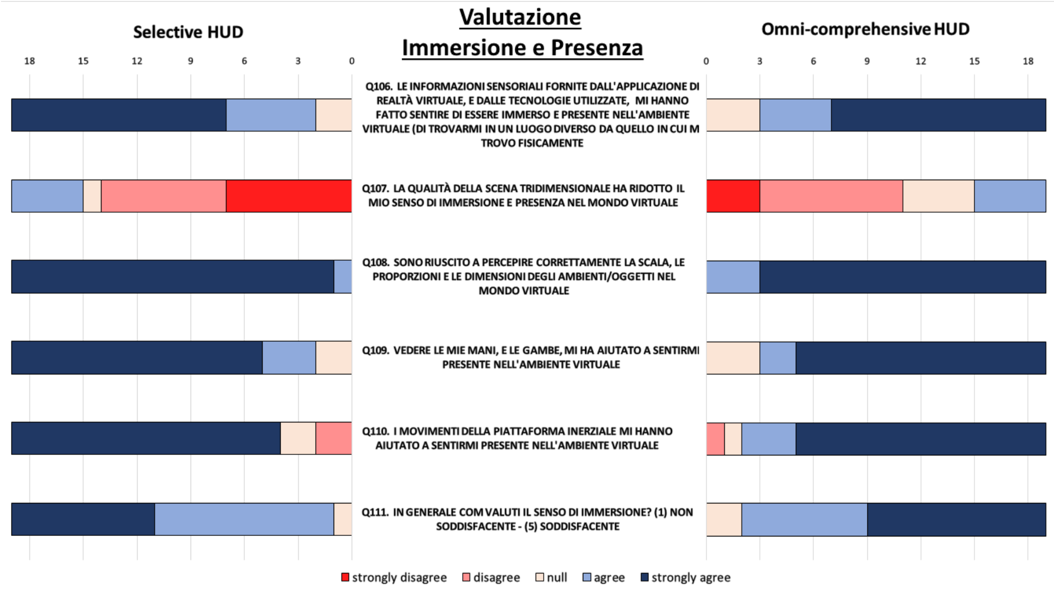


Figure 6.14: On the left the selective HUD, on the right omni-comprehensive HUD. Questions related to the evaluation of the immersion/presence. Scale starting from 1 to 5, (1) I do not agree with the statement - (5) I fully agree with the statement. Mann-Whitney Test, $^{*}p-value < 0.05$, $^{**}p-value < 0.01$, $^{***}p-value < 0.001$.

Immersion and Presence Assessment					
Question	HUD	Mean	Std. Dev.	Std. Error Mean	Mann-Withney p-value (two-tailed)
Q106	SEL	4.5263	.69669	.15983	p=.999
	OMN	4.4737	.77233	.17718	
Q107	SEL	2.1053	1.14962	.26374	p=.239
	OMN	2.4737	1.02026	.23406	
Q108	SEL	4.9474	.22942	.05263	p=.604
	OMN	4.8421	.37463	.08595	
Q109	SEL	4.6316	.68399	.15692	p=1.000
	OMN	4.5789	.76853	.17631	
Q110	SEL	4.4737	1.07333	.24624	p=.987
	OMN	4.5789	.83771	.19218	
Q111	SEL	4.3684	.59726	.13702	p=.769
	OMN	4.4211	.69248	.15887	

* statistical significance, p<.05

Table 6.14: Descriptive statistics and Mann-Whitney U Test.

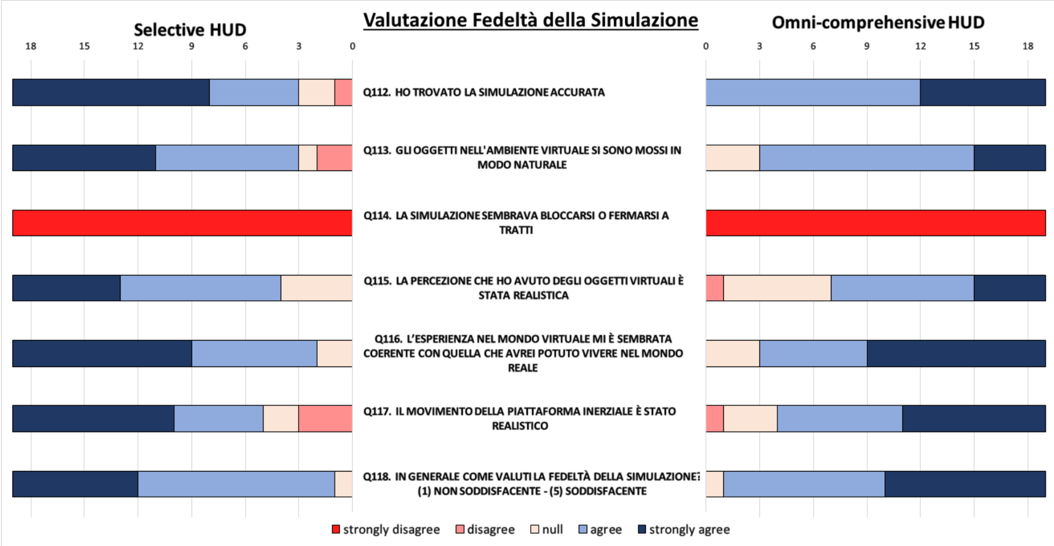


Figure 6.15: On the left the selective HUD, on the right omni-comprehensive HUD. Questions related to the evaluation of the fidelity of the simulated experience. Scale starting from 1 to 5, (1) I do not agree with the statement - (5) I fully agree with the statement. Mann-Whitney Test, $^{*}p-value < 0.05$, $^{**}p-value < 0.01$, $^{***}p-value < 0.001$.

Driving Simulation Fidelity Assessment

Question	HUD	Mean	Std. Dev.	Std. Error Mean	Mann-Withney p-value (two-tailed)
Q112	SEL	4.3684	.89508	.20535	p=.571
	OMN	4.3684	.49559	.11370	
Q113	SEL	4.1579	.95819	.21982	p=.359
	OMN	4.0526	.62126	.14253	
Q114	SEL	1.0000	,00000	0.00000	p=1.000
	OMN	1.0000	,00000	0.00000	
Q115	SEL	4.1053	.73747	.16919	p=.315
	OMN	3.7895	.85498	.19615	
Q116	SEL	4.4211	.69248	.15887	p=.966
	OMN	4.3684	.76089	.17456	
Q117	SEL	4.0526	1.12909	.25903	p=.994
	OMN	4.1579	.89834	.20609	
Q118	SEL	4.3158	.58239	.13361	p=.662
	OMN	4.4211	.60698	.13925	

* statistical significance, p<.05

Table 6.15: Descriptive statistics and Mann-Whitney U Test.

6.2 GSR Outcomes

In the following section results obtained from the GSR data have been reported. The section provide firstly the comparison between the user interfaces in term of $\Delta_{peak-to-peak}$, Δ_{max}, $\Delta_{accumulated}$ and Δ_{mean} features. Statistical analysis has been performed separately for each feature. Continuing, the next subsection provides the trend analysis separately for the user interface. Then, the previously described pre-post features are reported within the user interface. Additionally, the correlation analysis between the GSR and questionnaires is reported. Even the motion platform effect on GSR data is then analysed and showed with some consideration about the user 'presence'.

The main analysis of this thesis concerns the study of the main effect that HUD has on the EDA response.

6.2.1 Selective vs Omni-comprehensive HUD

Peak-to-Peak SCR Feature

Figure 6.16 reports the result of the analysis of the $\Delta_{peak-to-peak}$ feature between the user-interfaces. In general, the picture shows higher mean elevation in the selective HUD compared to the omni-comprehensive HUD. In addition, it can be observed that the figure shows the delta values for the test and control events; firsts are generally higher than control values, remembering that test events report the difference between *pre* and *post* in correspondence of an unexpected event during the simulation, while control events are computed from the pre-post difference for a random time value. Even these latter show higher absolute value for the selective sample. Independent t-test analysis performed on such data reports three events which are statistically different between each other, namely 'Car1', 'Car2', 'Man2' test events. These events represent the last three test events perceived as dangerous by the subjects, considering that the 'Scooter' event has not been perceived hazardous. Results of between-subject effects coming from the two-way repeated-measure ANOVA analysis are reported in Table 6.16. Significant results should have a $p - value < .05$. Analysis including control events reports a p-value of .161, meaning no significant difference, while analysis excluding control events report a significant main effect of HUD on $\Delta_{peak-to-peak}$ feature (p=.039). No significant interaction between HUD and events is reported. Moreover, within-subjects effects reports that event has significant effect on $\Delta_{peak-to-peak}$ value ($p < .001$) whether including and excluding control events. On Table 6.17 the independent t-test results analysis is presented.

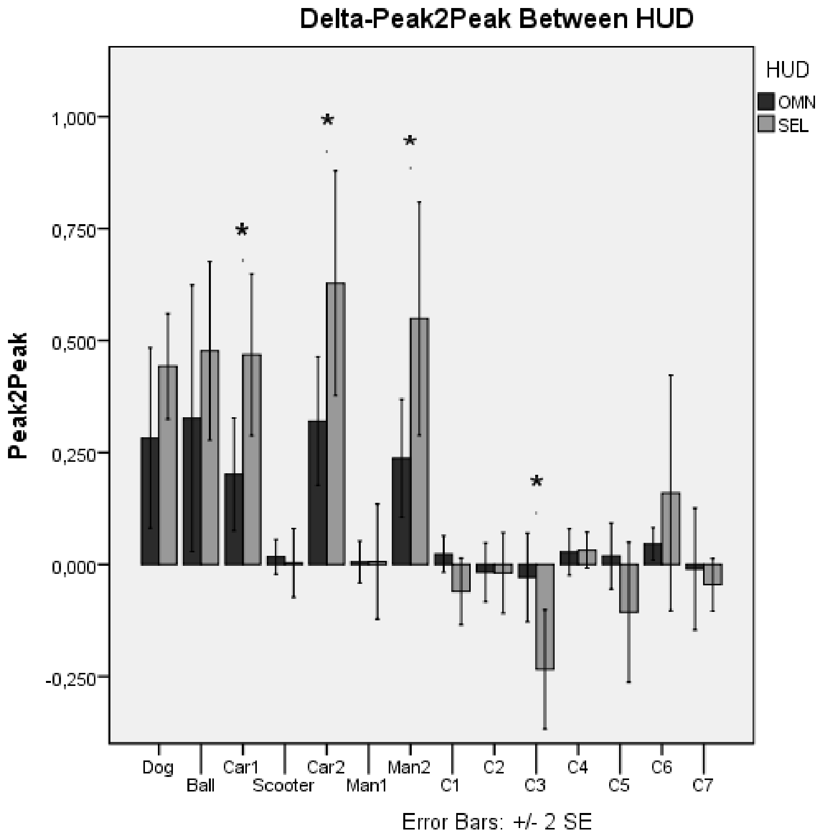


Figure 6.16: Estimated mean value of the $\Delta_{peak-to-peak}$ feature for the selective sample (gray) and omni-comprehensive (black) sample. Means with standard errors are represented.

Two-way RM ANOVA Peak to Peak	F statistic	p-value	
Within-Subjects Effects (Event)	F(5,148)=26.07	p=.000*	Test-Control Events
Between-Subjects Effects (HUD)	F(1,28)=2.08	p=.161	
Event*HUD	F(5,148)=2.71	p=.020*	
Between-Subjects Effects (HUD)	F(1,28)=4.72	p=.039*	Test Events
Within-Subjects Effects (Event)	F(6,168)=13.90	p=.000*	
Events*HUD	F(6,168)=1.74	p=.115	

Table 6.16: Two-way ANOVA with repeated measures run on the test and control events in the first row, while two-way ANOVA with repeated measures run only considering the test events, $\Delta_{peak-to-peak}$ feature. Significant results are marked for each events. Assumptions were tested, Mauchly's Test of Sphericity reported a significant p-value, therefore the Greenhouse-Geisser correction was considered when the control events were included. Excluding control events the Sphericity assumption was met ($p > .05$).

Delta-Peak2Peak Feature - Independent Samples Test					
Event	t-test for Equality of Means				
	t	df	Sig. (2-tailed)	Mean Diff.	Std. Error Diff.
Dog	1.371	28	.181	.160009	.116671
Ball	.839	28	.409	.150472	.179328
Car1	2.426	28	.022	.267268	.110148
Scooter	-0.310	28	.759	-0.013357	.043072
Car2	2.131	22.282	.044	.308076	.144580
Man1	.013	17.659	.990	.000889	.068296
Man2	2.137	20.722	.045	.311714	.145832
C1	-1.961	28	.060	-0.083374	.042509
C2	-0.025	28	.980	-0.001419	.055715
C3	-2.476	28	.020	-0.205519	.083021
C4	.137	28	.892	.004510	.032842
C5	-1.452	28	.158	-0.125568	.086495
C6	.854	28	.400	.113430	.132820
C7	-0.475	28	.639	-0.035201	.074151

Table 6.17: Independent T-Test outcomes. For *car2*, *man1* and *man2* test events the equal variance was not assumed.

Max GSR Feature

Moreover, the same analysis has been done for the Δ_{max} feature, Figure 6.17 reports the result. As above described, such a results show the difference of the mean value of the Δ_{max} feature computed for *pre* and *post* time interval in test and control event. Also for this feature, the test events present a higher elevation of *post* value for the selective user interface compared to the omni-comprehensive HUD. Also for this feature, delta absolute values are greater in selective sample for the control events. Additionally, test events present generally major response than the control events. Results of between-subject effects coming from the two-way repeated-measure ANOVA analysis are reported in Table 6.18. HUD showed a significant main effect on Δ_{max} value both including control events (p=.037) and excluding control events (p=.007). Events have a significant main including control events and excluding control events ($p < .001$). Instead, independent t-test analysis has shown a significant difference between user interface for the 'Car1', 'Car2' and 'Man2' test events, results are reported in Table 6.19.

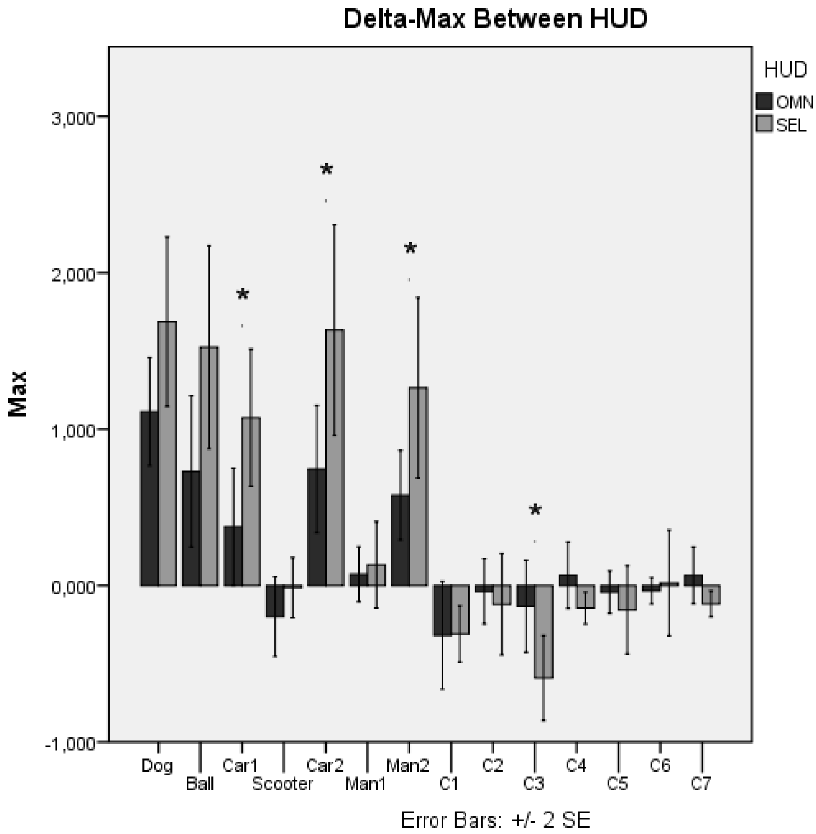


Figure 6.17: Estimated mean value of Δ_{max} feature for the selective sample (grey) and omni-comprehensive (black) sample. Means with standard errors are represented for each events.

Two-way RM ANOVA Max	F statistic	p-value	
Within-Subjects Effects (Event)	F(5,140)=26.75	p=.000*	Test-Control Events
Between-Subjects Effects (HUD)	F(1,28)=4.80	p=.037*	
Event*HUD	F(5,140)=3.37	p=.007*	
Within-Subjects Effects (Event)	F(6,168)=18.98	p=.000*	Test Events
Between-Subjects Effects (HUD)	F(1,28)=8.53	p=.007*	
Event*HUD	F(6,168)=1.44	p=.202	

Table 6.18: Two-way ANOVA with repeated measures run on the test and control events in the first row, while two-way ANOVA with repeated measures run only considering the test events, Δ_{max} feature. Significant result are marked. Assumptions were tested, Mauchly's Test of Sphericity reported a significant p-value, therefore the Greenhouse-Geisser correction was considered when the control events were included. Excluding control events the Sphericity assumption was met ($p > .05$).

Delta-Max Feature - Independent Samples Test					
Event	t-test for Equality of Means				
	t	df	Sig. (2-tailed)	Mean Diff.	Std. Error Diff.
Dog	1.790	28	.084	.575628	.321506
Ball	1.960	28	.060	.793336	.404775
Car1	2.420	28	.022	.697586	.288303
Scooter	1.155	28	.258	.185093	.160203
Car2	2.257	23.03	.034	.888759	.393807
Man1	.368	28	.716	.060262	.163909
Man2	2.135	20.50	.045	.686835	.321633
C1	.046	28	.964	.008911	.194217
C2	-0.429	28	.671	-0.082834	.192905
C3	-2.290	28	.030	-0.458603	.200290
C4	-1.784	28	.085	-0.209562	.117468
C5	-0.724	28	.475	-0.113233	.156455
C6	.286	15.75	.779	.049840	.174300
C7	-1.825	28	.079	-0.182529	.100042

Table 6.19: Independent T-Test outcomes. For *car2*, *man2* and *C6* events the equal variance was not assumed.

Accumulated GSR Feature

Figure 6.18 reports the result of the analysis of the $\Delta_{accumulated}$ feature between the user-interfaces. Generally, the picture shows higher mean elevation in the selective HUD compared to the omni-comprehensive HUD. In addition, even in this case, delta values for the test and control events are showed; firsts are generally higher than control values. These latter report higher absolute value for the selective sample. Independent t-test analysis, Table 6.21, performed on such data reports three events which are statistically different between each other, namely 'Ball', 'Car1' and 'Car2' test events. Results of between-subject effects coming from the two-way repeated-measure ANOVA analysis are reported in Table 6.20. Analysis including control events reports a significant main effect of HUD on $\Delta_{accumulated}$ feature (p=.042), and also excluding control events (p=.006). In addition, within-subjects effects reports that event has significant effect on $\Delta_{accumulated}$ value ($p < .001$) both including and excluding control events.

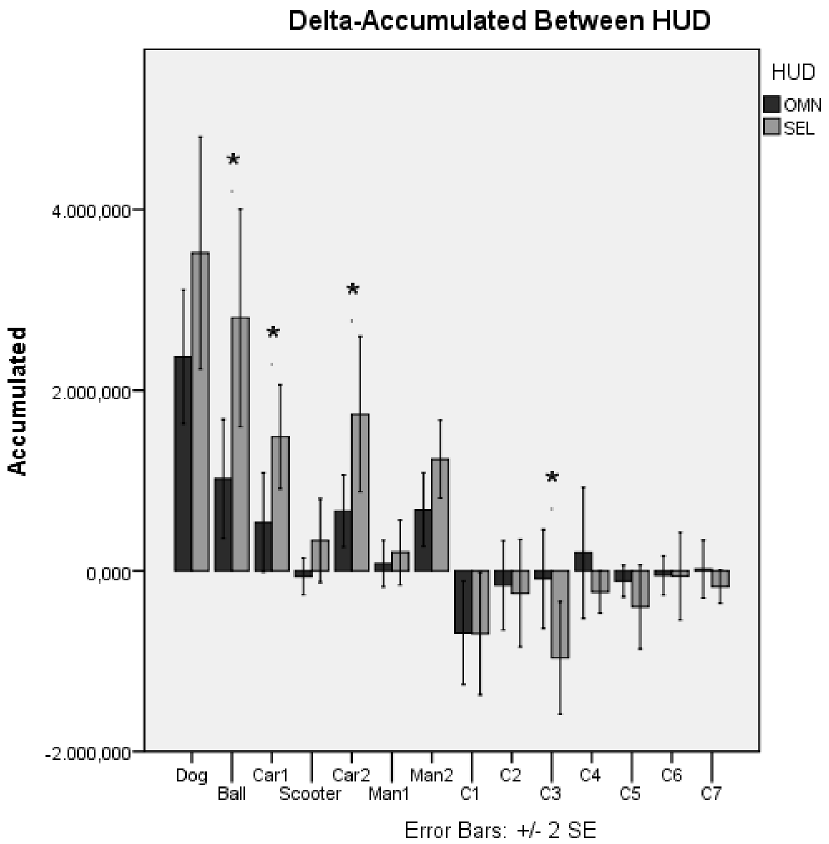


Figure 6.18: Estimated mean value of the $\Delta_{accumulated}$ feature for the selective sample (gray) and omni-comprehensive (black) sample. Means with standard errors are represented for each events.

Two-way RM ANOVA Accumulated	F statistic	p-value	
Within-Subjects Effects (Event)	F(4,137)=25.07	p=.000*	Test-Control Events
Between-Subjects Effects (HUD)	F(1,28)=4.55	p=.042*	
Event*HUD	F(4,137)=3.17	p=.010*	
Within-Subjects Effects (Event)	F(3,96)=22.86	p=.000*	Test Events
Between-Subjects Effects (HUD)	F(1,28)=9.02	p=.006*	
Event*HUD	F(3,96)=3.43	p=.149	

Table 6.20: Two-way ANOVA with repeated measures run on the test and control events in the first row, while two-way ANOVA with repeated measures run only considering the test events, $\Delta_{accumulated}$ feature. Significant result are marked. Assumptions were tested, Mauchly's Test of Sphericity reported a significant p-value, therefore the Greenhouse-Geisser correction was considered.

Delta-Accumulated - Independent Samples Test					
Event	t-test for Equality of Means				
	t	df	Sig. (2-tailed)	Mean Diff.	Std. Error Diff.
Dog	1.549	28	.133	1150.413218	742.807084
Ball	2.594	28	.015	1780.950291	686.588575
Car1	2.378	28	.024	947.311740	398.407183
Scooter	1.574	19.21	.132	398.207702	253.013252
Car2	2.263	28	.032	1073.185529	474.171345
Man1	.558	28	.581	123.537724	221.462374
Man2	1.887	28	.070	559.428510	296.531501
C1	-0.022	28	.982	-9.934304	444.804949
C2	-0.233	28	.817	-90.218385	386.867163
C3	-2.106	28	.044	-875.393597	415.668187
C4	-1.135	28	.266	-433.463715	381.992970
C5	-1.148	17.97	.266	-285.760832	248.925785
C6	-0.020	28	.984	-5.452564	265.988001
C7	-1.052	28	.302	-194.659209	185.012099

Table 6.21: Independent T-Test outcomes. For *scooter*, *C5* events the equal variance was not assumed.

Mean GSR Feature

Last feature considered is the Δ_{mean}, Figure 6.19 reports the result of the analysis between the user-interfaces. In general, data report higher mean elevation in the selective HUD compared to the omni-comprehensive HUD. Even in this case, delta values for the test and control events are showed; firsts are generally higher than control values. These latter report higher absolute value for the selective sample. Independent t-test analysis, Table 6.23, performed on such data reports three events which are statistically different between each other, namely 'Ball', 'Car1' and 'Car2' test events. Results of between-subject effects coming from the two-way repeated measure ANOVA analysis are reported in Table 6.22. Analysis including control events reports a significant main effect of HUD on Δ_{mean} feature (p=.041), and also excluding control events (p=.005). In addition, within-subjects effects reports that event has significant effect on $\Delta_{accumulated}$ value ($p < .001$) both including and excluding control events.

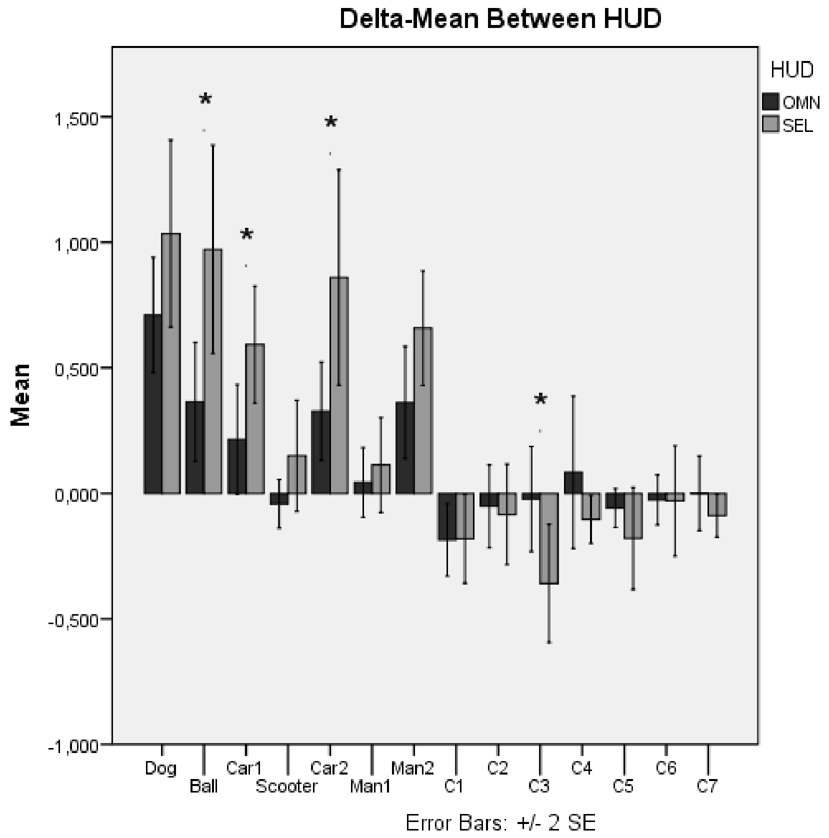


Figure 6.19: Estimated mean value of Δ_{mean} feature for the selective sample (grey) and omni-comprehensive (black) sample. Means with standard errors are represented for each events.

Two-way RM ANOVA Mean	F statistic	p-value	
Within-Subjects Effects (Event)	F(6,168)=20.78	p=.000*	Test-Control Events
Between-Subjects Effects (HUD)	F(1,28)=4.60	p=.041*	
Event*HUD	F(6,168)=3.24	p=.005*	
Within-Subjects Effects (Event)	F(6,168)=14.38	p=.000*	Test Events
Between-Subjects Effects (HUD)	F(1,28)=9.36	p=.005*	
Event*HUD	F(6,168)=1.36	p=.124	

Table 6.22: Two-way ANOVA with repeated measures run on the test and control events in the first row, while two-way ANOVA with repeated measures run only considering the test events, Δ_{mean} feature. Significant result are marked. Assumptions were tested, Mauchly's Test of Sphericity reported a significant p-value, therefore the Greenhouse-Geisser correction was considered when the control events were included. Excluding control events the Sphericity assumption was met ($p > .05$).

Delta-Mean Feature - Independent Samples Test					
Event	t-test for Equality of Means				
	t	df	Sig. (2-tailed)	Mean Diff.	Std. Error Diff.
Dog	1.478	23.28	.153	.323565	.218945
Ball	2.543	28	.017	.606791	.238625
Car1	2.365	28	.025	.377753	.159755
Scooter	1.591	19.14	.128	.192206	.120820
Car2	2.258	19.61	.036	.533246	.236163
Man1	.598	28	.555	.070103	.117234
Man2	1.854	28	.074	.296507	.159952
C1	.038	28	.970	.004390	.114380
C2	-0.251	28	.804	-0.032550	.129697
C3	-2.134	28	.042	-0.335834	.157363
C4	-1.179	28	.248	-0.187625	.159200
C5	-1.119	28	.273	-0.121700	.108796
C6	-0.031	28	.975	-0.003787	.120719
C7	-1.033	28	.311	-0.088798	.085989

Table 6.23: Independent T-Test outcomes. For *dog*, *scooter* and *car2* events the equal variance was not assumed.

6.2.2 GSR Trend

Trend analysis has been performed to evaluate the EDA behaviour during the driving simulations. User interfaces have been considered separately; then, Figure 6.20 reports the mean trend computed for the selective HUD, while Figure 6.21 shows the mean trend for the omni-comprehensive user interface. As previously reported in subsection 5.5.1, to allow the averaging the values are min-max scaled for each subject. Both figures show similar characteristics; recording in VR-baseline environment reports higher values than the baseline collected in non-virtual reality.

With the beginning of the driving simulation, EDA values increase. Continuing the simulation, users appear to get confidence in the automated driving system; decreasing EDA values are observable until the first test event is delivered. Such an event is very similar between user interfaces, considering that it is the first test event. Then, the trend is similar; however, differences in EDA response in test event are notable. The SCL has a slightly decreasing behaviour in the middle of the driving simulation; likely the subject has built trust on the vehicle, or he has acclimated to the VR environment. During the last part of the simulation, the driving context becomes to increase in term of complexity, a higher number of vehicles is present into the environment; a subsequent higher cognitive demand is required. For such reason, the last values of the EDA report a slightly increasing trend for both HMI.

Both users interfaces present SCR components as a response to the stimulus represented by the test events, even if qualitatively omni-comprehensive sample presents a smaller response.

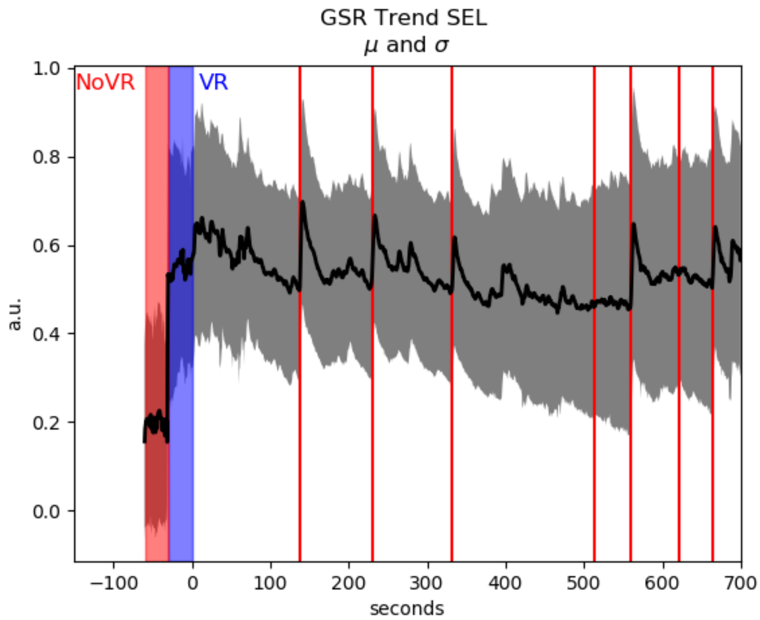


Figure 6.20: Black curve: Mean value. Grey curves: standard deviation value. Normalized raw GSR signal through the subject who tested the Selective HUD.

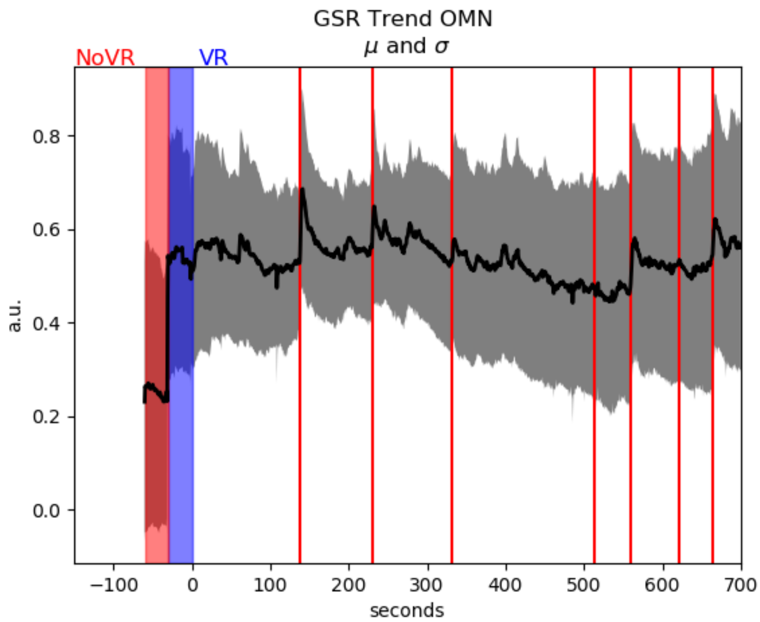


Figure 6.21: Black curve: Mean value. Grey curves: standard deviation value. Normalized raw GSR signal through the subject who tested the Omni-comprehensive HUD.

6.2.3 Selective HUD

The following subsection reports the selective user interface EDA results. Such considerations are done to evaluate if such physiological measure can provide support to evaluate users reactions about the virtual driving experience.

A qualitative analysis is effectuated to check the EDA response during a time interval of 10 seconds before and 10 seconds post the test event. Figure 6.22 reports the result of such a qualitative analysis, the picture represent information related to filtered EDA behaviour separately for each test event and jointly for all test events. More specifically, Figure 6.22a expresses the skin conductance response, obtained thanks to the application of a digital filter, for the seven test events: *dog*, *ball*, *car1*, *scooter*, *car2*, *man1* and *man2*. As already reported from questionnaire in the section which evaluates test events, such situations are not perceived dangerous at the same level, more in detail the most hazardous events evaluated are *car2* and *man2* while no perceived danger was reported for *scooter* and *man1*. EDA results, qualitatively reports similar outcomes, in fact, in Figure 6.22a, *car2* and *man2* test events shows the main response, while no EDA response is visible in *scooter* and *man1* events. SCR values are normalized for the first value in order to observe variation rather absolute value.

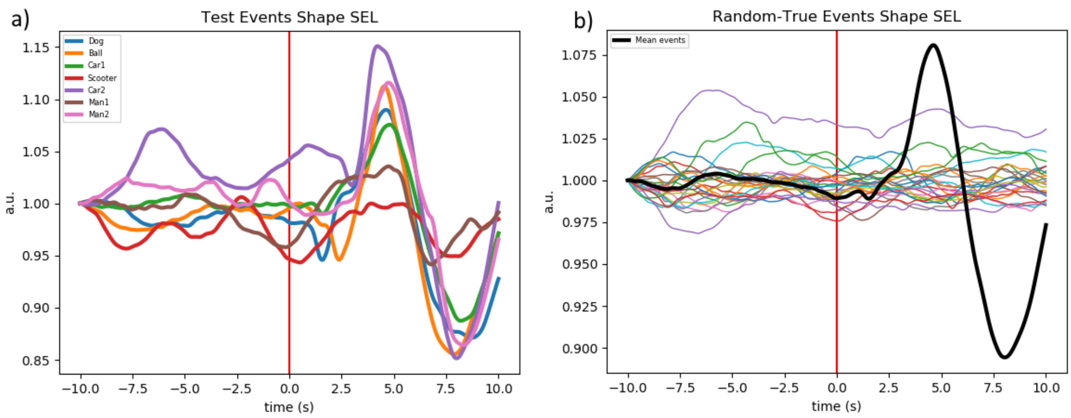


Figure 6.22: a) The mean of the SCR epochs for all the subject 10s prior and post a single test event. b) Black curve: mean of the SCR epochs for all the subject 10s prior and post of all test events. Coloured curves are the averaging result of 10 random sequence of fake test events epochs mediated each other.

On the other side, Figure 6.22b check the validity of such a biosignal to assess the test events. In this case, it is represented one main shape representing the mean value of all the subjects for all the test events in the 10 seconds prior and after the event. The figure reports an evident SCR component. The same identical algorithm has been followed to examine 30 random time-point sequences. Each time point sequence is composed of seven random time point, and these time point acted as a control sequence, thus averaged thought subjects and thought each time point within the sequence. Such random or control curves shows the effectiveness of EDA measurement to evaluate the user emotional response to a stimulus. Also such curves are normalised for the initial value to observe variation rather absolute values.

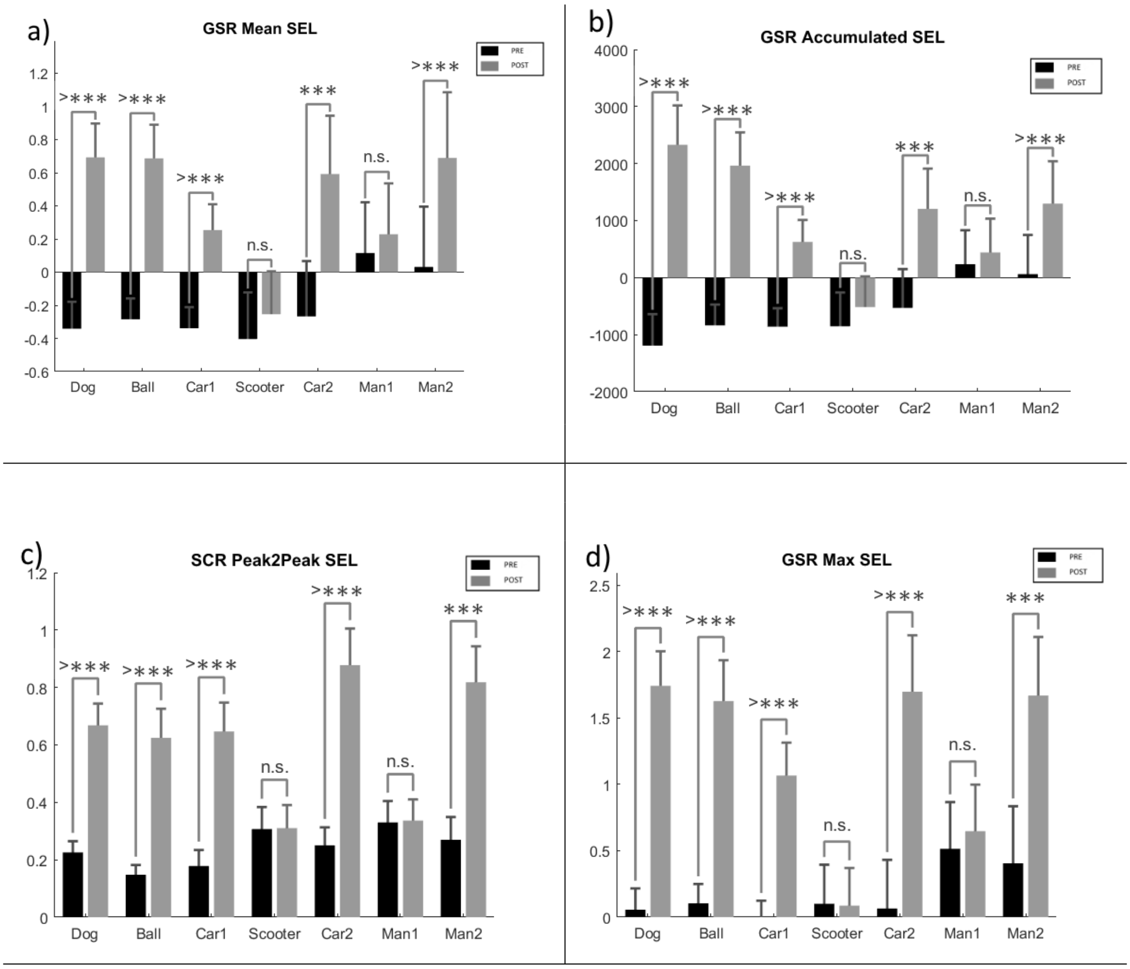


Table 6.24: a): *Mean* values computed for the selective sample in time interval prior and sequent the test event. Means are displayed with standard errors. b): *Accumulated* values computed for the selective sample in time interval prior and sequent the test event. Means are displayed with standard errors. c): *Peak-to-Peak* values computed for the selective sample in time interval prior and sequent the test event. Means are displayed with standard errors. d): *Max* values computed for the selective sample in time interval prior and sequent the test event. Means are displayed with standard errors.

Furthermore, a quantitative analysis of *pre-post* EDA behaviour has been performed. The four features have been separately examined to check the effectiveness of such signal's characteristics, Table 6.24 reports the outcomes of the just mentioned analysis.

Figure 6.24a reports the mean values and standard errors of the *GSR Mean* feature for the seven test events. As notable, almost all test events present an

PRE-POST RELATED T-TEST	Mean GSR	Accumulated GSR	Peak2Peak SCR	Max GSR
Dog	t:-5.546 p:0.000	t:-5.478 p:0.000	t:-7.511 p:0.000	t:-6.222 p:0.000
Ball	t:-4.690 p:0.000	t:-4.650 p:0.000	t:-4.774 p:0.000	t:-4.694 p:0.000
Car1	t:-5.073 p:0.000	t:-5.150 p:0.000	t:-5.185 p:0.000	t:-4.896 p:0.000
Scooter	t:-1.355 p:0.197	t:-1.454 p:0.168	t:-0.093 p:0.927	t:0.144 p:0.888
Car2	t:-4.003 p:0.001	t:-4.040 p:0.001	t:-5.004 p:0.000	t:-4.848 p:0.000
Man1	t:-1.199 p:0.251	t:-1.134 p:0.276	t:-0.103 p:0.920	t:-0.959 p:0.354
Man2	t:-5.754 p:0.000	t:-5.752 p:0.000	t:-4.217 p:0.001	t:-4.389 p:0.001

Table 6.25: Related t-test analysis of the selective sample.

elevation of the *post* values, mainly for 'Dog', 'Ball', 'Car2' and 'Man2' events. As it can be notable, such measure is computed on the z-score GSR, in fact, some values are negative, and they are not expressed in μS. Between *pre-post*, EDA values are significantly different, exception are provided by 'Scooter' and 'Man1' test events, Table 6.25 reports the statistic analysis outcomes. The main *pre-post* difference is observed for the 'Dog' test event.

Figure 6.24b, instead, reports the mean values and standard errors of the *GSR Accumulated* feature for the seven test events. From such a figure, all test events present an elevation of the *post* values. Even in this case, some measures are negative. *GSR Accumulated* feature is calculated summing the GSR values within the time interval; therefore the sum of negative values returns a negative value. Between *pre-post*, EDA values are significantly different, exception are provided by 'Scooter' and 'Man1' test events, Table 6.25 reports the statistic analysis outcomes. The main *pre-post* difference is observed for the 'Dog' test event.

Proceeding, Figure 6.24c reports the mean values and standard errors of the *SCR Peak-to-Peak* feature for the seven test events. As notable, almost all test events present an evident elevation of the *post* values. Between *pre-post*, EDA values are significantly different, exception are provided by 'Scooter' and 'Man1' test events, Table 6.25 reports the statistic analysis outcomes. The main *pre-post* difference is observed for the 'Car2' and 'Man2' test events.

The last Figure 6.24d, instead, shows the mean values and standard errors of the *GSR Max* feature for the seven test events. From such a figure, almost all test events present an increase of the *post* values. Between *pre-post*, EDA values are significantly different, exception are provided by 'Scooter' and 'Man1' test events, Table 6.25 reports the statistic analysis outcomes. Main *pre-post* differences are observed for the 'Dog','Ball', 'Car2' and 'Man2' test events.

All four features qualitatively and quantitatively analysed represents adequate indicator to assess the presence of the stimulus delivered through the virtual driving simulation.

6.2.4 Omni-comprehensive HUD

The following subsection reports the omni-comprehensive user interface EDA results. It has been performed the same analysis previously showed for the selective sample.

Firstly, a qualitative study is effectuated to examine the EDA response during a time interval of 10 seconds before and 10 seconds after the test event. Figure 6.23 reports the result of such analysis, the picture represents information related to filtered EDA behaviour separately for each test event and jointly for all test events. In particular, Figure 6.23a expresses the skin conductance response, obtained thanks to the application of a digital filter, for the seven test events: *dog*, *ball*, *car1*, textitscooter, *car2*, *man1* and *man2*. As already stated from the questionnaire in the section which judges test events, such situations are not perceived dangerous at an identical level, more specifically the most hazardous events valued are *car2* and *man2* while no perceived danger was reported for *scooter* and *man1*, slightly risk was judged the *car1* event. EDA results, qualitatively reports similar outcomes, in fact, in Figure 6.23a, *dog*, *ball* and *man2* test events shows the main response, while no EDA response is visible in *scooter* and *man1* events. SCR values are normalized for the first value in order to observe variation rather absolute value.

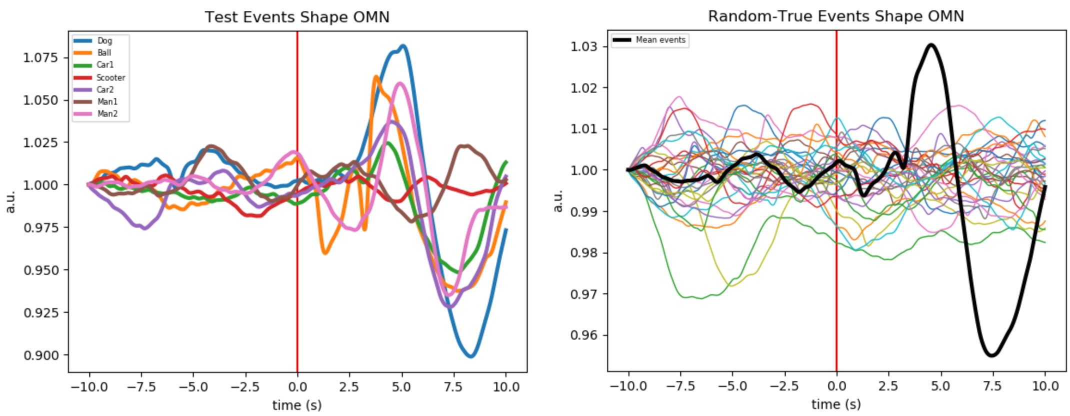


Figure 6.23: a) The mean of the SCR epochs for all the subject 10s prior and post a single test event. b) Black curve: mean of the SCR epochs for all the subject 10s prior and post of all test events. Coloured curves are the averaging result of 10 random sequence of fake test events epochs mediated each other.

On the other side, Figure 6.23b examines the validity of such a physiological measure to assess the test events. In this site, the figure denotes one main shape representing the mean value of all the subjects for all the test events in the 10 seconds preceding and following the event. The figure reports an evident SCR component. The same algorithm has been developed to examine 30 random time-point sequences. Each time point sequence is composed of 7 random time point; these time point acted as a control sequence, thus averaged thought subjects and thought each time point within the series. Such random or control curves confirm the effectiveness of EDA measurement to evaluate the user emotional response to a stimulus. Also such curves are normalised for the initial value in order to observe variation rather absolute values.

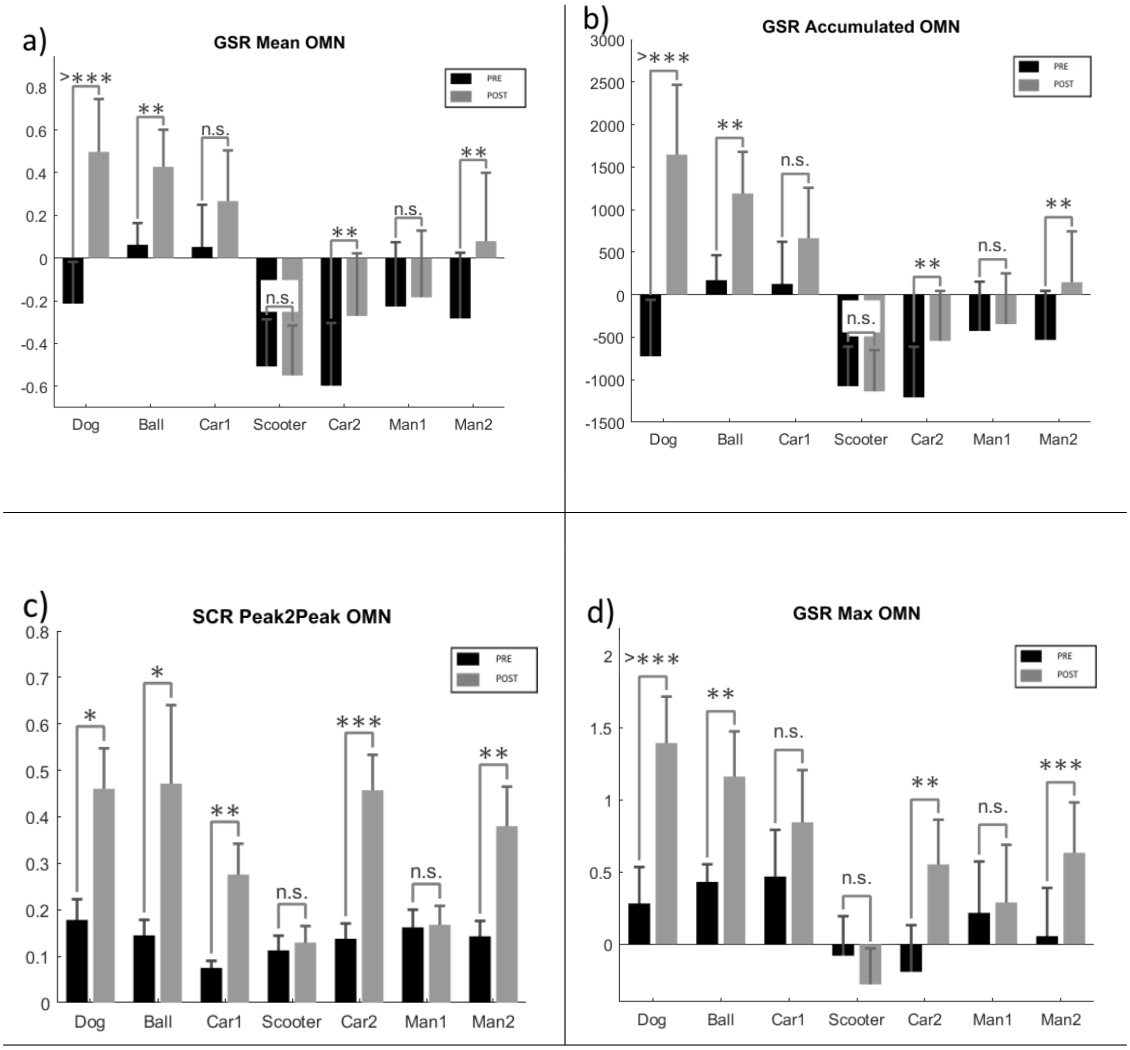


Table 6.26: a): *Mean* values computed for the omni-comprehensive sample in time interval prior and sequent the test event. b): *Accumulated* values computed for the omni-comprehensive sample in time interval prior and sequent the test event. c): *Peak-to-Peak* values computed for the omni-comprehensive sample in time interval prior and sequent the test event. d): *Max* values computed for the omni-comprehensive sample in time interval prior and sequent the test event. Means are displayed with standard errors.

From the observation of Figure 6.22 and Figure 6.23, some considerations can be done. The interesting difference between two user interfaces is that the mean shape curve, englobing all test events, for the omni-comprehensive user interface is tinier compared to the selective sample. Moreover, another

PRE-POST RELATED T-TEST	Mean GSR	Accumulated GSR	Peak2Peak SCR	Max GSR
Dog	t:-6.188 p:0.000	t:-6.371 p:0.000	t:-2.804 p:0.014	t:-6.439 p:0.000
Ball	t:-3.083 p:0.008	t:-3.093 p:0.008	t:-2.191 p:0.046	t:-3.021 p:0.009
Car1	t:-1.970 p:0.069	t:-1.960 p:0.070	t:-3.188 p:0.007	t:-2.008 p:0.064
Scooter	t:0.873 p:0.398	t:0.604 p:0.555	t:-0.867 p:0.401	t:1.557 p:0.142
Car2	t:-3.325 p:0.005	t:-3.308 p:0.005	t:-4.453 p:0.001	t:-3.658 p:0.003
Man1	t:-0.620 p:0.545	t:-0.630 p:0.539	t:-0.244 p:0.811	t:-0.828 p:0.422
Man2	t:-3.235 p:0.006	t:-3.321 p:0.005	t:-3.602 p:0.003	t:-4.039 p:0.001

Table 6.27: Related t-test outcomes for omni-comprehensive sample.

consideration is that 'Car1' EDA response event is more evident in the selective sample.

Moreover, even for the omni-comprehensive sample, a quantitative analysis of *pre-post* EDA behaviour has been performed. The four features have been separately studied to examine the effectiveness of such signal's properties, Table 6.26 reports the outcomes.

Figure 6.26a reports the mean values and standard errors of the *GSR Mean* feature for the seven test events. As it can be seen, all test events present an increase of the *post* values, mainly for 'Dog', 'Ball', and 'Man2' events. Between *pre-post*, EDA values are significantly different, exception are provided by 'Car1', 'Scooter' and 'Man1' test events, Table 6.27 reports the statistic analysis outcomes. The main *pre-post* difference is observed for the 'Dog' test event.

Figure 6.24b, instead, shows the mean values and standard errors of the *GSR Accumulated* feature for the seven test events. From such a figure, all test events present an elevation of the *post* values. Between *pre-post*, EDA values are significantly different, even here exception are represented by'Car1', 'Scooter' and 'Man1' test events, Table 6.27 reports the statistic analysis outcomes. The main *pre-post* difference is observed for the 'Dog' test event.

Continuing, Figure 6.24c addresses the mean values and standard errors of the *SCR Peak-to-Peak* feature for the seven test events. Almost all test events present an evident elevation of the *post* values. Between *pre-post*, EDA values are significantly different, exception are provided by 'Scooter' and 'Man1' test events, Table 6.27 reports the statistic analysis outcomes. The main *pre-post* differences are observed for the 'Dog' , 'Ball', 'Car2' and 'Man2' test events.

Finally, the last Figure 6.24d, instead, shows the mean values and standard errors of the *GSR Max* feature for the seven test events. From the figure, all test events exhibit an increment of the *post* values. Between *pre-post*, EDA values are significantly different, exception are provided by 'Car1', 'Scooter' and 'Man1' test events, Table 6.27 reports the statistic analysis outcomes. Main *pre-post* differences are observed for the 'Dog' test event.

Even for the omni-comprehensive sample, the analysed four features qualitatively and quantitatively represents an adequate indicator to assess the presence of the stimulus delivered through the virtual driving simulation and to what extent.

6.2.5 GSR vs Questionnaire

A qualitative analysis of the correlation between SCR signal and questionnaire response is notable in 6.24. Instead, a more appropriate quantitative analysis of dependence between questionnaire and GSR response for the respective user interface is computed building a multinomial regression model.

Figures 6.24a and 6.24b have been already presented in the previous section; they represent the mean shape over the subjects for each test event, while Figure 6.24c expresses the mean value of the questionnaire response at the question which accounts the perceived risk. It can be observed that test events are generally rated more dangerous by selective users than omni-comprehensive users.

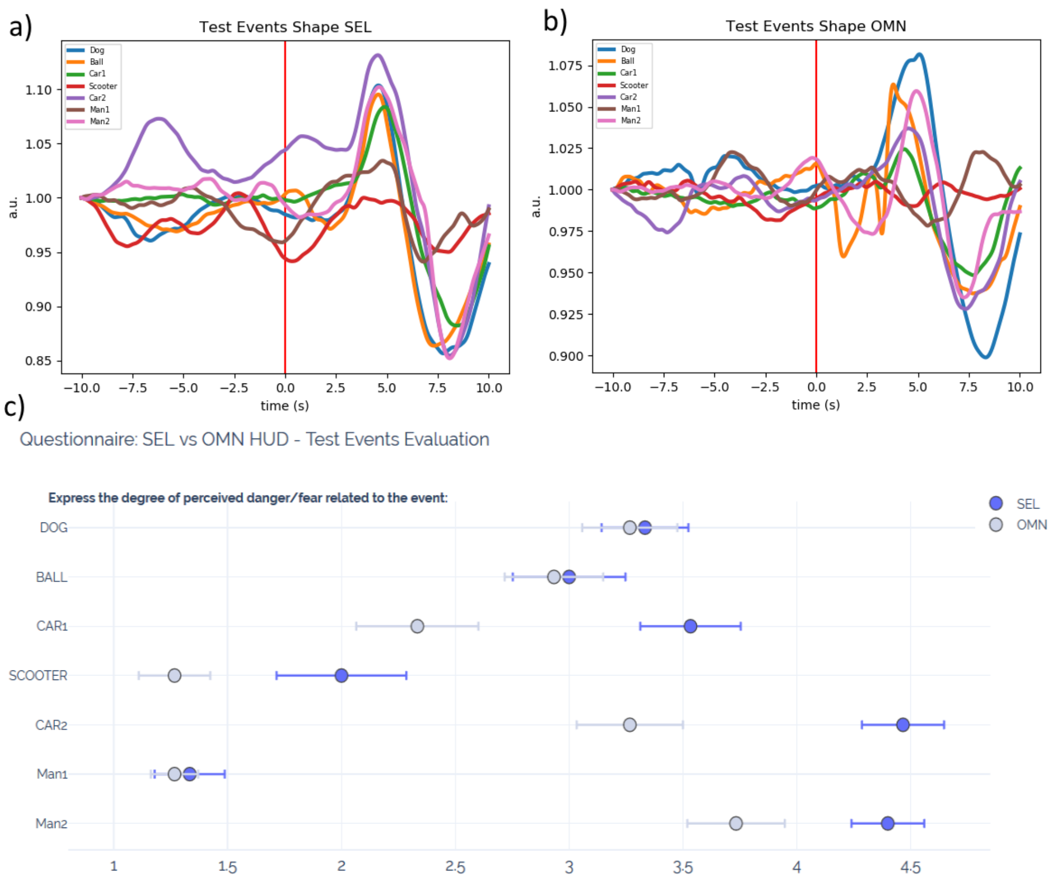


Figure 6.24: Study2 summary: qualitative correlation between the SCR responses and the questionnaire ratings.

Peak-to-peak data confirm these findings; skin conductance responses are in amplitude higher for the selective sample. The most hazardous event rated from the selective users is the *car2* test event and the *man2* event for the omni-comprehensive sample, while in *car2* the higher SCR for the selective HUD is measured and in *man2* for omni-comprehensive HUD. Moreover, Figure 6.24c shows that selective and omni-comprehensive user interfaces start to differentiate each other from the second event; the *ball* test event.

As previously described in section 5.7, a multinomial regression model has been created for each sample of the user interface and each test event; results are presented in Table 6.28.

Significant results obtained for each test events category express that predictor variable have a significant main effect on the goodness of fit of the classificator built. This is a measure of correlation between the discrete ratings of the questionnaire and continuous measurements represented by the SCR $\Delta_{peak-to-peak}$ feature for each user interface. No significant main effect has been found in *dog* and *man1* for the selective sample, in *scooter* test event for the omni-comprehensive sample.

MULTINOMIAL REGRESSION	SELECTIVE	OMNICOMPRENSIVE
Dog	$\chi^2(21)=18.42$ $p=.622$	$\chi^2(21)=37.05$ $p=.017$*
Ball	$\chi^2(28)=44.69$ $p=.024$*	$\chi^2(21)=37.96$ $p=.013$*
Car1	$\chi^2(21)=38.04$ $p=.013$*	$\chi^2(21)=35.13$ $p=.027$*
Scooter	$\chi^2(21)=34.10$ $p=.035$*	$\chi^2(14)=18.83$ $p=.172$
Car2	$\chi^2(14)=25.60$ $p=.029$*	$\chi^2(21)=37.56$ $p=.015$*
Man1	$\chi^2(14)=18.83$ $p=.172$	$\chi^2(7)=17.40$ $p=.015$*
Man2	$\chi^2(14)=27.83$ $p=.015$*	$\chi^2(21)=32.33$ $p=.054$

Table 6.28: Study2: table representing the multinomial regression model between GSR data and subjective response at the perception of risk/fear question., *p value<0.05, **p value<0.01.

In order to have another confirmation of the correlation between the GSR and the questionnaire outcomes, also the Study1 results are reported in Figure 6.25 and in Table 6.29 making the same reasoning effectuated for the Study2 upper described.

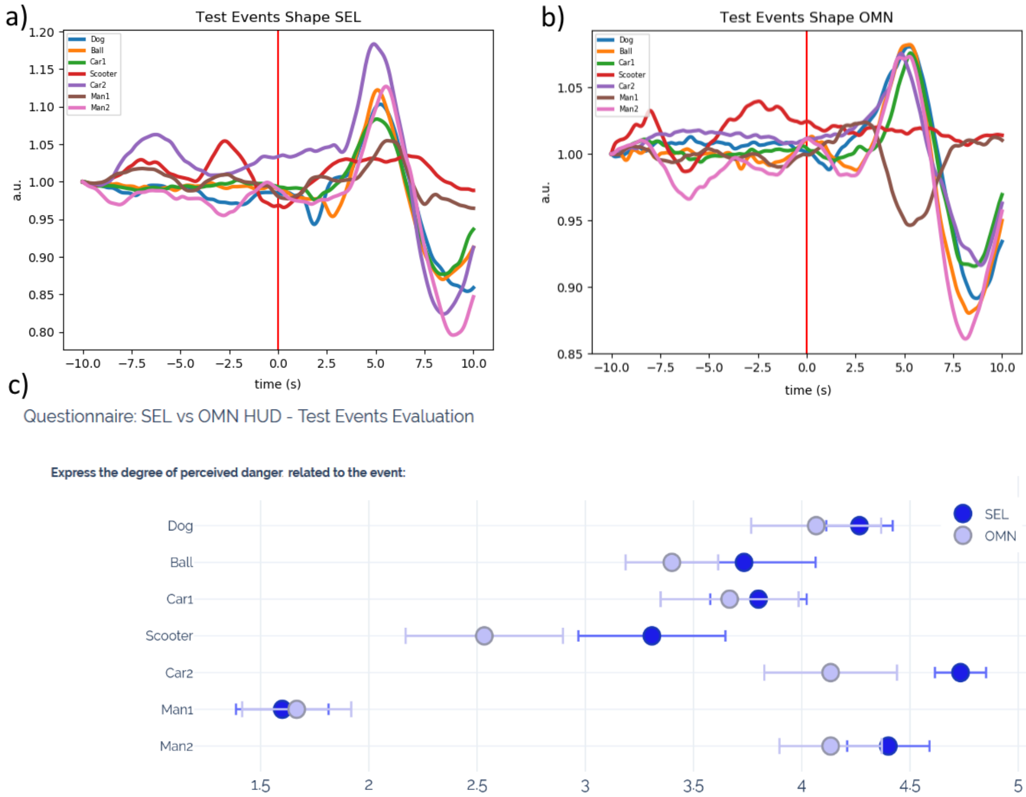


Figure 6.25: Study1 summary: qualitative correlation between the SCR response and the questionnaire ratings. The strongest responses in filtered GSR signal are present in the most hazardous rated situation for the relative user interface.

MULTINOMIAL REGRESSION STUDY1	SELECTIVE	OMNICOMPRENSIVE
Dog	$\chi^2(14)=11.09$ $p=.680$	$\chi^2(21)=35.83$ $p=.023*$
Ball	$\chi^2(28)=43.51$ $p=.031*$	$\chi^2(21)=35.47$ $p=.025*$
Car1	$\chi^2(21)=36.31$ $p=.020*$	$\chi^2(28)=43.105$ $p=.034*$
Scooter	$\chi^2(28)=38.46$ $p=.090$	$\chi^2(28)=45.74$ $p=.019*$
Car2	$\chi^2(7)=17.40$ $p=.015*$	$\chi^2(21)=34.79$ $p=.030*$
Man1	$\chi^2(14)=28.51$ $p=.012*$	$\chi^2(21)=32.33$ $p=.054$
Man2	$\chi^2(14)=29.10$ $p=.010**$	$\chi^2(14)=22.51$ $p=.069$

Table 6.29: Study1: table representing the multinomial regression model between GSR data and subjective response at the perception of risk question., *p value<0.05, **p value<0.01.

6.2.6 Motion Platform Test

In this section are shown the results obtained utilising the same algorithm used for the selective and omni-comprehensive GSR data analysis. This approach has been used in order to check the validity of the GSR data obtained during the selective and omni-comprehensive HUD test sessions testing the possibility of *Movement artifacts* due to the movement of the motion platform. The core idea is to compare the GSR data obtained during the selective or omni-comprehensive test session with the GSR data recorded while the motion platform was following the same movements performed in simulated autonomous driving but without wearing the headset HTC-VIVE and with closed eyes, therefore without being in virtual reality or boing focused on the simulated road.

Figure 6.26 report the results of the $\Delta_{accumulated}$ feature collected for the user experience during the virtual driving simulation (i.e Test) and while seating on the active motion platform and the user held eyes closed (i.e. Control). Such analysis is performed on the same six subjects for both test sessions; therefore a related t-test analysis has been conducted. Results show a significant difference between virtual driving experience and only motion platform active.

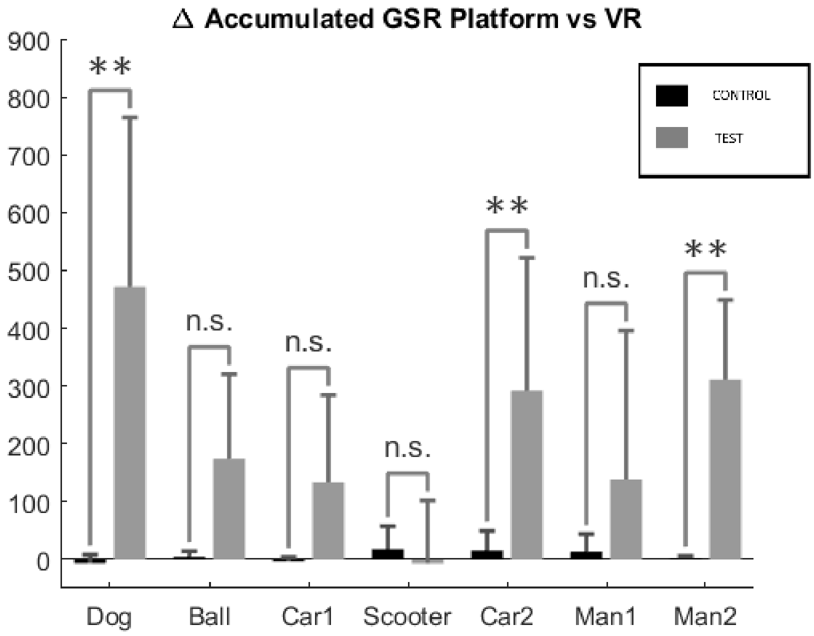
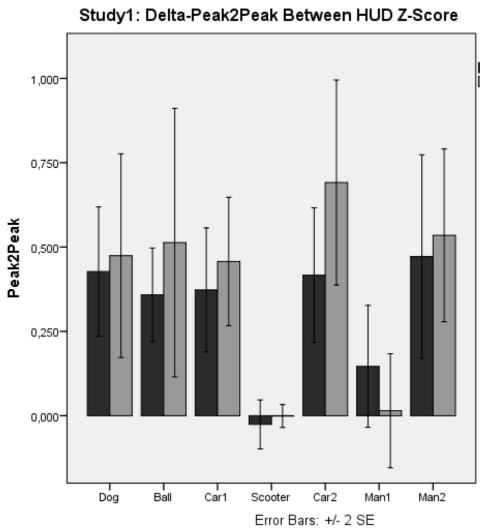
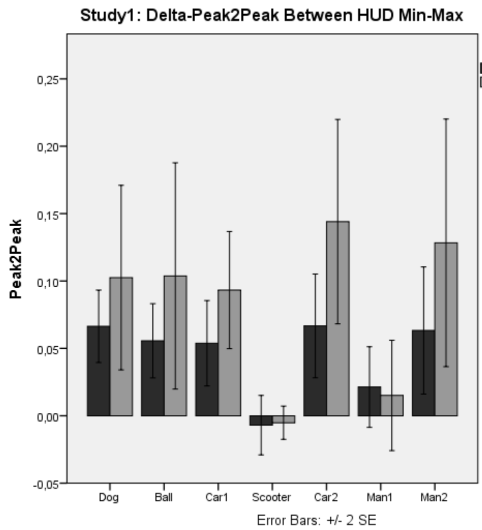


Figure 6.26: Delta difference of the accumulated GSR value computed for 10s prior and post a test event. Black bar is the No VR dataset, gray the VR dataset

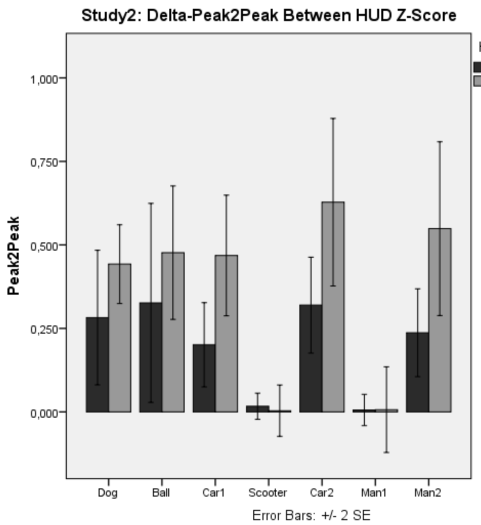
6.2.7 Study1 vs Study2



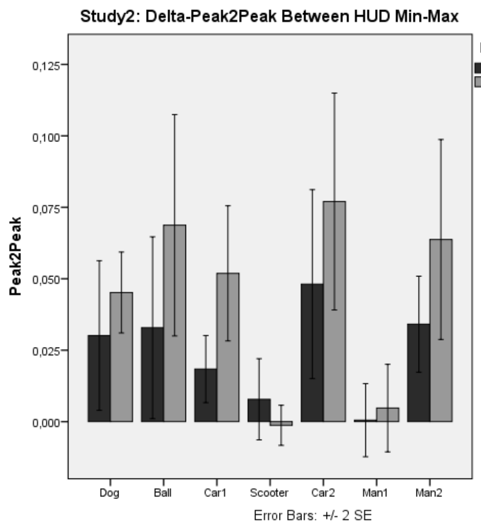
(a) Study1 $\Delta_{peak-to-peak}$ feature computed from the z-score standardised and filtered GSR.



(b) Study1 $\Delta_{peak-to-peak}$ feature computed from the minmax normalized and filtered GSR.



(c) Study2 $\Delta_{peak-to-peak}$ feature computed from the z-score standardised and filtered GSR.



(d) Study2 $\Delta_{peak-to-peak}$ feature computed from the minmax normalized and filtered GSR.

Figure 6.27: Normalization and standardisation outcomes comparing Study1 and Study2.

Such a comparison serves as a qualitative validation of outcomes found in previous chapters. Looking at Figure 6.27 the same trend is observable between Study1 and Study2; selective users present a higher elevation of *post* features compared to omni-comprehensive user interface. Moreover, such figure provides also a comparison between two different data pre-processing to accounts for the intrinsic inter-individual differences into the GSR biosignal. Both data preparation gives comparable results, for its low variability, the z-score standardised GSR has been considered for all the analysis presented in this thesis.

6.3 HR Outcomes

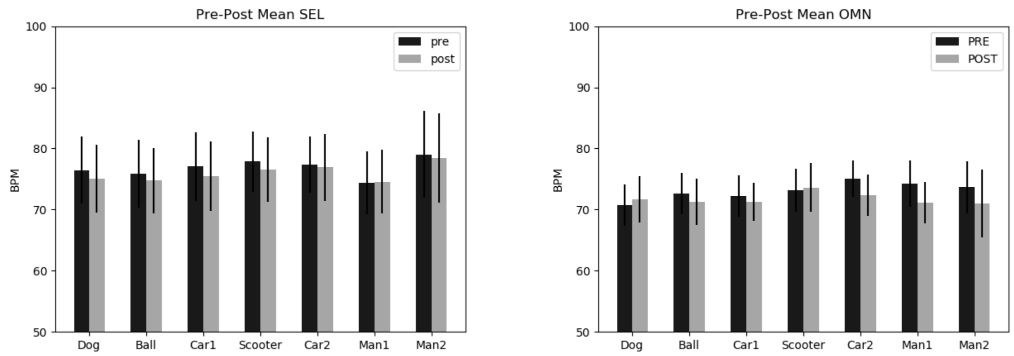
In the current section the result obtained from the analysis of the heart rate physiological signal are reported. Remembering that a similar analysis to Johnson et al. 2011 [47] has been performed, it was evaluated of the mean heart rate under test event condition. Such analysis is performed separately for the user interfaces.

Additionally, frequency analysis results are reported for each user interface.

6.3.1 Time Domain

Results obtained averaging the HRV data 15 seconds before and post the test situation for the selective HUD are presented in Figure 6.28a, while the mean heart rate outcome for the omni-comprehensive user interface is reported in Figure 6.28b.

As can be notable from such figures, no significant mean heart rate elevation is present.



(a) Histogram which shows the HRV mean value of the 15s before (black) and after(gray) the test situation. Selective HUD. No significant main effects were found.

(b) Histogram which shows the HRV mean value of the 15s before (black) and after(gray) the test situation. Omni-comprehensive HUD. No significant main effects were found.

Figure 6.28: Study2 outcomes. Separated mean heart rate during *pre-post* event, time interval of 15s. Means and standard errors are displayed.

6.3.2 Frequency Domain

In the following subsection, frequency domain analysis results are reported. Figure 6.29 shows the comparison between the selective and omni-comprehensive HUD, considering the Delta difference between the LF/HF ratio computed in the 50s before and 50s after the test event. Consequently, separates results are reported for the user interface observing directly at the 'pre' and 'post' LF/HF values.

Results obtained show no statistically significant difference in LF/HF ratio in correspondence of events. However, it is observable that Delta ratio in the selective HUD is generally higher than the omni-comprehensive user interface, in particular for 'Car1' and 'Car2' events for which an evident difference between the two HUD is notable. A no statistical significance can be due to the high variance within the selective sample which lead to obtain p-values greater than 0.05.

Instead, within the selective user interface, Figure 6.30a, in three events an elevation of the LF/HF ratio is observable between *pre* and *post*. Greater values in *post* are present in 'Dog', 'Car1' and 'Car2' event. Minor unexpected events, i.e. 'Scooter' and 'Man1', do not promote LF/HF elevation. However, differences between pre and post values are not statistically significant.

On the other side, within the omni-comprehensive HUD, Figure 6.30b, in four events an elevation of the LF/HF ratio is observable between *pre* and *post*. Greater values in *post* are present in 'Dog', 'Ball', 'Scooter' and 'Car2' event. Only the pre-post difference in the 'Dog' event is statistically significant (paired t-test analysis).

It has to be noted that the last test event, named 'Man2' is excluded from such analysis because the event is located at about 30 s from the end of the simulation; an HRV spectrum analysis of 30s time-interval is not that much reliable [79].

Two-way conducted ANOVA with repeated mesures did not report significant main effect between and within subjects, Table 6.30. Paired t-test conducted on *pre-post* evaluation of mean heart rate and LF/HF, also in this case no significant effects were found.

Selective vs Omni-comprehensive HUD

Two-way RM ANOVA LF/HF	F statistic	p-value	
Within-Subjects Effects (Event)	F(2,33)=.32	p=.773	
Between-Subjects Effects (HUD)	F(1,14)=.50	p=.492	Study2
Event*HUD	F(3,42)=.571	p=.638	

Table 6.30: Two-way ANOVA with repeated measures conducted on the test events. Assumptions were tested, Mauchly's Test of Sphericity reported a significant p-value, therefore the Greenhouse-Geisser correction was considered.

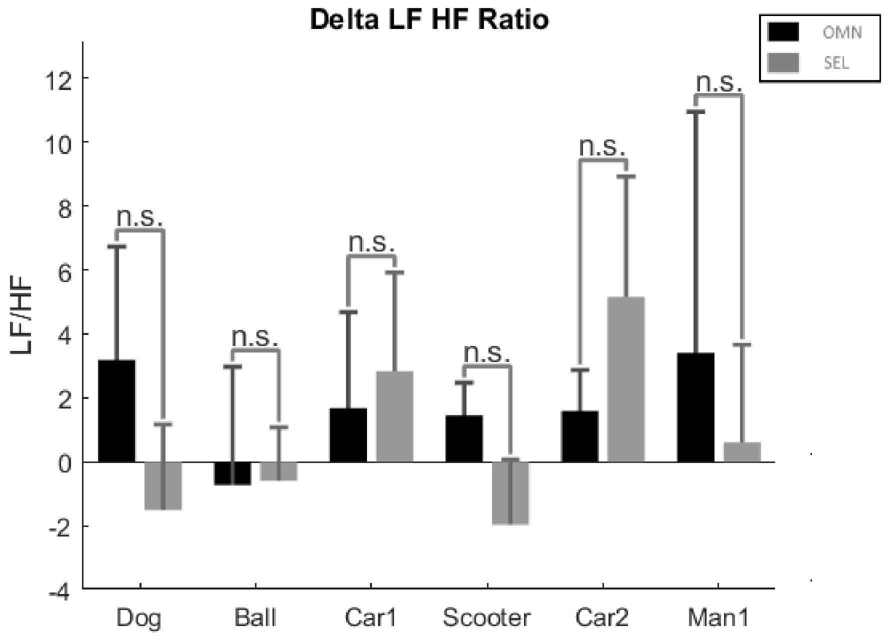
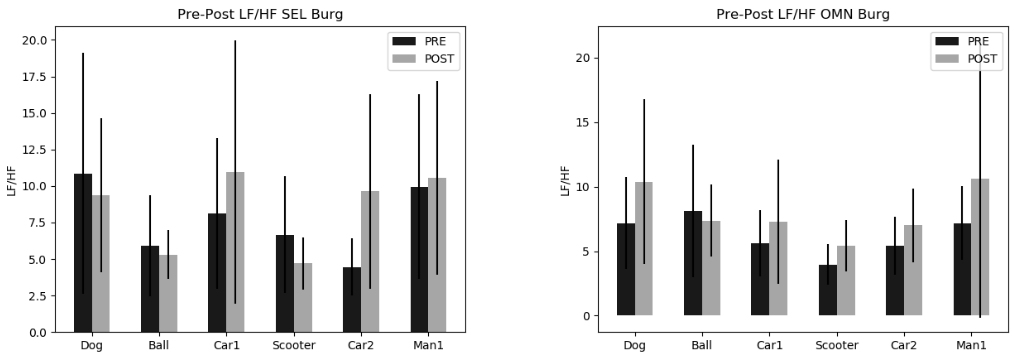


Figure 6.29: Study2 outcome. Histogram which shows the delta LF/HF ratio between pre and post the test event for selective user interface (grey) and omni-comprehensive user interface (black). The segment epoch is set to 50s. Means and standard errors are displayed. No significant main effects were found.



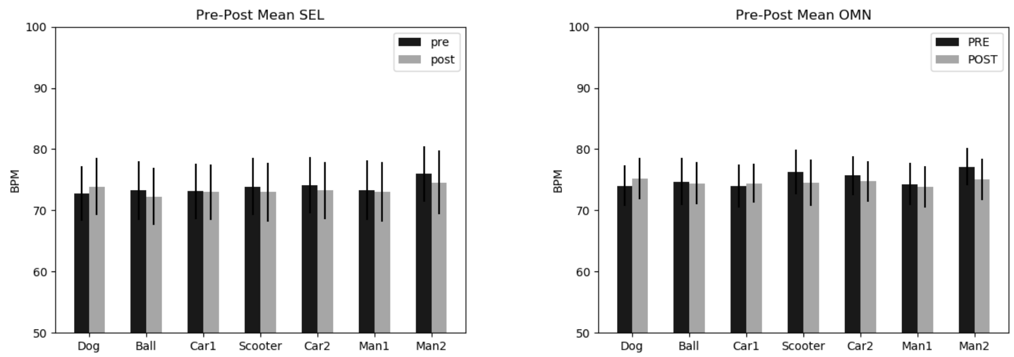
(a) Histogram which shows the LF/HF ratio *pre* and *post* the test event for the selective HUD. The segment epoch is set to 50s. No significant main effects were found.

(b) Histogram which shows the LF/HF ratio *pre* and *post* the test event for omni-comprehensive HUD. The segment epoch is set to 50s. No significant main effects were found.

Figure 6.30: Study2 outcomes. Separated LF/HF ratio analysis during *pre-post* event, time interval of 50s. Means and standard errors are displayed.

6.3.3 Study1 HR Outcomes

Also the *Study1* results are reported to qualitative compare the outcomes with the *Study2*. No significant effects are reported.



(a) Histogram which shows the HRV mean value of the 15s before (black) and after(gray) the test situation. Selective HUD. No significant main effects were found.

(b) Histogram which shows the HRV mean value of the 15s before (black) and after(gray) the test situation. Omni-comprehensive HUD. No significant main effects were found.

Figure 6.31: Study1 outcomes. Separated mean heart rate during *pre-post* event, time interval of 15s. Means and standard errors are displayed.

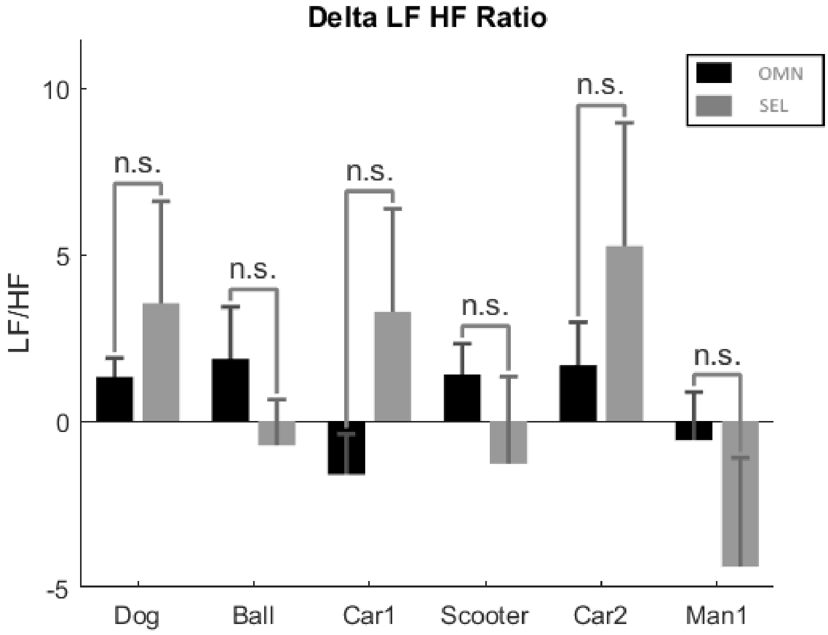
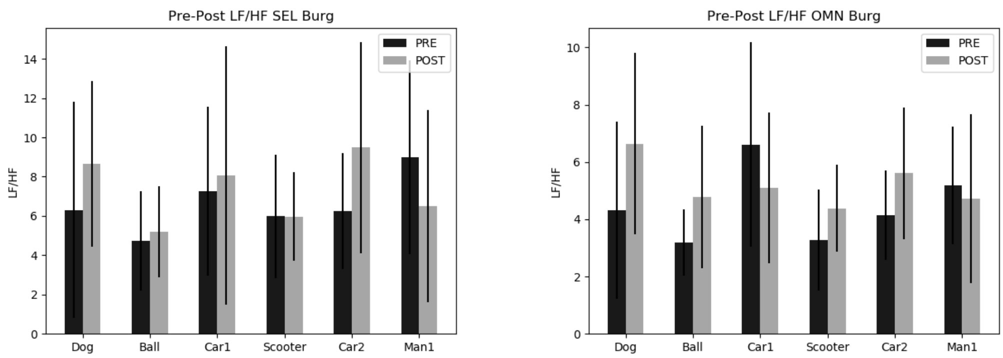


Figure 6.32: Study1 outcome. Histogram which shows the delta LF/HF ratio between pre and post the test event for selective user interface (grey) and omni-comprehensive user interface (black). The segment epoch is set to 50s. Means and standard errors are displayed. No significant main effects were found.



(a) Histogram which shows the LF/HF ratio *pre* and *post* the test event for the selective HUD. The segment epoch is set to 50s. No significant main effects were found.

(b) Histogram which shows the LF/HF ratio *pre* and *post* the test event for omni-comprehensive HUD. The segment epoch is set to 50s. No significant main effects were found.

Figure 6.33: Study1 outcomes. Separated LF/HF ratio analysis during *pre-post* event, time interval of 50s. Means and standard errors are displayed.

Chapter 7

Conclusions & Future Developments

7.1 Conclusions

Over the space of six months, the main aim of this thesis project has been achieved. The study focused on the evaluation of which specific user interface helps significantly to build trust in a fully automated vehicle.

Therefore, to this end, literature research was conducted to study the state of art of the autonomous sector, principally concerning the automotive field. Nowadays, six SAE classification levels exist; starting from level 0 which considers no automotion and ending at level 5 that accounts the fully automated vehicles in which the system wholly replaces the driver.

Moreover, the literature research also has dealt with the analysis of the metrics to evaluate the user-experience during the driving task. Principally two kinds of measurements have been acknowledged: the compilation of a *pre-post* questionnaire, which accounts for the subjective evaluation of specific and general situations during the driving, and objective measurements represented by physiological signals, i.e. *heart rate variability* and *electrodermal activity*. These physiological measures have been established to be valid indicators of the sympathetic nervous system activity since it is highly correlated with the affective states. Such measures can serve as real-time marker of the emotional state [47][57].

However, the available automotive market does not provide fully automated vehicles; therefore to reproduce and test the user-experience within an entirely controlled environment an already developed immersive virtual driving simulator has been employed with integrated autonomous driving modality,

An unresolved issue of this upcoming technology is the social dilemma related to automation in daily-life. American researches report that a slight majority of Americans would not ride a self-driving car if given the chance, lack of trust is the main concern [9]. The study by Ekman et al.[12] stated that *human-machine interaction* has an essential role in supporting such forthcoming technology to build confidence and trust by supplying situational awareness information to the users. To this end, the research is very active in the use of AR-HUD (Augmented Reality Head Up Display) to act as a link between subject and machine, since this technology can provide a support enhancing the user-experience in automated vehicles [21].

The works focused on the analysis of two user interface: selective and omni-comprehensive, the first refers to an HMI which displays a subset of information (in particular the ones which affects the actual driving), while the latter refers to a more informative user interface which provides all the information about objects which are within a detection range of 150 meters.

The virtual driving simulation provided the reproduction of an urban scenario, i.e. San Francisco, with scheduled events. More in details, seven test events have been pre-programmed to occur, namely *Dog*, *Ball*, *Car1*, *Scooter*, *Car2*, *Man1*, *Man2*. For these test events, a potential hazardous driving situation has been simulated.

Objective results reported that HUD has a significant main effect on the EDA response, more specifically looking at $\Delta_{peak-to-peak}$, Δ_{max}, $\Delta_{accumulated}$ and Δ_{mean} features ($p < .05$). Events had a significant main effect on the EDA response ($p < .000$). This last finding is also confirmed from the investigation performed within the user interface; four *pre-post* features where studied within the HUD and for *Dog*, *Ball*, *Car1*, *Car2* and *Man2* test events the paired t-test analysis reported that these test events had a significant main effect on the four features ($p < .05$). The trend separately analysed for each HUD of the minmax normalised EDA qualitatively confirms these outcomes.

Concerning the heart rate variability physiologic signal, neither for the mean heart rate nor for the spectral information about the LF/HF ratio the user interface has reported a significant main effect. A main possible reason can be distinguished: the limit of the available biosensor: it gave only digital output when the heartbeat was received; therefore almost none post-processing analysis can be performed. It would have been better to have the raw ECG signal to extract more accurately biosignal characteristics for reliability concerns.

Moreover, to obtain other confirmations about the user-experience, the analysis of the subjective responses collected from the questionnaire has been carried out. From this analysis it is possible to understand why EDA had a significantly different behaviour under test events circumstances. Section of the questionnaire which evaluates the test events and the section which accounts the overall assessments of the user-experience are particularly meaningful for these purposes. The omni-comprehensive sample significantly noticed the potential danger previously to influence the current driving ($p < .001$). Therefore the HUD was able to put the user on alert about that potential danger since the selective user interface presents information only once the object influence the actual driving. The significant main effect that user interface had on the question which accounts the evaluation of the utility of HUD information to build trust represents a further relevant finding ($p < .001$). Moreover, users who faced the omni-comprehensive HUD where significantly more in agreement to have a real-road driving experience on a fully automated vehicle. However, the omni-comprehensive HUD was assessed to be slightly too excessive in term of quantity of information compared to the selective user interface ($p < .001$).

Additionally, to validate the findings of the EDA continuous and objective measures and subjective measurements collected in the questionnaire, a classification model was built separately for each HUD and each test events. Classificator models take as predictors the continuous measurement and the output is the predicted questionnaire response of the question which accounts the perceived risk/fear during the test events. Both interfaces in five test events have reported that predictors had a significant main effect on the goodness of fit of the classificator ($p < .05$). Such a result is considered to be a measure of correlation between the questionnaire responses and EDA outcomes.

In conclusion, users evaluate as slightly excessive all the information presented by the omni-comprehensive user interface. However, such information is considered to be useful to enhance users situational awareness and to build trust in the fully automated vehicles.

7.2 Future Work

The first main thought which comes up about the future work is to derive the heart rate variability differently, i.e. using the raw ECG signal. Such a signal enables the evaluation of more reliable features thanks to the possibility of a more accurate post-processing. The lack of HR significant findings is probably due to the limitation provided by the available heart rate sensor; it could be overcome through recording electrodes placed on the skin surface around the thoracic area in order to derive the RR interval more accurately.

Concerning the driving simulation, it could also be interesting to evaluate different scenarios, like for instance a non-urban environment, and evaluate through objective measurements how the users' affective states change while passing from a non-urban to an urban scenario or vice-versa.

Finally, another idea is to study the ecological validation of the immersive driving simulator through these physiological measurements. Such a validation needs a one-to-one comparison between simulated and on-road experience. Generally, the primary approach is to reproduce in the virtual environment the same path followed in the on-road experience. Therefore a specific track is required for all the subjects as well as the simulated reproduction of the real-road experience. Additionally, to evaluate the user response, two approaches are possible: an event-related analysis; therefore assessing specific situations during the on-road test session and simulated reproduction looking at the physiological response or the study of the mean trend that physiological markers have both in simulated and real experience.

Appendix A

Questionario di valutazione dell'esperienza utente

A.1 Domande Personali

1. Età
2. Sesso

Quanto spesso utilizzi/hai utilizzato:

3. Strumenti per la realtà virtuale immersiva(ad es. Oculus Rift, HTC Vive ecc.)? (mai, raramente, qualche volta, spesso, ogni giorno)
4. Simulatori di guida (ad es. Assetto Corsa, rFactor ecc.)? (mai, raramente, qualche volta, spesso, ogni giorno)
5. Sarei disposto a partecipare ad un'esperienza di guida autonoma a bordo di un' auto reale (decisamente in disaccordo, in disaccordo, non saprei, d'accordo, decisamente decisamente d'accordo)

A.2 Autovalutazione dello stato di salute pre-simulazione

Quanto sei affetto dai seguenti sintomi in questo momento? (per nulla, lievemente, moderatamente, intensamente)

6. Malessere in generale
7. Affaticamento
8. Mal di testa
9. Occhi affaticati

10. Dffcoltà di messa a fuoco
11. Salivazione aumentata
12. Sudorazione
13. Nausea
14. Diffoltà di concentrazione
15. Visione sfocata
16. Capogiro con occhi aperti
17. Capogiro con occhi chiusi
18. Vertigini
19. Fastidio allo stomaco

A.3 Autovalutazione dello stato di salute post-simulazione

Quanto sei affetto dai seguenti sintomi in questo momento? (per nulla, lievemente, moderatamente, intensamente)

20. Malessere in generale
21. Affaticamento
22. Mal di testa
23. Occhi affaticati
24. Dffcoltà di messa a fuoco
25. Salivazione aumentata
26. Sudorazione
27. Nausea
28. Diffoltà di concentrazione
29. Visione sfocata
30. Capogiro con occhi aperti
31. Capogiro con occhi chiusi
32. Vertigini
33. Fastidio allo stomaco

A.4 Valutazione della simulazione di guida autonoma

(completamente in disaccordo, in disaccordo, né in disaccordo né d'accordo, d'accordo, completamente d'accordo)

34. L'auto autonoma ha mostrato delle adeguate capacita decisionali
35. L'auto autonoma e andata in difficoltà in caso di cambiamenti inaspettati dell'ambiente
36. L'auto autonoma mi è sembrata intelligente
37. In generale, valuteresti le capacità di guida autonoma dell'auto nell'esperienza come: per nulla soddisfacente, poco soddisfacente, mediamente soddisfacente, soddisfacente, molto soddisfacente?

A.5 Situazioni di test

Durante l'esperienza in Realtà Virtuale si sono verificate alcune situazioni di guida 'particolari'.

A.5.1 Situazione #1: Cane che attraversa la strada

38. Quanto hai percepito pericolosa/spaventosa le seguente situazione durante la simulazione? (decisamente non pericolosa - decisamente pericolosa)
39. La situazione mi ha colto di sorpresa (completamente in disaccordo - completamente d'accordo)
40. Sono riuscito a vedere il potenziale pericolo prima che influenzasse il comportamento del veicolo (es. rallentare/frenare) (completamente in disaccordo - completamente d'accordo)
41. L'interfaccia mi ha fornito informazioni utili a prevedere il pericolo (completamente in disaccordo - completamente d'accordo)
42. Se d'accordo, quali?

A.5.2 Situazione #2: Palla lanciata da un bambino

43. Quanto hai percepito pericolosa/spaventosa le seguente situazione durante la simulazione? (decisamente non pericolosa - decisamente pericolosa)

44. La situazione mi ha colto di sorpresa (completamente in disaccordo - completamente d'accordo)

45. Sono riuscito a vedere il potenziale pericolo prima che influenzasse il comportamento del veicolo (es. rallentare/frenare) (completamente in disaccordo - completamente d'accordo)

46. L'interfaccia mi ha fornito informazioni utili a prevedere il pericolo (completamente in disaccordo - completamente d'accordo)

47. Se d'accordo, quali?

A.5.3 Situazione #3: Macchina che si si inserisce da destra causa corsia bloccata

48. Quanto hai percepito pericolosa/spaventosa le seguente situazione durante la simulazione? (decisamente non pericolosa - decisamente pericolosa)

49. La situazione mi ha colto di sorpresa (completamente in disaccordo - completamente d'accordo)

50. Sono riuscito a vedere il potenziale pericolo prima che influenzasse il comportamento del veicolo (es. rallentare/frenare) (completamente in disaccordo - completamente d'accordo)

51. L'interfaccia mi ha fornito informazioni utili a prevedere il pericolo (completamente in disaccordo - completamente d'accordo)

52. Se d'accordo, quali?

A.5.4 Situazione #4: Il sorpasso dello scooter

53. Quanto hai percepito pericolosa/spaventosa le seguente situazione durante la simulazione? (decisamente non pericolosa - decisamente pericolosa)

54. La situazione mi ha colto di sorpresa (completamente in disaccordo - completamente d'accordo)

55. Sono riuscito a vedere il potenziale pericolo prima che influenzasse il comportamento del veicolo (es. rallentare/frenare) (completamente in disaccordo - completamente d'accordo)

56. L'interfaccia mi ha fornito informazioni utili a prevedere il pericolo (completamente in disaccordo - completamente d'accordo)

57. Se d'accordo, quali?

A.5.5 Situazione #5: Macchina che taglia la strada

58. Quanto hai percepito pericolosa/spaventosa le seguente situazione durante la simulazione? (decisamente non pericolosa - decisamente pericolosa)

59. La situazione mi ha colto di sorpresa (completamente in disaccordo - completamente d'accordo)

60. Sono riuscito a vedere il potenziale pericolo prima che influenzasse il comportamento del veicolo (es. rallentare/frenare) (completamente in disaccordo - completamente d'accordo)

61. L'interfaccia mi ha fornito informazioni utili a prevedere il pericolo (completamente in disaccordo - completamente d'accordo)

62. Se d'accordo, quali?

A.5.6 Situazione #6: Pedone che attraversa fuori dalle strisce

63. Quanto hai percepito pericolosa/spaventosa le seguente situazione durante la simulazione? (decisamente non pericolosa - decisamente pericolosa)

64. La situazione mi ha colto di sorpresa (completamente in disaccordo - completamente d'accordo)

65. Sono riuscito a vedere il potenziale pericolo prima che influenzasse il comportamento del veicolo (es. rallentare/frenare) (completamente in disaccordo - completamente d'accordo)

66. L'interfaccia mi ha fornito informazioni utili a prevedere il pericolo (completamente in disaccordo - completamente d'accordo)

67. Se d'accordo, quali?

A.5.7 Situazione #7: Pedone che attraversa in maniera inaspettata

68. Quanto hai percepito pericolosa/spaventosa le seguente situazione durante la simulazione? (decisamente non pericolosa - decisamente pericolosa)

69. La situazione mi ha colto di sorpresa (completamente in disaccordo - completamente d'accordo)

70. Sono riuscito a vedere il potenziale pericolo prima che influenzasse il comportamento del veicolo (es. rallentare/frenare) (completamente in disaccordo - completamente d'accordo)

71. L'interfaccia mi ha fornito informazioni utili a prevedere il pericolo (completamente in disaccordo - completamente d'accordo)

72. Se d'accordo, quali?

A.6 Valutazione dell'interfaccia HUD ai fini della consapevolezza del contesto

In questa sezione vogliamo valutare in generale quanto l'interfaccia HUD sia stata efficace nell'aiutarti a comprendere quanto stava accadendo. (decisamente in disaccordo, in disaccordo, non saprei, d'accordo, decisamente decisamente d'accordo)

73. I "bounding box" mi hanno aiutato a capire che il veicolo aveva "preso in carico" il semaforo e capito come gestirlo

74. Le etichette (lampeggianti) hanno aiutato a capire che il veicolo aveva "preso in carico" il semaforo e capito come gestirlo

75. I "bounding box" mi hanno aiutato a capire che il veicolo aveva "preso in carico" il segnale stradale e capito come gestirlo

76. Le etichette (lampeggianti) mi hanno aiutato a capire che il veicolo aveva "preso in carico" il segnale stradale e capito come gestirlo

77. I "bounding box" mi hanno aiutato a capire che il veicolo aveva "preso in carico" il potenziale ostacolo (pedone, animale, ecc.) e capito come gestirlo

78. Le etichette mi hanno aiutato a capire che il veicolo aveva "preso in carico" il potenziale ostacolo (pedone, animale, ecc.) e capito come gestirlo

79. I "bounding box" mi hanno aiutato a capire che il veicolo aveva "preso in carico" le altre auto potenzialmente condizionanti la guida e capito come gestirle

80. Le etichette mi hanno aiutato a capire che il veicolo aveva "preso in carico" le altre auto potenzialmente condizionanti la guida e capito come gestirle

81. La linea di pianificazione del percorso della tua auto è stata di aiuto alla comprensione delle intenzioni del veicolo

82. Le linee di pianificazione del percorso delle altre auto del traffico mi hanno aiutato a capire che il veicolo aveva ”preso in carico” la loro presenza e capito come gestirle

83. La colorazione dei bounding box riferite al grado di pericolosità di potenziali ostacoli mi hanno aiutato a capire che il veicolo li aveva ”presi in carico” e capito come gestirli

84. L’avviso sonoro di presa in carico di un semaforo/segnale stradale mi ha aiutato a capire che il veicolo aveva ”preso in carico” la situazione

85. L’ avviso sonoro in caso di pericolo mi ha aiutato a capire che il veicolo aveva ”preso in carico” la situazione e capito come gestirla

A.7 Valutazione dell’interfaccia HUD - Quantità informazioni

Il numero di informazioni mostrate dal sistema HUD per tale informazione è risultato in (1 = Quantità insufficiente, 3 = Quantità adeguata, 5 = Quantità eccessiva):

86. Bounding box ed etichette per i semafori

87. Bounding box ed etichette per i segnali stradali

88. Bounding box ed etichette per eventuali ostacoli (pedoni, animali ecc.)

89. Bounding box ed etichette per le altre auto del traffico

90. Linee di pianificazione del percorso delle auto

91. Avviso sonoro di rilevazione di un semaforo/segnale stradale

92. Avviso sonoro in caso di pericolo

A.8 Carico cognitivo

In questa sezione vogliamo valutare il carico cognitivo dell’interfaccia HUD. (decisamente in disaccordo, in disaccordo, non saprei, d’accordo, decisamente decisamente d’accordo)

93. Ho trovato faticoso (impegnativo) individuare le informazioni nell’interfaccia HUD

94. Se d’accordo, perché?

95. Ho trovato frustrante (stressante, irritante) individuare le informazioni nell'interfaccia HUD

96. Se d'accordo, perché?

97. In generale il numero di informazioni mostrate dall'interfaccia HUD mi è sembrato eccessivo

98. In generale la comprensibilità / la qualità delle informazioni mostrate dall'interfaccia HUD mi è sembrata adeguata

A.9 Domande sull'intera esperienza

(decisamente in disaccordo, in disaccordo, non saprei, d'accordo, decisamente decisamente d'accordo)

99. Le informazioni mostrate dall'interfaccia HUD sono state d'aiuto nello stabilire la fiducia nell'auto

100. Durante la simulazione, le informazioni mostrate dall'interfaccia HUD mi hanno aiutato a capire perché l'auto stesse effettuando una determinata azione

101. Durante la simulazione, le informazioni fornite dall'interfaccia HUD mi hanno aiutato a farmi sentire tranquillo ed a mio agio

102. Durante la simulazione, grazie alle informazioni mostrate dall'interfaccia HUD ho avuto la percezione che l'auto avesse il pieno controllo della situazione

103. L'interfaccia HUD è riuscita ad informarmi correttamente prima che il potenziale pericolo influenzasse la guida

104. In generale, l'interfaccia HUD mi ha fornito informazioni utili a prevedere le situazioni di pericolo

105. Dopo questa esperienza, sarei disposto a partecipare ad un'esperienza di viaggio in guida totalmente autonoma a bordo di un'auto reale

A.10 Senso di immersione e presenza

(completamente in disaccordo, in disaccordo, né in disaccordo né d'accordo, d'accordo, completamente d'accordo)

106. Le informazioni sensoriali fornite dall'applicazione di Realtà Virtuale, e dalle tecnologie utilizzate, mi hanno fatto sentire di essere immerso nell'ambiente virtuale (di trovarmi in un luogo diverso da quello in cui mi trovo fisicamente)

107. La qualità della scena tridimensionale ha ridotto il mio senso di immersione e presenza nel mondo virtuale

108. Sono riuscito a percepire correttamente la scala, le proporzioni e le dimensioni degli ambienti/oggetti nel mondo virtuale

109. Vedere le mie mani, e le gambe, mi ha aiutato a sentirmi presente nell'ambiente virtuale

110. I movimenti della piattaforma inerziale mi hanno aiutato a sentirmi presente nell'ambiente virtuale

111. In generale valuteresti il tuo senso di immersione e presenza come: per nulla soddisfacente (1), poco soddisfacente (2), mediamente soddisfacente (3), soddisfacente (4), molto soddisfacente (5)?

A.11 Fedeltà della simulazione

(completamente in disaccordo, in disaccordo, né in disaccordo né d'accordo, d'accordo, completamente d'accordo)

112. Ho trovato la simulazione accurata

113. Gli oggetti nell'ambiente virtuale si sono mossi in modo naturale

114. La simulazione sembrava bloccarsi o fermarsi a tratti

115. La percezione che ho avuto degli oggetti virtuali è stata realistica

116. L'esperienza nel mondo virtuale mi e sembrata coerente con quella che avrei potuto vivere nel mondo reale

117. Il movimento della piattaforma inerziale è stato realistico

118. In generale valuteresti la fedeltà della simulazione come: per nulla soddisfacente, poco soddisfacente, mediamente soddisfacente, soddisfacente, molto soddisfacente?

Bibliography

[1] Raspberry adc. https://fortronic.it/auto-connesse-innovazione-digitale/, 2018.

[2] Muhammad Awais, Nasreen Badruddin, and Micheal Drieberg. A hybrid approach to detect driver drowsiness utilizing physiological signals to improve system performance and wearability. *Sensors*, 17(9):1991, 2017.

[3] Jennifer Healey, Rosalind W Picard, et al. Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on intelligent transportation systems*, 6(2):156–166, 2005.

[4] B. C. Zanchin, R. Adamshuk, M. M. Santos, and K. S. Collazos. On the instrumentation and classification of autonomous cars. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 2631–2636, Oct 2017.

[5] Fabian Kröger. *Automated Driving in Its Social, Historical and Cultural Contexts*, pages 41–68. Springer Berlin Heidelberg, Berlin, Heidelberg, 2016.

[6] Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. https://www.sae.org/standards/content/j3016_201609/, 2019.

[7] Jelena Kocic, Nenad Jovičić, and Vujo Drndarevic. Sensors and sensor fusion in autonomous vehicles. 11 2018.

[8] *Self driving safety report*. General Motors, 2018.

[9] Aaron Smith and Monica Anderson. Americans attitudes toward driverless vehicles. http://www.pewinternet.org/2017/10/04/americans-attitudes-toward-driverless-vehicles/, 2017.

[10] iQ Intel Italy. https://iq.intel.it/fiducia-e-guida-autonoma-come-ottenere-la-fiducia-delle-persone-trasportate/, 2018.

[11] J. Weast. La fiducia umana nell'auto robot. http://blog.ilgiornale.it/lombardo/2017/08/25/la-fiducia-umana-nella-auto-robot/.http://blog.ilgiornale.it/lombardo/2017/08/25/la-fiducia-umana-nella-auto-robot/., journal=Blog.ilgiornale.it, year=2017.

[12] Fredrick Ekman, Mikael Johansson, and Jana Sochor. Creating appropriate trust in automated vehicle systems: A framework for hmi design. *IEEE Transactions on Human-Machine Systems*, 48(1):95–101, 2018.

[13] Mica R Endsley. Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, 42(3):462–492, 1999.

[14] S Debernard, C Chauvin, R Pokam, and Sabine Langlois. Designing human-machine interface for autonomous vehicles. *IFAC-PapersOnLine*, 49(19):609–614, 2016.

[15] Ola Benderius, Christian Berger, and Victor Malmsten Lundgren. The best rated human–machine interface design for autonomous vehicles in the 2016 grand cooperative driving challenge. *IEEE Transactions on Intelligent Transportation Systems*, 19(4):1302–1307, 2018.

[16] Richie Jose, Gun A Lee, Mark Billinghurst, and OzCHI 2016 28th Australian Computer-Human Interaction Conference. A comparative study of simulated augmented reality displays for vehicle navigation. 2016.

[17] Jake Holmes. Every car infotainment system available in 2018. https://www.cnet.com/roadshow/news/car-infotainment-system-automotive-tech-guide-2018/, 2018.

[18] Nithya Palaniyappan. Continental aims to put augmented reality on your car windshield — digit.in. https://www.digit.in/car-tech/continental-aims-to-put-augmented-reality-on-your-car-windshield-29716.html, 2018.

[19] Anup Doshi, Shinko Yuanhsien Cheng, and Mohan Manubhai Trivedi. A novel active heads-up display for driver assistance. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 39(1):85–93, 2009.

[20] Zeljko Medenica, Andrew L Kun, Tim Paek, and Oskar Palinko. Augmented reality vs. street views: a driving simulator study comparing two emerging navigation aids. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, pages 265–274. ACM, 2011.

[21] Lotfi Abdi and Aref Meddeb. Driver information system: a combination of augmented reality and deep learning. In *Proceedings of the Symposium on Applied Computing*, pages 228–230. ACM, 2017.

[22] Renate Haeuslschmid, Laura Schnurr, Julie Wagner, and Andreas Butz. Contact-analog warnings on windshield displays promote monitoring the road scene. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 64–71. ACM, 2015.

[23] Minh Tien Phan, Indira Thouvenin, and Vincent Fremont. Enhancing the driver awareness of pedestrian using augmented reality cues. In *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pages 1298–1304. IEEE, 2016.

[24] UBI Mobility. How augmented reality can help us accept autonomous cars. *printed from http://www. ubimobility. org/how-augmented-reality-can-help-us-accept-autonomous-cars/on Apr*, 19, 2017.

[25] Laurino Antonello. Strumenti di simulazione in realtà virtuale per la valutazione dell'esperienza utente in veicoli a guida autonoma. 2018.

[26] how artificial intelligence can create a real world simulation for autonomous driving 2018. https://www.forbes.com/sites/jenniferhicks/2018/07/10/how-artificial-intelligence-can-create-a-real-world-simulation-for-autonomous-driving/#471389d42884, 2018.

[27] test di sicurezza virtuale per auto a guida autonoma da nvidia drive constellation 2018. https://www.nvidia.com/it-it/self-driving-cars/drive-constellation/, 2018.

[28] CARLA Team. Carla. http://carla.org/, 2018.

[29] genivi alliance 2018. https://www.genivi.org/, 2018.

[30] Gerard J Blaauw. Driving experience and task demands in simulator and instrumented car: a validation study. *Human Factors*, 24(4):473–486, 1982.

[31] Samuel Mudd. Assessment of the fidelity of dynamic flight simulators. *Human Factors*, 10(4):351–358, 1968.

[32] Ernest J. (Ernest James) McCormick and Human engineering McCormick, Ernest J. (Ernest James). *Human factors engineering*. New York ; [Maidenhead] : McGraw-Hill, 3rd ed edition, 1970. First ed. published in 1957 under title: Human engineering; 4th ed. under title: Human factors in engineering.

[33] Evi Blana. Driving simulator validation studies: A literature review. 1996.

[34] H Alm. Driving simulators as research tools: A validation study based on the vti driving simulator. 1996.

[35] David Shinar and Adi Ronen. Validation of speed perception and production in a single screen simulator. *Advances in transportation studies*, (SI 1):51–56, 2007.

[36] Erwin R Boer. Experiencing the same road twice: A driver centered comparison between simulation and reality. In *DSC2000 (Driving Simulation Conference), Paris, France*, 2000.

[37] Alvah Bittner, Ozgur Simsek Jr, William Levison, and John Campbell. On-road versus simulator data in driver model development driver performance model experience. *Transportation Research Record: Journal of the Transportation Research Board*, (1803):38–44, 2002.

[38] Francesco Bella. Driving simulator for speed research on two-lane rural roads. *Accident; analysis and prevention*, 40(3):1078—1087, May 2008.

[39] Adi Ronen and David Shinar. Validation of speed perception and production in a single screen simulator. *Advances in transportation studies*, (SI 1):51–56, 2007.

[40] Nadia Mullen, Judith Charlton, Anna Devlin, and Michel Bedard. *Simulator validity: behaviours observed on the simulator and on the road*, pages 1 – 18. CRC Press, Australia, 1st edition, 2011.

[41] Nazan Aksan, Sarah D Hacker, Lauren Sager, Jeffrey Dawson, Steven Anderson, and Matthew Rizzo. Correspondence between simulator and on-road drive performance: implications for assessment of driving safety. *geriatrics*, 1(1):8, 2016.

[42] Francesco Bella. Validation of a driving simulator for work zone design. *Transportation Research Record: Journal of the Transportation Research Board*, (1937):136–144, 2005.

[43] Francesco Bella. Driving simulator for speed research on two-lane rural roads. *Accident Analysis & Prevention*, 40(3):1078–1087, 2008.

[44] Stuart T Godley, Thomas J Triggs, and Brian N Fildes. Driving simulator validation for speed research. *Accident analysis & prevention*, 34(5):589–600, 2002.

[45] S. Helman and N. Reed. Validation of the driver behaviour questionnaire using behavioural data from an instrumented vehicle and high-fidelity driving simulator. *Accident Analysis & Prevention*, 75:245 – 251, 2015.

[46] Joshua Hoffman, John Lee, Timothy Brown, and Daniel McGehee. Comparison of driver braking responses in a high-fidelity simulator and on a test track. *Transportation Research Record: Journal of the Transportation Research Board*, (1803):59–65, 2002.

[47] Michel J Johnson, Tammem Chahal, Arne Stinchcombe, Nadia Mullen, Bruce Weaver, and Michel Bedard. Physiological responses to simulated and on-road driving. *International journal of Psychophysiology*, 81(3):203–208, 2011.

[48] Hoe Lee, Torbjorn Falkmer, Lorna Rosenwax, Richard Cordell, Andrew Granger, Barry Vieira, and Alfred Lee. Validity of driving simulator in assessing drivers with parkinson's disease. *Advances in transportation studies*, (SI 1):81–90, 2007.

[49] Ying Wang, Bruce Mehler, Bryan Reimer, Vincent Lammers, Lisa A D'Ambrosio, and Joseph F Coughlin. The validity of driving simulation for assessing differences between in-vehicle informational interfaces: A comparison with field testing. *Ergonomics*, 53(3):404–420, 2010.

[50] GR Watts and AR Quimby. Design and validation of a driving simulator for use in perceptual studies. Technical report, 1979.

[51] Bryan Reimer, Lisa A D'Ambrosio, Joseph F Coughlin, Michael E Kafrissen, and Joseph Biederman. Using self-reported data to assess the validity of driving simulation data. *Behavior research methods*, 38(2):314–324, 2006.

[52] Dianne Parker, James T Reason, Antony SR Manstead, and Stephen G Stradling. Driving errors, driving violations and accident involvement. *Ergonomics*, 38(5):1036–1048, 1995.

[53] Matthew Lombard and Theresa Ditton. At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, 3(2), 1997.

[54] Christophe Deniaud, Vincent Honnet, Benoit Jeanne, and Daniel Mestre. The concept of "presence" as a measure of ecological validity in driving simulators. *Journal of Interaction Science*, 3(1):1, 2015.

[55] Jan Törnros. Driving behaviour in a real and a simulated road tunnel—a validation study. *Accident Analysis & Prevention*, 30(4):497–503, 1998.

[56] Andras Kemeny and Francesco Panerai. Evaluating perception in driving simulation experiments. *Trends in cognitive sciences*, 7(1):31–37, 2003.

[57] Luis Eudave and Miguel Valencia. Physiological response while driving in an immersive virtual environment. In *Wearable and Implantable Body*

Sensor Networks (BSN), 2017 IEEE 14th International Conference on, pages 145–148. IEEE, 2017.

[58] Shinji Kajiwara. Evaluation of driver's mental workload by facial temperature and electrodermal activity under simulated driving conditions. *International Journal of Automotive Technology*, 15(1):65–70, 2014.

[59] Krystyna Zużewicz, Danuta Roman-Liu, Maria Konarska, Paweł Bartuzi, Krzysztof Matusiak, Dariusz Korczak, Zbigniew Lozia, and Marek Guzek. Heart rate variability (hrv) and muscular system activity (emg) in cases of crash threat during simulated driving of a passenger car. *International journal of occupational medicine and environmental health*, 26(5):710–723, 2013.

[60] Yutao Ba, Wei Zhang, Qinhua Wang, Ronggang Zhou, and Changrui Ren. Crash prediction with behavioral and physiological features for advanced vehicle collision avoidance system. *Transportation Research Part C: Emerging Technologies*, 74:22–33, 2017.

[61] Yutao Ba, Wei Zhang, Gavriel Salvendy, Andy SK Cheng, and Petya Ventsislavova. Assessments of risky driving: a go/no-go simulator driving task to evaluate risky decision-making and associated behavioral patterns. *Applied ergonomics*, 52:265–274, 2016.

[62] Daniele Ruscio, Luca Bascetta, Alessandro Gabrielli, Matteo Matteucci, Dedy Ariansyah, Monica Bordegoni, Giandomenico Caruso, and Lorenzo Mussone. Collection and comparison of driver/passenger physiologic and behavioural data in simulation and on-road driving. In *Models and Technologies for Intelligent Transportation Systems (MT-ITS), 2017 5th IEEE International Conference on*, pages 403–408. IEEE, 2017.

[63] Mel Slater, Christoph Guger, Guenter Edlinger, Robert Leeb, Gert Pfurtscheller, Angus Antley, Maia Garau, Andrea Brogni, and Doron Friedman. Analysis of physiological responses to a social situation in an immersive virtual environment. *Presence: Teleoperators and Virtual Environments*, 15(5):553–569, 2006.

[64] Cristoph Guger, Guenter Edlinger, Robert Leeb, and Gert Pfurtscheller. Heart-rate variability and event-related ecg in virtual environments. 2004.

[65] Faizul Rizal Ismail, Nor Kamaliana Khamis, Mohd Zaki Nuawi, Dieter Schramm, and Benjamin Hesse. Measurement of heart rate to determine car drivers' performance impairment in simulated driving: An overview. *Jurnal Teknologi*, 78(6-9):15–23, 2016.

[66] Michael Meehan, B Insko, M Whitton, and FP Brooks. An objective surrogate for presence: Physiological response. In *3rd International Workshop on Presence*, 2000.

[67] Paul Ekman, Robert W Levenson, and Wallace V Friesen. Autonomic nervous system activity distinguishes among emotions. *Science*, 221(4616):1208–1210, 1983.

[68] Kim J Vicente, D Craig Thornton, and Neville Moray. Spectral analysis of sinus arrhythmia: A measure of mental effort. *Human factors*, 29(2):171–182, 1987.

[69] J Gordon Betts et al. Anatomy & physiology.(2013). *Open Stax College*.

[70] Mary Boudreau Conover. *Understanding electrocardiography*. Elsevier Health Sciences, 2002.

[71] Mohamed Hammad, Asmaa Maher, Kuanquan Wang, Feng Jiang, and Moussa Amrani. Detection of abnormal heart conditions based on characteristics of ecg signals. *Measurement*, 125, 05 2018.

[72] ROBIN Callister, N Omar Suwarno, and DOUGLAS R Seals. Sympathetic activity is influenced by task difficulty and stress perception during mental challenge in humans. *The Journal of physiology*, 454(1):373–387, 1992.

[73] David G Gilbert, Charles J Meliska, Richard Welser, and Steven L Estes. Depression, personality, and gender influence eeg, cortisol, beta-endorphin, heart rate, and subjective responses to smoking multiple cigarettes. *Personality and Individual Differences*, 16(2):247–264, 1994.

[74] Jason Gregoire, Stephanie Tuck, Richard L Hughson, and Yoshiharu Yamamoto. Heart rate variability at rest and exercise: influence of age, gender, and physical training. *Canadian Journal of Applied Physiology*, 21(6):455–470, 1996.

[75] U Rajendra Acharya, K Paul Joseph, Natarajan Kannathal, Choo Min Lim, and Jasjit S Suri. Heart rate variability: a review. *Medical and biological engineering and computing*, 44(12):1031–1051, 2006.

[76] https://www.datasci.com/solutions/cardiovascular/heart-rate-variability-(hrv), 2018.

[77] Fred Shaffer and JP Ginsberg. An overview of heart rate variability metrics and norms. *Frontiers in public health*, 5:258, 2017.

[78] Task Force. Standards of measurement, physiological interpretation and clinical use. task force of the european society of cardiology and the

north american society of pacing and electrophysiology. *Circulation*, 93(5):1043–1065, 1996.

[79] Lizawati Salahuddin, Jaegeol Cho, Myeong Gi Jeong, and Desok Kim. Ultra short term analysis of heart rate variability for monitoring mental stress in mobile settings. In *2007 29th annual international conference of the ieee engineering in medicine and biology society*, pages 4656–4659. IEEE, 2007.

[80] https://blog.adafruit.com/2012/09/14/new-product-polar-heart-rate-transmitter/, 2018.

[81] Jaakko Malmivuo and Robert Plonsey. *Bioelectromagnetism. 27. The Electrodermal Response*, pages 428–434. 01 1995.

[82] Don C Fowles. The eccrine system and electrodermal activity. *Psychophysiology: Systems, processes, and applications*, 1:51–96, 1986.

[83] Robert Edelberg, Theodore Greiner, and Neil R Burch. Some membrane properties of the effector in the galvanic skin response. *Journal of Applied Physiology*, 15(4):691–696, 1960.

[84] Wolfram Boucsein. *Electrodermal activity*. Springer Science & Business Media, 2012.

[85] Gaetano Valenza, Antonio Lanata, and Enzo Pasquale Scilingo. The role of nonlinear dynamics in affective valence and arousal recognition. *IEEE transactions on affective computing*, 3(2):237–249, 2012.

[86] Roberto Zangróniz, Arturo Martínez-Rodrigo, José Pastor, María López, and Antonio Fernández-Caballero. Electrodermal activity sensor for classification of calm/distress condition. *Sensors*, 17(10):2324, 2017.

[87] Bernd Figner, Ryan O Murphy, et al. Using skin conductance in judgment and decision making research. *A handbook of process tracing methods for decision research*, pages 163–184, 2011.

[88] Michael E Dawson, Anne M Schell, and Diane L Filion. The electrodermal system. *Handbook of psychophysiology*, 2:200–223, 2007.

[89] Grove gsr sensor seeed wiki. http://wiki.seeedstudio.com/Grove-GSR_Sensor/, 2017.

[90] Renate Häuslschmid, Max von Buelow, Bastian Pfleging, and Andreas Butz. Supportingtrust in autonomous driving. In *Proceedings of the 22nd international conference on intelligent user interfaces*, pages 319–329. ACM, 2017.

[91] Doronzo Dario. Interfacce uomo-macchina per veicoli a guida autonoma: Studio della user experience in scenari virtuali. 2018.

[92] https://www.gamespot.com/articles/valve-and-htc-reveal-vive-vr-headset/1100-6425606/, 2018.

[93] https://www.roadtovr.com/latest-vive-shipping-with-tweaked-base-stations-redesigned-packaging/, 2018.

[94] https://www.vrheads.com/exposing-magic-behind-htc-vive-controller, 2018.

[95] VIVE SYSTEM. Vive — business edition and commercial vr. https://enterprise.vive.com/us/BE/.

[96] Davinia Rizzo and M.R. Blackburn. Use of bayesian networks for qualification planning: A predictive analysis framework for a technically complex systems engineering problem. *Procedia Computer Science*, 61:133–140, 12 2015.

[97] https://www.atomicmotionsystems.com/motion-systems/, 2018.

[98] Orlando Games. Talon simulations motion simulator and vr experts. http://talonsimulations.com/, 2018.

[99] Barney Dalgarno and Mark JW Lee. What are the learning affordances of 3-d virtual environments? *British Journal of Educational Technology*, 41(1):10–32, 2010.

[100] Arduino yun. https://store.arduino.cc/arduino-yun, 2017.

[101] Raspberry pi 3 model b. https://www.raspberrypi.org/products/raspberry-pi-3-model-b/, 2017.

[102] Raspberry adc. https://learn.adafruit.com/raspberry-pi-analog-to-digital-converters/mcp3008, 2015.

[103] B Patrao, Samuel Pedro, and P Menezes. How to deal with motion sickness in virtual reality. *Sciences and Technologies of Interaction, 2015 22nd*, pages 40–46, 2015.

[104] Robert S Kennedy, Norman E Lane, Kevin S Berbaum, and Michael G Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3):203–220, 1993.

[105] Roy S Kalawsky. Vruse—a computerised diagnostic tool: for usability evaluation of virtual/synthetic environment systems. *Applied ergonomics*, 30(1):11–25, 1999.

[106] Kristin E Schaefer. Measuring trust in human robot interactions: Development of the "trust perception scale-hri". In *Robust Intelligence and Trust in Autonomous Systems*, pages 191–218. Springer, 2016.

[107] Nasa. https://humansystems.arc.nasa.gov/groups/TLX/.

[108] RM Taylor. Situational awareness rating technique (sart): The development of a tool for aircrew systems design. In *Situational Awareness*, pages 111–128. Routledge, 2017.

[109] Pysimplegui. https://pypi.org/project/PySimpleGUI/.

[110] Numpy. http://www.numpy.org/.

[111] Matplotlib. https://matplotlib.org/.

[112] Rajiv Ranjan Singh, Sailesh Conjeti, and Rahul Banerjee. Assessment of driver stress from physiological signals collected under real-time semi-urban driving scenarios. *International Journal of Computational Intelligence Systems*, 7(5):909–923, 2014.

[113] Kyung Hwan Kim, Seok Won Bang, and Sang Ryong Kim. Emotion recognition system using short-term monitoring of physiological signals. *Medical and biological engineering and computing*, 42(3):419–427, 2004.

[114] Gershon Ben-Shakhar. Standardization within individuals: A simple method to neutralize individual differences in skin conductance. *Psychophysiology*, 22(3):292–299, 1985.

[115] DT Lykken, R Rose, B Luther, and M Maley. Correcting psychophysiological measures for individual differences in range. *Psychological Bulletin*, 66(6):481, 1966.

[116] John Parker Burg. Maximum entropy spectral analysis. In *37ˆ¡ th¿ Annual International Meeting, Soc. of Explor. Geophys., Oklahoma City, Okla., Oct. 31, 1967*, 1967.

[117] Nargess Nourbakhsh, Fang Chen, Yang Wang, and Rafael A Calvo. Detecting users' cognitive load by galvanic skin response with affective interference. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 7(3):12, 2017.

[118] IBM SPSS. https://www.ibm.com/analytics/spss-statistics-software?mhq=spss&mhsrc=ibmsearch_a.

[119] vassarstats: statistical computation web site 2019. http://vassarstats.net/, 2019.

Thanks to her who has the power to affect my emotion.
To make me smile. To make me believe.
To my love.