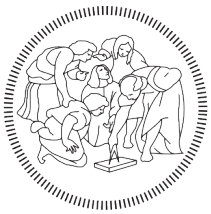


POLITECNICO
MILANO 1863

POLITECNICO DI MILANO
Department of Electronics, Information and Bioengineering
Doctoral program in information technology

Studying user acceptance of autonomous driving in the wild

Supervisor:
Prof. Matteo Matteucci
Tutor:
Prof. Luca Bascetta

Doctoral Dissertation of:
Alessandro Gabrielli

Cycle XXXII

To my family

Abstract

The car's automation technology is progressing fast and has already reached a very good level of safety. Because of the advantages that autonomous cars will introduce, the question of how a driver wants to be driven becomes more and more important to ensure driving comfort for the passive driver and, so, fast and wide acceptance of this technology.

The I.DRIVE (Interaction between Driver, Road Infrastructure, Vehicle and Environment) Lab of Politecnico di Milano, where this research was conducted, aims at developing inter-disciplinary proficiency required for the analysis and modeling of behavioral aspects due to the interaction between driver, vehicle, infrastructure, and environment.

In this research thesis, we have developed a software and hardware platform that allows studying, in a real car, the interaction between driver, car, road infrastructure and environment. We focused on driver stress and the driving factors that impact driver stress to increase the little available knowledge in this field and help the acceptance of car automation in the future.

We instrumented a vehicle with different sensors to acquire vehicle's and environmental's data. To assess the driver's stress with a continuous and objective measure, we acquired the driver's physiological signals; we developed software able to acquire the sensor's signals, and store them in a database. Furthermore, our software analyzes the vehicle and environmental data to extract information about the surroundings and stress indexes from the physiological data. Finally, we used the data in a correlation analysis using the K-nn algorithm to evaluate the correlation between stress indexes and driving features.

To validate our framework, six drivers performed a manual drive, an autonomous drive, and a manual drive as passengers. We validated our platform with the data acquired in these experiments. Moreover, we evaluate the correlation using our software pipeline.

Results demonstrate that longitudinal jerk, the angular velocity on x-axis, and the linear acceleration on y-axis have a high correlation with the skin conductance's phasic component and the respiration rate that are two stress indexes. The ratio between low frequency and high frequency and the heart rate of the electrocardiogram showed a low correlation with all driving features studied.

Acknowledgements

I would like to thanks Stefania Vecchini, Stefania Milani, Marco Colgiago, Roberto Bottini, Silvia Saini, Camillo D'errico, Marco Battisti, Gianluca Recita, Simone Lepre and Ewerton Lopes. During these years, they had supported me and believed in me when I didn't and have helped me overcome the difficulties I met.
I would like to thank the friends I met and worked with in the AirLab during this PhD journey, which made this work period better and less heavy. In particular, I would like to thanks: Simone Mentasti, Riccardo Bertoglio, Michele Bertoni, Giulio Fontana and Luca Morreale.
I would like to thanks also all the people I met during the summer schools and the business trips I have done that have enriched me from a professional and human point of view.
Special thanks to my family, my parents Roberto and Patrizia and my brother Simone that have supported me in every decision I made in my life and helped me to reach this goal.
Finally, I would like to thanks professor Matteo Matteucci, that makes this PhD possible.

Contents

List of Figures

List of Tables

Glossary

AB autonomic balance. 27

ANS Autonomic Nervous System. 10

ARMA Auto-Regressive Moving Average. V, XII, 65, 66, 68–70

BN Bayesian network. IV, 30–32

DBN dynamic Bayesian network. XI, 29, 31

EDA electrodermal activity. III, 20–22, 29, 30

EKG electrocardiogram. IV, 4, 5, 10, 11, 25, 27–30, 33, 42, 47, 49, 51, 52, 60, 83, 89, 91, 93, 124

EMG electromyography. 25–28, 33

GPS global positioning system. XII, 4, 38, 39, 46, 60

GSR galvanic skin response. 4, 10, 128

HF High Frequency. 11, 27, 28, 51, 60

HR heart rate. VI, VII, IX, XII, 4, 5, 11, 25, 27–29, 31, 52, 60, 80–83, 87–91, 94, 95, 98, 99, 110, 114–117, 123

HRV heart rate variability. IX–XI, 11–13, 28, 30, 51, 114, 118–121, 128

I.DRIVE Interaction between Driver, Road Infrastructure, Vehicle and Environment. III, IV, 1–3, 7, 20, 21, 35, 36, 41, 42, 45, 79, 92, 94, 128, 129

IBI inter-beat interval. 25

IMU inertial measurement unit. 4, 40, 60

LF Low Frequency. 11, 27, 28, 51, 60

LF/HF ratio between low frequency (LF) and high frequency (HF). VI, VII, 4, 11, 12, 28, 51, 60, 79–83, 90, 91, 94, 99–102, 123

LIDAR light detection and ranging. 4, 36

MF medium Frequency. 28

NNSBM Nearest Neighbor Statistic Based Method. V, 65, 66, 68–71

NTP Network Time Protocol. 39

PC phasic component. VI–VIII, XII, 4, 5, 12, 52, 60, 76, 81–88, 90, 91, 94–97, 99, 104–110, 123, 124, 128

PCC Pearson Correlation Coefficient. V, 65, 68–70

PNS Parasympathetic Nervous System. 10, 11, 51

RMSE Root Mean Square Error. 66

ROS Robot Operating System. IV, 39, 41, 45–49

rPPG Remote PhotoPlethysmoGraphy. 129

RR respiration rate. 4

SC skin conductance. III, IV, VI–VIII, XII, 4, 5, 10, 12–14, 21, 25–28, 31, 33, 42, 47, 52–54, 60, 76, 80–86, 88, 90, 91, 93–97, 99, 104–110, 123, 124, 128

SNS Sympathetic Nervous System. 10–13, 51

TC tonic component. VI, 4, 12, 52, 60, 80, 82, 83, 90, 91

CHAPTER

1. Introduction

In the past two decades, autonomous cars have become a reality for many car manufacturers and software companies, and, nowadays, there are many ongoing tests on autonomous cars in our streets. Indeed, the car's automation technology is progressing fast and has already reached an excellent safety level.

As autonomous cars will reduce pollution, reduce road traffic, and increase road safety when commercially available, it becomes crucial to have a fast and wide acceptance of this technology. Understanding how a driver wants to be driven becomes a key factor [5].

In 2014, SAE international, a standards developing organization, provided a taxonomy with detailed definitions for six levels of driving automation, ranging from no driving automation (level 0) to full driving automation (level 5), in the context of motor vehicles and their operation on roadways [3] (Figure 1.1). With the higher car automation levels (level 3-5 of SAE), the human role changes from active driver to passengers. The lack of controls perceived as passengers can lead to a higher stress level and discomfort if the car does not behave as expected [5]. If passengers do not feel relaxed and comfortable in autonomous driving conditions, they probably will not use it, and so, the acceptance and usage of this technology will be delayed [6].

Therefore, to guarantee the success and the acceptance of this technology, it becomes essential to focus on the passenger's comfort, as a psychophysiological aspect and not as an ergonomic issue, where there is little available knowledge, and it is also important to understand how to increase the comfort and reduce the stress from the physiological point of view.

The present research is a part of the I.DRIVE Lab (Interaction between Driver, Road Infrastructure, Vehicle and Environment) that aims to develop

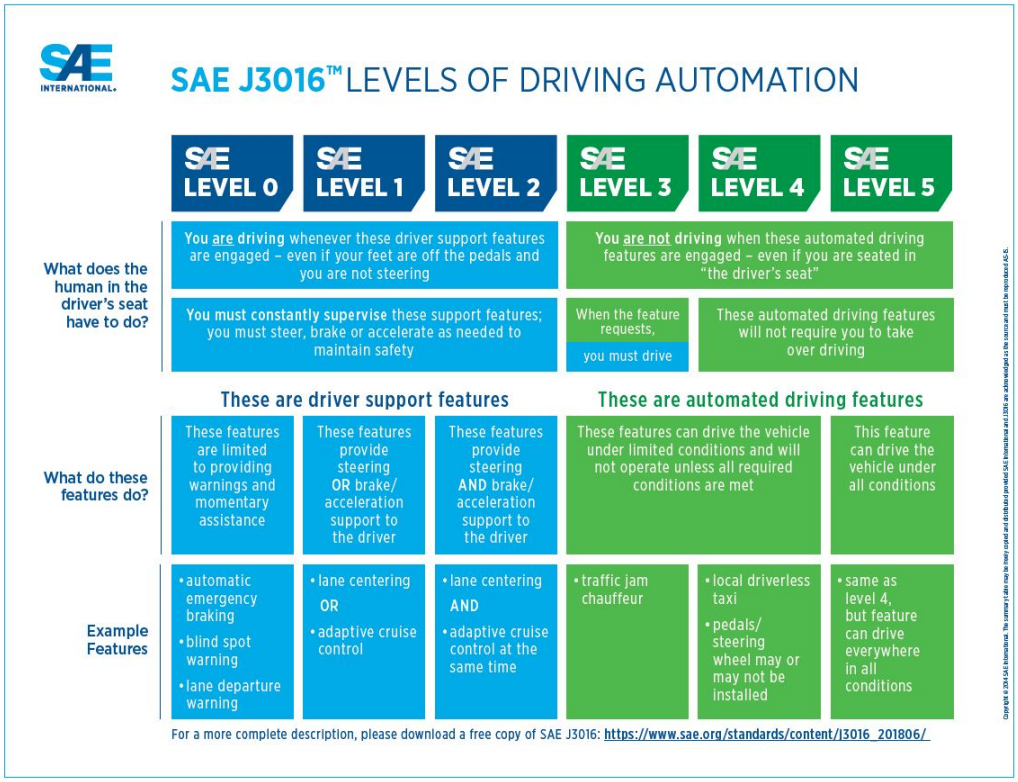


Figure 1.1: An infographic of the six levels defined in the SAE standard

inter-disciplinary competencies for the analysis and modeling of behavioral aspects related to the interaction between driver, vehicle, infrastructure, and the environment in conventional or autonomous vehicles, where the driver becomes, de facto, a passenger (so the term driver refers to both driver and passenger). The laboratory is composed of a fixed structural component, based on a Virtual Reality simulator (Figure 1.2), and a mobile component, based on an instrumented vehicle (Figure 1.3) that are the core of this research.

While we used the I.DRIVE simulator to compare the physiological answer between a real car and a simulator [33], this research aimed to develop a platform that allows studying, in a real environment, the interaction between driver and car, focusing on the driver stress and the driving factor that impact on driver stress, to increase the knowledge in this field and help the acceptance of the car automation technology in the future.

Figure 1.2: The I.DRIVE fixed-based driving simulator during an acquisition

Figure 1.3: The instrumented Tazzari zero car used in this research

The majority of previous research studies that have studied driver stress and workload have been conducted in driving simulators. Driving simulators have many benefits, like doing experiments in a controlled environment, studying dangerous maneuvers or dangerous situations, and repeating the same experiment with different participants. Still, the findings in a simulator may not be easily generalizable to real-life. Indeed, in most simulators, it

is impossible to feel the body's acceleration as in a real car, and the fact that a crash will not cause real harm makes a driver drive faster and less careful [33].

Since there are only a small number of on-road driver's stress studies available, our research focuses on real environments in which the results on driver's stress and what influences the driver's stress may be more accurate. However, the data are more challenging to measure and analyze.

To achieve our goals, we have used an instrumented Tazzari Zero car (Figure 1.3). In the Tazzari we installed cameras, an inertial measurement unit (IMU), a global positioning system (GPS) and a light detection and ranging (LIDAR). From these sensors, we obtain some information such as speed, acceleration and jerk of the car in the three dimensions, number and category of obstacles detected, their position with respect to the car and, for the obstacles in front of the car, their time headway with respect to the car.

Moreover, most previous research in this field uses only questionnaires or interviews to understand how stressful a drive is. Indeed, measuring the passenger/driver discomfort in real-time is not an easy task since the passenger/driver, in our case, is not in a controlled environment, but it is in a highly dynamic environment. There are few examples of real-time discomfort acquisition in literature, such as using a handset control [17] [18] (Figure 1.4). However, it is possible to use a handset only in an autonomous, not in a manual drive, and, in our study, we want to be free to acquire the stress level in both autonomous and manual drive. For these reasons, the real comfort of a passenger/driver is usually acquired after driving with questionnaires or interviews.

Questionnaires and interviews are not a direct stress measure since they are subjective and are retrospective information on the passenger's stress, and, for this reason, they may have strong biases. In the platform developed, in order to acquire a stress metric, physiological sensors are used. In particular, in our car, we have installed sensors to monitor the physiological signals of the driver/passengers, such as the electrocardiogram (EKG), galvanic skin response (GSR) and respiration. From these sensors we are able to extract stress indexes such as the heart rate (HR), the ratio between low frequency (LF) and high frequency (HF) (LF/HF) [41] [45], respiration rate (RR) [46] and the tonic component (TC) and phasic component (PC) of the skin conductance (SC)[23] [45].

Moreover, the physiological sensors are an objective, direct and continuous stress measure. This research has used questionnaires and interviews

Figure 1.4: An example of handset control for real time discomfort acquisition used in the research [17] [18]

only to compare and validate the acquired data analysis results.

With the platform developed in this research, we increased the knowledge of passenger comfort in a real instrumented car from the physiological perspective overcoming previous research limits of using simulators and questionnaires. To validate the entire data acquisition and processing pipeline and ensure that all data acquired is useful for future and better analysis, we had done a correlation analysis on data collected during an acquisition campaign with 6 participants from 24 to 32 years old.

For correlation analysis, we used the K-nn algorithm [26]. This algorithm allows studying the correlation between a time series and event sequences. Time series were retrieved from the stress indexes while the event sequences were extracted from the driving features. Furthermore, this algorithm provides information on the delay relationship between the time series and the event sequences and if the event sequences cause an increase or a decrease in the time series values. We tested the algorithm on the data acquired from our platform.

Results demonstrates that the phasic component extracted from the skin conductance and the heart rate extracted from the EKG are the only stress indexes showing a high correlation with the driving features. Furthermore, results show that the driving factors that impact the driver's stress the most are the longitudinal jerk, the speed, and the angular velocity on y-axis.

1.1 Research questions and objectives

This research's main goal was to develop and validate a platform, allowing the on-road study of the driver's comfort as a physiological issue and the driving features that impact the drive's physiological comfort. Indeed, in the field of driver/passenger's comfort, especially from the physiological point of view, some points needed an investigation and a deepening, such as:

- Which physiological data provides more information on the driver's stress level considering a real drive's dynamic environment?
- Which are the driving factors in a manual drive that impact the driver's comfort/stress?
- Are the stressful driving factors discovered in manual drive the same in an autonomous drive?
- Are the general stress in an autonomous drive higher than a manual drive due to the lack of control?
- Are there stressful driving factors in common for different drivers?

There are already answers to some of these questions in literature but are often limited to simulations [17] [18] [19] [39] [5] [9] [8] or are limited to specific maneuvers [34] [7]. With our system, we extended the answer to these questions to the real environment and developed a tool which will enable further studies in this field

1.2 Thesis outline

The structure of the thesis is the following:

- Chapter 2 briefly describes the physiological signals and some of the stress indexes that are possible to extract from them. Afterward, this chapter gives a review of the literature. We survey the previous studies in the field of driver/passenger stress acquisition and the driving factors that impact this stress to highlight the problems and limitations in the existing literature and how our research contribution has improved these studies.

- Chapter 3 describes the I.DRIVE hardware. It describes in detail the characteristics of the sensors used in the instrumented car. Furthermore, it describes the fixed-base simulator hardware.
- Chapter 4 describes the software component developed and used for the data acquisition, the extraction of additional features from raw data and the data analysis. Moreover, the entire data processing pipeline is shown.
- Chapter 5 describes how the acquisition campaign was done. In particular, it describes the participants who took part in the acquisition and the experimental protocol used in our study. Furthermore, the achieved results are shown and discussed.
- Chapter 6 concludes the thesis, suggesting and detail further direction for our research.

C H A P T E R

2. Literature review

This chapter briefly describes the physiological signals and some of the stress indexes possible to extract from them to understand this chapter better. Afterward, we survey the previous studies in the field of driver/passenger stress acquisition and the driving factors that impact this stress to highlight the problems and limitations in the existing literature and how our research contribution has improved these studies. In the first part of the literature review, we show the existing simulation studies and, in the second part, we show the previous studies conducted on real vehicles. At the end of each section, we compare our research with the existing literature to highlight how our study improved upon these studies.

2.1 Physiological signals and stress indexes

In 2015, Shahsavarani published on International Journal of Medical Reviews a review [38] of stress trying to cover all facts and theories. In psychological sciences stress can be defined as a feeling of mental press and tension. Many formal definitions have been proposed in recent years, some of these definitions can be found in Table 2.1. We use Shahsavarani's 2013 [37] definition *"Any effect of change in surrounding environment on living being which results in disruption of homeostasis (internal balance) of that living being is called stress."* Stress has a positive form that can improve bio psychological health and facilitate performance, but high level of stress could result in biological, psychological and social problems. Stress can be external with environmental sources, as could be in our work, or caused by internal perceptions of the individual. Other factor that influences the level of stress is the exposure time of stressors that could be acute or short-term, chronic or

Author(s)	Stress Definition
Shalev, Yehuda, & McFarlane, (2000) (11)	Stress is a normal psychophysiological response to events which result in the sense of threat, sadness, Dysphoria, and imbalance in people.
American Psychiatric Association (2014) (12)	Stress is described as a sense of being overwhelmed, worry, destruction, press, exhaustion, and lethargy. Therefore, stress can influence people in every age, sex, race, and situation and can result in both physical and psychological health
McEwen (2007) (13)	(In medicine and biology, stress is called to any physical, psychological, and/or emotional factor which results in physical, and/or psychological tension
Behnoudi (2005) (14)	Stress is a situation in which individual is forced to act, and cannot bear the received mental tension. In other words, stress means readjustment of individual with new situations and conditions. Whenever a change occur in life, individual is confronting with stress.
Kumari, et al. (2009) (15)	Stress is bodily response to any demand. Stress could be caused by either good or bad experiences.
Falsetti, Monier, & Resnick (2005) (16)	Stress is any unpleasant emotional experience which is accompanied with predictable biochemical, physiological, and behavioral changes
Silverman, et al. (2010) (17)	Stress is a bodily reaction to a change which needs response, regulation, and/or physical, psychological, and or emotional adaptation. Stress could derive from any situation, condition, thought, and/or state; just need to cause frustration, anger, nervousness, and or anxiety.
Sarafino (2002) (18)	Stress is considered as a situation which is the result of interactions of individuals and their surrounding environments and causes disharmony between situational demands and biopsychosocial resources.
Lazarus, & Folkman (1984) (19); Lazarus (1990) (20)	Stress is an exclusive relation between person and her/his surrounding environment which she/he perceives as taxing, or is gone far beyond her/his coping resources and threatening her/his health.
McEwen (2004; 1999; 1998) (21, 22, 23)	Stress is equivalent to allostasis. Allostasis is the process and capability of gaining stability in the moment of change. Whenever body stress systems are activated in response to high levels of stress, allostasis is evacuated and body exposes to harm.
Shahsavarani, et al. (2013) (5)	Any effect of change in surrounding environment on living being which results in disruption of homeostasis (internal balance) of that living being is called stress.

Table 2.1: Stress definitions in Literature [38]

long-term, for our work we consider only short-term.

The Autonomic Nervous System (ANS) plays an important role in translating stress into a Physiological response. The ANS is composed of the Parasympathetic Nervous System (PNS) and Sympathetic Nervous System (SNS), two branches that are both tonically active with opposing activities. The activity of the Sympathetic Nervous System drives what is called the "fight or flight" response. The "fight or flight" response to emergency or stress increased heart rate, muscle contraction, vasoconstriction, bronchodilation, sweating and many others. The Parasympathetic Nervous System, is activated as a secondary response in what is called "rest and digest" and involves maintaining homeostasis. The hypothalamic-pituitary-adrenal (HPA) axis is a neuroendocrine system that mediates a stress response.

The physiological signals useful to determine the level of stress most used in driving studies are the electrocardiogram (EKG), the respiration rate and the skin conductance (SC) also known as galvanic skin response (GSR).

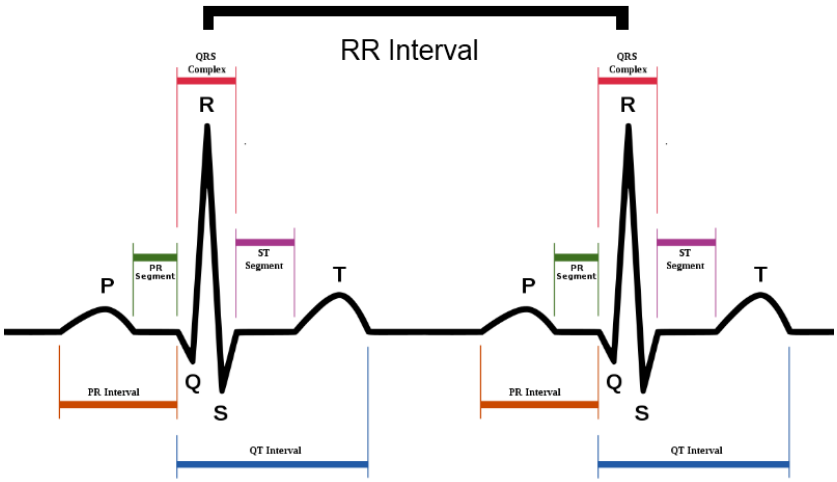


Figure 2.1: An example of two QRS complex and the RR interval

Electrocardiogram

The electrocardiogram (EKG) measures the heart electrical activity. Since the Sympathetic Nervous System activity affect the heart rate, from this signals it is possible to extract different stress indexes as the heart rate (HR) and the heart rate variability (HRV) in time or in frequency domain [41] [45].

For both of these stress indexes it is important to find in the EKG signals what is called QRS complex (Figure 2.1). The QRS complex corresponds to the depolarization of the right and left ventricles of the heart and contraction of the large ventricular muscles. In adults, the QRS complex normally lasts 60 to 100 ms. The time between two consecutive R peeks is called RR interval and it is the base for the extraction of the heart rate and the heart rate variability (Chapter 4.2)

The heart rate is the speed of the heartbeat measured by the number of contractions (beats that correspond to R peeks) of the heart per minute (bpm). The heart rate variability, instead, measures the variation in the time interval between heartbeats (RR interval).

The most widely used methods to analyze the HRV can be grouped under time-domain (Table 2.2) and frequency-domain (Table 2.3). The ratio between low frequency (LF) and high frequency (HF) (LF/HF) estimates the ratio between SNS and PNS. The effects of varying cardiac sympathetic and cardiac parasympathetic nerve activity on LF/HF are non-linear [11], this

Parameter	Unit	Description
SDNN	ms	Standard deviation of NN intervals
SDRR	ms	Standard deviation of RR intervals
SDANN	ms	Standard deviation of the average NN intervals for each 5 min segment of a 24 h HRV recording
SDNN index (SDNNI)	ms	Mean of the standard deviations of all the NN intervals for each 5 min segment of a 24 h HRV recording
pNN50	%	Percentage of successive RR intervals that differ by more than 50 ms
HR Max – HR Min	bpm	Average difference between the highest and lowest heart rates during each respiratory cycle
RMSSD	ms	Root mean square of successive RR interval differences
HRV triangular index		Integral of the density of the RR interval histogram divided by its height
TINN	ms	Baseline width of the RR interval histogram

Table 2.2: HRV time-domain measures [36]

leads to an accurate analysis of LF/HF in order to obtain valid information.

Skin conductance

The skin conductance measures the variation in the electrical characteristics of the skin. As we have seen before, if the SNS is highly aroused, sweat gland activity also increases, which increases skin conductance. So, from this signal it is possible to measure the stress of a person [45] [23].

The skin conductance is divided into two components: the phasic component (PC) and the tonic component (TC). The tonic component is responsible for the slow changes in the skin conductance signal and it is considered the background level of activity. The phasic component, instead, is responsible for rapid changes in the skin conductance signal. The tonic component reflect general changes in autonomic arousal while the phasic component is an

Parameter	Unit	Description
ULF power	ms^2	Absolute power of the ultra-low-frequency band (≤0.003 Hz)
VLF power	ms^2	Absolute power of the very-low-frequency band (0.0033–0.04 Hz)
LF peak	Hz	Peak frequency of the low-frequency band (0.04–0.15 Hz)
LF power	ms^2	Absolute power of the low-frequency band (0.04–0.15 Hz)
LF power	nu	Relative power of the low-frequency band (0.04–0.15 Hz) in normal units
LF power	%	Relative power of the low-frequency band (0.04–0.15 Hz)
HF peak	Hz	Peak frequency of the high-frequency band (0.15–0.4 Hz)
HF power	ms^2	Absolute power of the high-frequency band (0.15–0.4 Hz)
HF power	nu	Relative power of the high-frequency band (0.15–0.4 Hz) in normal units
HF power	%	Relative power of the high-frequency band (0.15–0.4 Hz)
LF/HF	%	Ratio of LF-to-HF power

Table 2.3: HRV frequency-domain measures [36]

event-related stress measure related with SNS activity.

Another stress index derived from the skin conductance is the orienting response [20]. The orienting response measures a fast change in the skin conductance detecting the onset and peak of individual responses, as shown in Figure 2.2. Individual responses are considered an orienting response only if the first derivative of the response exceeds a threshold that depends on the application. Responses occurring less than one second after a previous response are counted as a continuation of that response. Once an orienting response is detected three features are extracted and are: the magnitude of the response (O_M), the duration of the response (O_D) and the area of the response (O_A) computed as $O_A = \frac{1}{2} * O_M * O_D$ (Figure 2.2).

Respiration

The respiration signals measured expansion and contraction of the chest cavity. The respiration is related to the stress [28]. Two stress indexes are extracted from the respiration signals: the respiration rate and the respiration

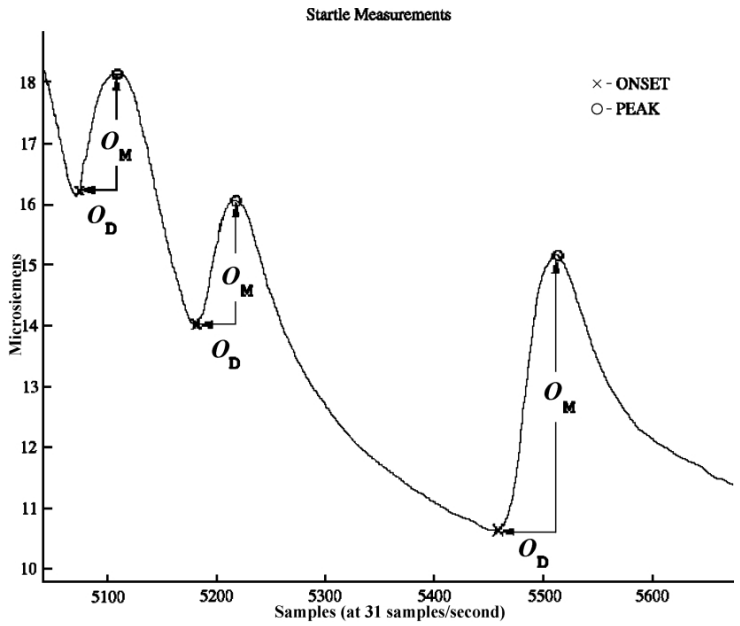


Figure 2.2: An example of orienting response features extracted from the skin conductance (SC) [20]

signal's power spectral density information. The respiration rate is the rate at which breathing occurs. The frequency information in the respiration signal is computed using the power spectral density function, extracting the average energy in each of the first four 0.1 Hz bands of the power spectral density range 0.1-0.4 Hz. Picard et al. [28] showed that the three higher frequency bands of the respiration signal are more related to the stress.

2.2 Driver comfort and trust studies in the driving simulator

In literature, there is driver stress research using simulators. These research benefit from doing experiments in a controlled virtual environment, allowing them to study dangerous maneuvers or dangerous situations safely. Furthermore, it is easier and less time expensive to repeat the same experiment with different participants. The disadvantage of using a simulation instead of a real car is that it is impossible to feel the body's acceleration in most sim-

ulators. Even when the simulator would allow to provide the human with the feelings of some acceleration, it is not easy to reproduce the same feeling and reaction of a real car. For this reason, still, the findings in a simulator may not be easily generalizable to real-life. Nevertheless, there are studies on driver comfort focusing on the improvement of simulators. Some of these studies use real car data to validate and compare the simulation results.

Two types of simulators exist, the fixed-base (Figure 2.3a) and the moving-base simulator (Figure 2.3b). The main difference between the two types is that with the moving-base, it is possible to reproduce some motion feedback due to physical motion mechanism, while the fixed-base simulators have no physical motion mechanisms. Most of the time, the moving-base simulator's motion feedback needs to be as similar as possible to the reality if we want to reproduce the same feeling we experience in real life, but this cannot always be attained [8]. The motion system's parametrization is not a simple task and sometimes requires focusing only on the motion feedback we are interested in.

[8] [9] used the same advanced moving-base simulator platform to find the best motion system's parametrization. The simulator used in these studies is composed of a hexapod that moves along a linear rail and can simulate the motion feedback of a real car (Figure 2.4). Depending on the simulator's orientation with respect to the rail, the simulator can simulate better the longitudinal or the lateral motion.

In the first of these two studies, Bellem et al. [8] assessed the behavioral validity of two simulator configurations comparing with a real test-track. They focus on two maneuvers: the lane change maneuvers for the lateral motion and the deceleration maneuvers for the longitudinal motion. In both of the maneuvers, they tried six different parametrizations for the simulation. The participants had to provide a comfort rating on a scale from 1 to 7 for each scenario.

The data acquired about the participant's comfort showed that the configuration with the simulator settings with the vehicle positioned parallel to the linear rail has a relative and absolute validity. With the simulator settings with the vehicle positioned transversely to the linear rail, it is not possible to demonstrate a relative and absolute validity. Furthermore, Bellem et al. showed that the validity of a simulator highly depends on the motion system's parametrization and that it was necessary to apply a scaling factor for both lateral and longitudinal motion cues with respect to the real car due to the underestimation of speed in virtual environments.

(a)

(b)

Figure 2.3: An example of the two types of simulators-: (a) moving-base (b) fixed-base

In the same simulator, Bellem et al. [9] studied the acceleration and jerk configurations that maximize the comfort in the lane change, acceleration, and deceleration maneuvers. They applied three different acceleration profiles for each maneuver arranged in three pairs for pairwise comparison. The participants stated which of the two varieties presented in a direct pair com-

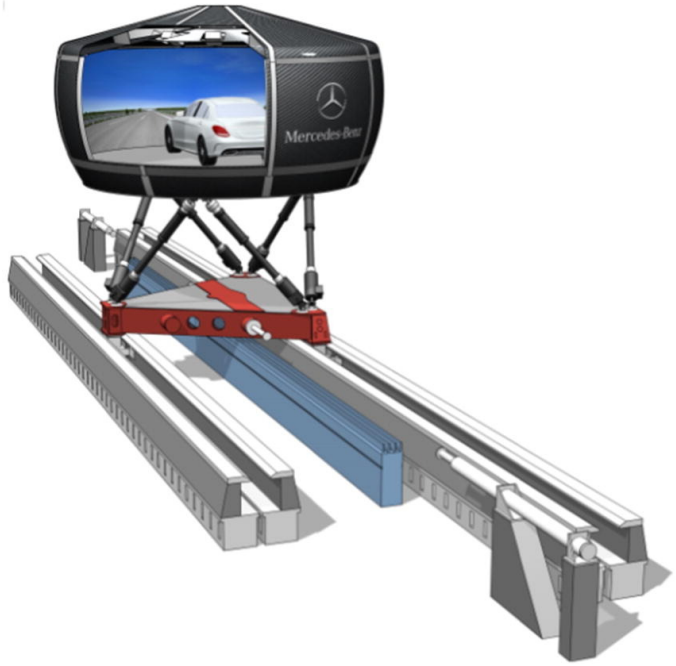
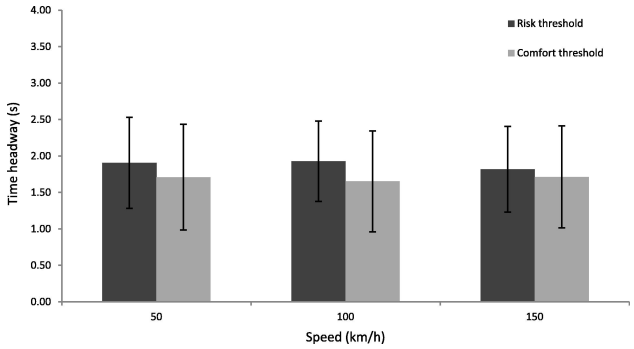


Figure 2.4: The advanced moving-base simulator [8] [9]

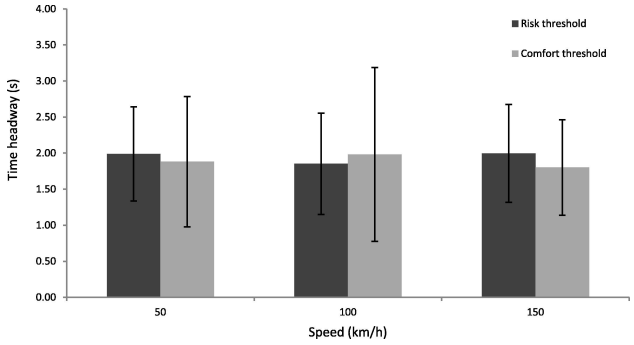
parison they preferred.

Bellem et al. found that the driving style perceived as comfortable in the maneuvers tested are characterized by low jerks and early action in situations in which a criticality might arise and a softer onset in uncritical situations. They also demonstrate that a comfortable autonomous car does not necessarily have to drive as a human driver since passive drivers prefer the safest possible lane change alternative disregarding whether actual risk is immanent or not. Furthermore, in the deceleration maneuver, the approach with the most deceleration in the first part of the maneuver should not be used in automated driving style. This is surprising, as this behavior was reported as natural behavior in literature [25].

In the Siebert et al. [39] work, using a fixed-base simulator, authors studied the time headways threshold that divides the feeling of risk and the feeling of comfort in adaptive cruise control and a self-driving car. They studied the time headway (i.e., s the distance between vehicles measured in time) for three different speeds: 50, 100 and 150 km/h. They varied the time headways from 4s to 0.5s and from 0.5s to 4s, asking the participant to report a



(a)



(b)

Figure 2.5: The mean time headway thresholds and standard deviation in: (a) adaptive cruise control, (b) self-driving [39]

change in their comfort/risk experience.

They found that the mean time headway threshold was between 1.5s and 2.0s and that there was no difference between different speeds. Furthermore, they found that there was no significant difference between time headways of self-driving (Figure 2.5b) and distance-assisted driving (Figure 2.5a). Finally, they suggest that the car needs to adapt to the individual driver's time headway threshold with a high automation level.

In the Scherer et al. [34] work, they focused on modeling a customized automated driving style based on manual driving that makes the driving experience comfortable for the driver.

Scherer et al. [34] used a driving simulator to derive the participant's driving parameters, and then, they tested, in a field study, the parameters essential for the perceived comfort discovered in simulation.

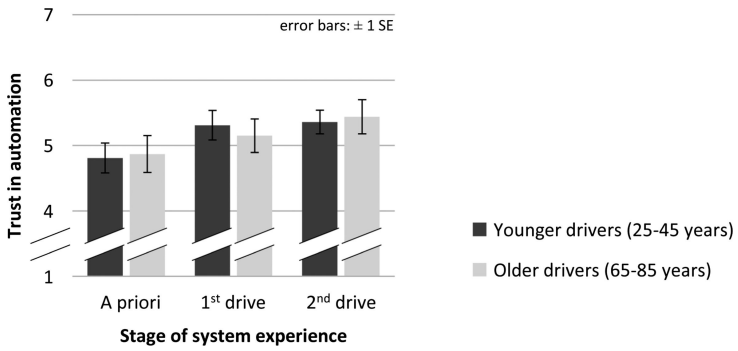


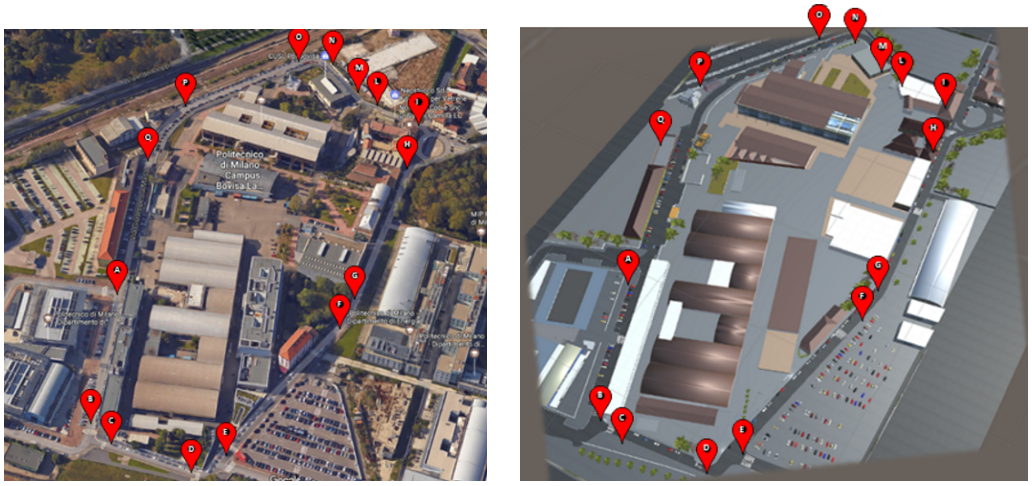
Figure 2.6: Development of younger and older drivers' trust in highly automated driving [19]

The participants drove manually in the simulator, and afterward, this drive was played as a replay so that the participants experienced their own driving style in a highly automated drive. Furthermore, three automated driving styles were presented to the participants, one based on their own driving style and two on other participant's driving style. Scherer et al. assessed the comfort and enjoyment with questionnaires and interviews and the discomfort during autonomous drives with a handset control (Figure 1.4). Together with the comfort data, the simulator data highlighted that longitudinal control parameters were important for the participant's comfort.

The field study on a real car had focused its attention on the braking and acceleration parameters. The same handset control was installed in the test vehicle, like in the driving simulator, to record the perceived comfort. This study showed only preliminary results that indicate that driving parameters concerning longitudinal control have a high impact on subjectively perceived comfort.

In the Hartwich et al. [19] work, the authors examined the development of drivers' trust and acceptance regarding highly automated driving at different stages of system experience in a driving simulator and on a real vehicle. The participants performed four drives in the simulator: a training drive, a manual drive and two highly automated drives. Subsequently, they conducted a test track study to provide a more physically realistic experience. The test track is limited to acceleration and deceleration maneuvers. Questionnaires assessed the trust and acceptance in all the drives.

Analyzing the data, they found that the drivers' acceptance and trust in automation increased after the first system experience (Figure 2.6). The



(a) (b)

Figure 2.7: Aerial view of the of the urban scenario tested for the comparison between (a) the real-life driving and (b) the 3D virtual simulation, reconstructed in Unity 3D using GPS coordinates to segment the different driving scenes. The red pins represents the Landmarks used to define the 15 driving segments

trend continued in the subsequent test track study. Indeed, trust and acceptance remained stable and did not decrease after the real experience.

In the Ruscio et al. [33] study, developed in the I.DRIVE lab, we presented a methodological and conceptual description of a first validation study performed on vehicle data (travel time and speed) and physiological data (electrodermal activity (EDA)) using the I.DRIVE simulator (Figure 3.6) and the I.DRIVE vehicle (Figure 1.3) equipped with the same sensors.

To test the relative and absolute validity of the data acquired in simulation and the real vehicle, we selected an urban portion of the City of Milan - Italy, as a testing scenario. The urban circuit was chosen to include a wide range of turning radii and roadway widths that all together require the driver to adjust his behavior to safely and efficiently face the circuit. The circuit was composed of 15 coherent segments: 8 straights parts, 6 turns and 1 roundabout. The beginning and end parts of each segment were isolated, considering 15 landmarks from GPS coordinates. The landmarks were also used to rebuild the same urban portion of Milan in the virtual simulator, using Unity 3D (Figure 2.7).

In our study, we conducted the pilot test on one male participant that drove on both virtual and real scenarios. Descriptive statistic (Mean and standard deviation) were computed for the overall time, speed and electrodermal activity. We calculated the Pearson Correlation Coefficient to assess the relationship among the distribution of the dependent variables across the 15 different driving scene between the two experimental settings (I.DRIVE virtual simulation, I.DRIVE real-life vehicle). We used the non-parametrical Mann-Whitney U test to test the presence/absence of significant differences between the null hypothesis of random distribution of the dependent measures (time, speed and electrodermal activity) and the observed values in the two experimental settings.

We found only a relative validity for the travel time (Figure 2.8a) and the speed (Figure 2.8b) since, in simulation, the speed is always underestimated. The reactivity scores of electrodermal activity confirmed previous research that indicated different absolute values for skin conductance activity in virtual simulations. However, the pattern of reactivity scores could provide relative validity (Figure 2.8c).

In the Hartwich et al. [17] and [18] works, authors showed the effects of driving automation and driving style familiarity on driving comfort, enjoyment and system acceptance. In a fixed-base simulator, they studied how comfortable and enjoyable is the autonomous drive compared to manual drive. Furthermore, they studied if a familiar autonomous driving style (i.e., a driving style that corresponds to each drivers' manual driving style) leads to higher driving comfort and enjoyment. The participants had to manually drive in different scenarios while the simulator recorded the driving data. Afterward, to simulate the autonomous drive with the participants' driving style, they replayed the participant manual drives recorded in simulation.

On another day, the participants experienced their autonomous driving style as a familiar style and two unfamiliar automated drives based on the other participants' manual drives. They selected young and older drivers to see if comfort and enjoyment differ with age. They assessed the acceptance, comfort and enjoyment with questionnaires and the discomfort during autonomous drive with a handset control (Figure 1.4). They concluded that automation increased both age groups' comfort but decreased younger drivers' enjoyment (Figure 2.9a). Moreover, younger drivers showed higher comfort, enjoyment and acceptance with familiar autonomous driving styles, whereas older drivers preferred unfamiliar driving styles (Figure 2.9b).

In the Basu et al. [5] work, they used a simulator to understand the

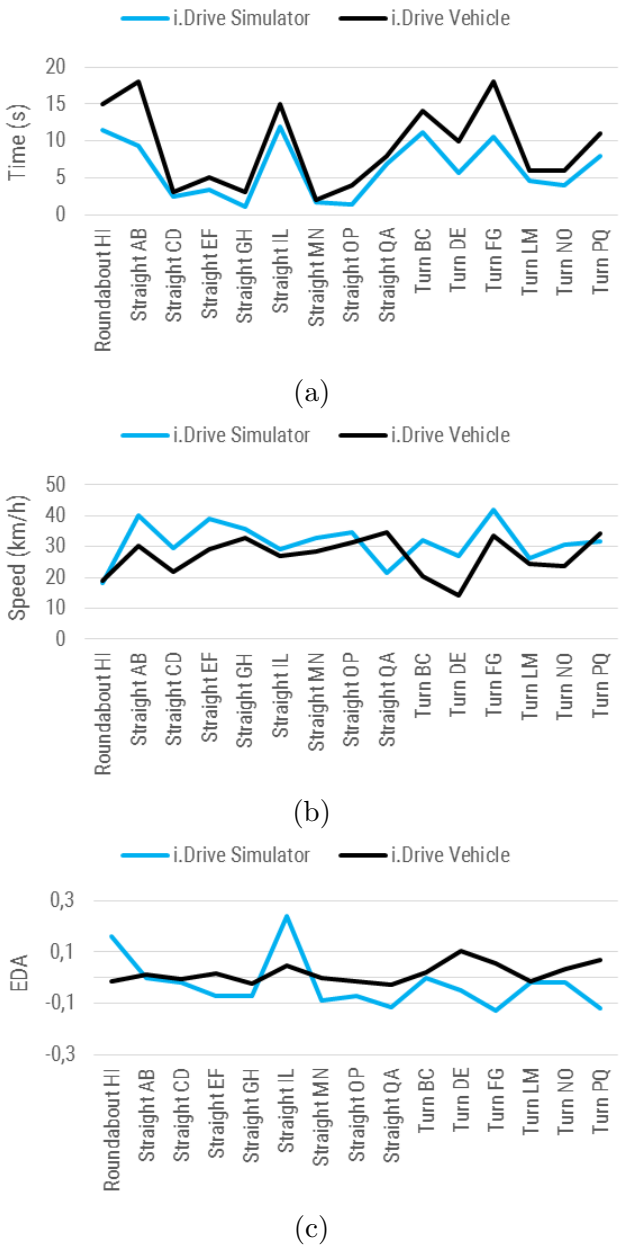
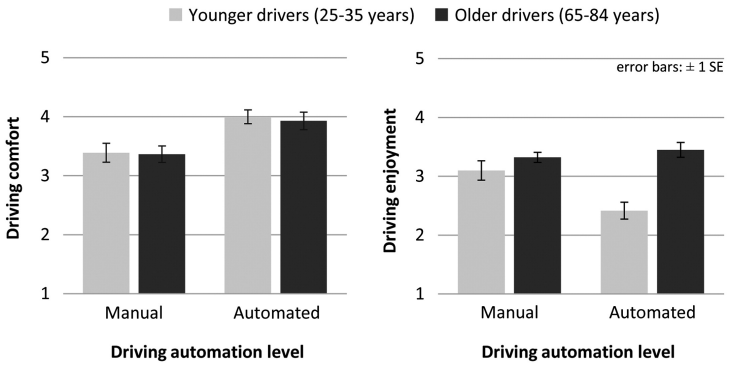
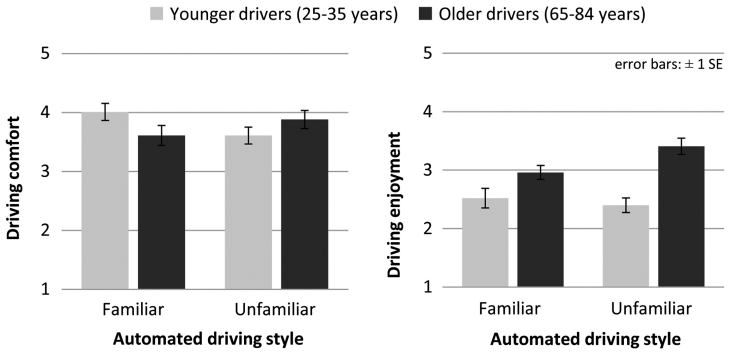


Figure 2.8: Dependent variables measured across the experimental setting for virtual driving and real-life driving: (a) time, (b) speed and (c) electrodermal activity (EDA)



(a)



(b)

Figure 2.9: Driving comfort and enjoyment rated by the younger and older drivers comparing:(a) manual and the own autonomous driving style, (b) the automated driving style familiarity [17] [18]

autonomous drive defensiveness preferred by drivers with respect to their own driving style. The participants demonstrated their manual driving in different scenarios, while the simulator collected data. In the second part, the participants tried four autonomous driving styles with different levels of defensiveness: aggressive, defensive, own style (participants did not know it is their own), "distractor" style (i.e., a different participant's style).

After the participants had tried each autonomous driving style, they conducted an interview asking participants to rate each driving style in terms of comfort and similarity with their own drive style.
They found that, overall, participants prefer a different driving style than their own, and, typically, they prefer a more defensive style than their own

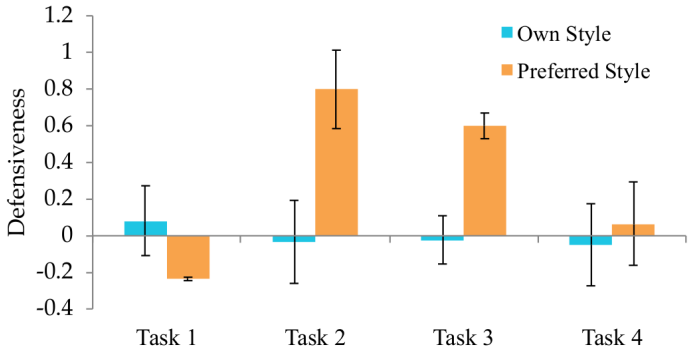


Figure 2.10: Mean Defensiveness Score Across Participants [5]

when they are passengers (Figure 2.10).

In the researches described above, it is possible to observe that it is difficult to reproduce the same feeling and reaction of a real car [8] [33]. The drivers' feeling and reactions also depend on the parametrization, and the configuration of the simulator [8].

Several studies had to validate the driving simulators comparing the data coming from them with the data of a real vehicle [8] [34] [19] [33]. All these works found that there is only a relative validity between data coming from the simulation and a car. Therefore, this confirms that the findings in a simulator may not be easily generalizable to real-life. Despite that, some findings discovered in simulation remain interesting [5] [17] [18] [19] [9] and, for some type of study, the simulation are the only way to perform a study due to safety reasons [39].

With the platform developed in our research, we can extend and validate the simulation study results in a real environment. Indeed, in most simulation studies, comfort data are assessed using questionnaires and interviews. Our research improved the comfort/stress information coming from questionnaires and interviews using an objective stress measurement acquired from the physiological sensors. Physiological sensors provide us better and more detailed information about the driver's state during a drive.

A study in simulation [5] supports one of our research hypotheses that an autonomous drive is more stressful than a manual drive due to the lack of control. Moreover, a study [9] demonstrated that an autonomous car does not necessarily have to drive as a human driver. However, two studies [17] [18]

challenged our research hypotheses showing that younger drivers perceived higher comfort, enjoyment and acceptance with familiar autonomous driving styles. We investigated this point showing that autonomous cars are, in general, more stressful than a manual drive, and so we disagree with the studies [17] [18]

2.3 Driver comfort studies in real vehicles

In literature, there are also few driver stress researches conducted on real cars. These research studies the driver's comfort level and the driving events and parameters that affect the driver's stress level. The advantage of performed studies in real environments is that the results are more accurate with respect to simulations, but it is more difficult to measure and analyze data. Furthermore, all the on-road studies are performed on manual drives since it is difficult to find roads where it is possible to drive autonomously.

In the Bellem et al. [7] study, authors identified the driving parameters to differentiate between three different driving styles: everyday, comfortable and dynamic driving. The study was performed driving a car manually in an urban, rural and highway environment. Bellem et al. asked the participants to drive in a fashion that corresponds to their idea of everyday, comfortable and dynamic driving style. Throughout the test runs, participants rated the maneuvers immediately after they completed them.

The results showed that the discriminating metrics of driving style were the same in both urban, rural (Table 2.4) and highway condition (Table 2.5) with only a few minor exception. The acceleration, jerk, quickness (i.e., the swiftness with which a maneuver takes place) and headway distance in seconds were the essential metrics to differentiate between the three driving styles. Furthermore, they highlighted a higher sensitivity to acceleration and consequently to jerk with higher speed.

The Healey et al. [21] work performed a preliminary study quantifying the driver's stress using physiological signals in a real drive. They collected information about the driving situation's context with cameras and drivers' stress information using different physiological sensors. Four types of signals were measured during the driving task: skin conductance (SC), respiration, electrocardiogram (EKG) and electromyography (EMG). After processing the physiological signals, they extracted some stress index: the heart rate (HR) and inter-beat interval (IBI) from the EKG, the rising edge with a steep

		Overall effect		Everyday – comfortable		Everyday – dynamic		Comfortable – dynamic	
		F	*p*	*t*	*p*	*t*	*p*	*t*	*p*
Accelerating from non-zero speed	Acceleration	48.49	<.001***	5.06	.001**	−5.30	<.001***	−9.34	<.001***
	Jerk pressing gas pedal	17.64	<.001***	3.90	.010**	−2.69	.009**	−5.67	<.001***
	Jerk releasing gas pedal	34.87	<.001***	−4.29	.315	5.46	<.001***	6.48	<.001***
	Quickness	26.71	<.001***	3.40	.028*	−4.25	<.001***	−6.61	<.001***
Lane change	Acceleration	19.77	<.001***	4.35	.009**	−3.02	.002**	−5.81	<.001***
	Jerk	10.53	<.001***	2.94	.046*	−2.03	.071	−4.60	<.001***
	Quickness	14.32	<.001***	2.46	.071	−2.95	.005**	−5.59	<.001***
Following	Headway distance in seconds	31.47	<.001***	−2.51	.012*	5.07	<.001***	8.72	<.001***
	Standard lane deviation	3.10	.049*	0.65	1.000	−1.67	.313	−2.57	.052

* $p < .05$.
** $p < .01$.
*** $p < .001$.

Table 2.4: Metrics able to distinguish between driving styles on urban and rural roads [7]

		Overall effect		Everyday – comfortable		Everyday – dynamic		Comfortable – dynamic	
		F	*p*	*t*	*p*	*t*	*p*	*t*	*p*
Decelerating behind a moving target	Acceleration	7.00	.002**	−1.88	.066	1.71	.092	3.81	.001**
	Jerk releasing gas pedal	0.07	.934	–	–	–	–	–	–
	Jerk pressing brake pedal	2.18	.134	–	–	–	–	–	–
	Jerk releasing brake pedal	3.33	.046*	2.15	.130	0.06	.957	−2.77	.054
	Jerk pressing gas pedal	1.42	.247	–	–	–	–	–	–
	Quickness	7.48	.001**	2.08	.264	−1.62	.180	−4.06	.001**
	Minimum headway distance in seconds	3.88	.025*	−1.69	.420	1.03	.799	2.77	.020*
Accelerating from non-zero speed	Acceleration	13.30	<.001***	3.46	.004**	−1.89	.150	−5.35	<.001***
	Jerk pressing gas pedal	3.09	.053	–	–	–	–	–	–
	Jerk releasing gas pedal	5.33	.007**	−3.48	.006**	−0.86	1.000	2.73	.057
	Quickness	9.25	<.001***	3.11	.027*	−1.61	.250	−4.32	<.001***
Lane change to the left	Acceleration 1	2.28	.118	–	–	–	–	–	–
	Acceleration 2	3.25	.051	–	–	–	–	–	–
	Jerk onset	0.14	.869	–	–	–	–	–	–
	Jerk lane switch	6.85	.003**	1.22	1.000	−2.56	.029*	−3.14	.003**
	Jerk end	5.25	.010*	2.29	.153	−1.19	.591	−3.22	.008**
	Quickness	3.97	.028*	1.45	.595	−1.44	.351	−2.93	.024*
Lane change to the right	Acceleration 1	5.77	.008**	2.35	1.000	−2.05	.088	−2.74	.013*
	Acceleration 2	4.78	.017*	0.80	1.000	−2.06	.087	−2.49	.035*
	Jerk onset	4.19	.026*	1.56	1.000	−1.74	.187	−2.39	.038*
	Jerk lane switch	6.17	.006**	2.05	.815	−1.93	.120	−3.05	.008**
	Jerk end	3.82	.035*	0.96	1.00	−1.79	.166	2.22	.062
	quickness	4.91	.015*	2.68	.745	−1.57	.265	−2.72	.016*
Following	Headway distance in seconds	7.77	.001**	−1.64	.270	2.80	.024*	4.07	.001**
	Standard lane deviation	1.00	.374	–	–	–	–	–	–

* $p < .05$.
** $p < .01$.
*** $p < .001$.

Table 2.5: Metrics able to distinguish between driving styles in a highway setting [7]

slope of the SC, the respiration rate (RR) from respiration and a trigger of muscle activity using EMG. With these drivers' stress and driving context information, they preliminary discovered that multiple sensors could help

Feature	Rank	Correct	Feature	Rank	Correct
$\mu_{\mathcal{R}}$	1	62.2%	$\sigma_{\mathcal{E}}$	6	53.5 %
$\mu_{\mathcal{G}}$	2	62.0%	ΣS_A	7	53.0 %
ΣS_D	3	58.5%	HR	8	52.6 %
$\mu_{\mathcal{E}}$	4	58.3%	AB	9	52.5 %
S_F	5	57.6%	$\sigma_{\mathcal{R}}$	10	50.2 %
ΣS_M	5	57.6%	$\sigma_{\mathcal{R}}$	11	48.3 %

Table 2.6: Recognition stress performance of each individual feature [20]

Optimal Selected Feature Set	SFFS kNN
AB, HR, ΣS_D, ΣS_M, $\mu_{\mathcal{E}}$, $\mu_{\mathcal{R}}$, $\sigma_{\mathcal{R}}$	88.6 %
AB, HR, ΣS_D, ΣS_M, S_F, $\mu_{\mathcal{G}}$, $\sigma_{\mathcal{G}}$	88.4 %

Table 2.7: Recognition stress performance using two different features combination [20]

discriminate reaction to driving events. Afterward, they discovered that individuals' responses to similar driving conditions could vary from day to day.

Healey and Picard [20] continued this work, improving the previous study's indexes used to detect the driver's stress. They extracted the mean and variance of three signals: SC (μ_G, σ_G), respiration (μ_R, σ_R), EMG (μ_ε, σ_ε). They extracted the orienting response features from skin conductance, including the frequency of occurrence (S_F), the sum of durations ($\sum S_D$), the sum of magnitudes ($\sum S_M$) and the sum of the estimated areas ($\sum S_A$ where $S_A = \frac{1}{2} * S_M * S_D$. Two features were extracted from the EKG signal: the heart rate and the autonomic balance (AB) that was calculated by finding the ratio of Low Frequency (LF) energy to High Frequency (HF) energy in a five-minute window of the heart rate spectrum. Furthermore, the individual drivers' stress of the events encountered during the drives were assessed with questionnaires. Each feature was tested to see how well it performed to recognize the stress with a linear classifier (Table 2.6). Afterward, a second analysis was performed combining multiple physiological indexes (Table 2.7). The results showed that using a combination of features, performance for recognizing driver's stress improves significantly.

Recognized as	Recognition Results Labeled as			Rate
	Low	Medium	High	
Low	36	0	0	100%
Medium	0	36	1	94.7%
High	0	2	37	97.4%

Table 2.8: Confusion matrix for the recognition algorithm with 5min intervals [22]

In a similar study, Healey and Picard [22] analyzed physiological data during a real-world driving in rest, highway and city driving conditions, extracting features from 5 minutes non-overlapping intervals of data to distinguish between three levels of driver stress (low, medium, high). They used questionnaires to validate general stress levels extracted from the acquired data.

They used four types of physiological sensors: EKG, EMG, SC and respiration. Statistical features were calculated for each segment: normalize mean of EMG and normalized mean and variance for respiration, heart rate and SC. Four spectral power features were calculated from the respiration signal representing the energy in each of four bands(0-0.1, 0.1-0.2, 0.2-0.3 and 0.3-0.4 Hz). From SC, four orienting response features were calculated (Figure 2.2). They also extracted the heart rate variability (HRV) from the EKG, and from the HRV, they extracted the ratio between low frequency (LF) and high frequency (HF) (LF/HF) feature. Afterward, they trained a recognition algorithm that used a linear discriminant function to classify each 5min windows' features in one of the three stress classes (Table 2.8).

A second analysis compared continuous features, calculated at 1-second intervals. For this continuous analysis they used: mean and variance of the EMG (μ_ε, σ_ε), SC (μ_G, σ_G), respiration (μ_R, σ_R), the mean of HR (μ_H), the LF/HF on a 300s and a 100s window (L100, L300) and the LF+MF/HF (MF stand for the medium Frequency) ratio in a 100s and 300s window (M100, M300). To validate the stress levels in this second analysis, they manually analyzed the driver videotapes by advancing them at 1s intervals, recording the number of stress indicators (stops, turning, bumps in the road, head-turning and gaze changes) in each frame. Finally, they did a correlation analysis between the stress metric created from the video and the stress metrics coming from physiological signals in the 1s segments (Table 2.9).
The result showed that three stress levels could be recognized with a high

Day	L100	L300	M100	M300	HR	$\mu_{\mathcal{E}}$	$\sigma^2_{\mathcal{E}}$	$\mu_{\mathcal{G}}$	$\sigma^2_{\mathcal{G}}$	$\mu_{\mathcal{R}}$	$\sigma^2_{\mathcal{R}}$	w
S1-2	0.53	0.61	0.53	0.64	0.34	0.22	0.01	0.75	0.09	-0.53	0.04	0.01
S1-3	0.45	0.45	0.44	0.42	0.35	0.04	0.01	0.77	0.08	-0.49	0.04	0.00
S1-4	0.45	0.58	0.47	0.60	0.53	0.14	0.06	0.71	0.18	-0.33	0.26	0.01
S1-5	0.41	0.35	0.22	0.09	0.46	0.30	0.08	0.85	0.22	-0.22	0.15	0.01
S1-6	0.62	0.62	0.59	0.62	0.31	0.32	0.09	0.74	0.00	-0.56	0.16	0.01
S1-7	0.46	0.36	0.41	0.31	0.52	0.28	0.04	0.77	0.23	-0.23	0.16	0.01
S2-2	0.49	0.66	0.55	0.69	0.49	0.02	0.03	0.13	0.00	-0.24	0.15	-0.01
S2-4	0.22	0.29	0.13	0.17	0.41	0.27	0.01	0.59	0.12	0.12	0.18	0.00
S3-2	0.74	0.73	0.75	0.74	0.44	0.20	0.06	0.78	0.20	0.17	0.25	-0.01
S3-4	0.46	0.41	0.48	0.48	0.38	0.16	0.06	0.77	0.15	0.59	0.19	0.01
S3-5	0.41	0.51	0.44	0.50	0.35	0.09	0.00	0.81	0.20	0.21	0.01	-0.02
S3-6	0.44	0.53	0.44	0.51	0.40	0.20	0.04	0.73	0.14	0.67	0.24	0.03
S3-7	0.35	0.35	0.39	0.35	0.29	0.22	0.08	0.78	0.16	0.44	0.12	-0.01
R2-1	0.41	0.58	0.39	0.54	0.30	0.20	0.06	0.47	0.06	0.10	0.03	0.00
R3-1	0.32	0.42	0.35	0.41	0.30	0.16	0.13	0.45	0.08	0.03	0.10	0.01
R4-1	0.49	0.55	-0.08	-0.19	0.76	0.37	0.09	-0.07	0.03	-0.28	0.22	-0.03
μ-zs	0.52	0.60	0.49	0.57	0.48	0.17	0.03	0.99	0.08	-0.42	0.10	-0.01
μ-zt	0.50	0.56	0.45	0.49	0.45	0.20	0.05	0.81	0.12	-0.03	0.15	0.00

Table 2.9: Correlation coefficient between the stress metric created from the video and variable from the sensors. A set of random numbers "w" was also correlated with the video metric for each drive. [22]

level of accuracy using 5 min intervals (Table 2.8) and that heart rate and skin conductivity metrics provided the highest overall correlations with 1s intervals (Table 2.9).

Rigas et al. [31] used a Dynamic Bayesian Network to estimate car drivers stress in a real environment. They used the EKG and the electrodermal activity (EDA) sensors. The features extracted were calculated over a 10-second window and were: heart rate variation from baseline (HRvB) and the EDA mean of first absolute differences (MFAD). The HRvB provides the instant variation of the heart rate from the estimated baseline, and the MFAD provides a measure of rapid EDA activity. They used video recordings and the driver's self-annotation to annotate the driving event and driver stress (Normal, Low and medium stress). The driver reported his state after experiencing a driving event. The annotated stress is assigned at a specific time segment.

They observed that using a Naive Bayes classifier on the collected data the classification results are poor (Table 2.10). Afterward, they trained a dynamic Bayesian network (DBN) (Figure 2.11) and showed that the overall accuracy improved with this approach (Table 2.11).

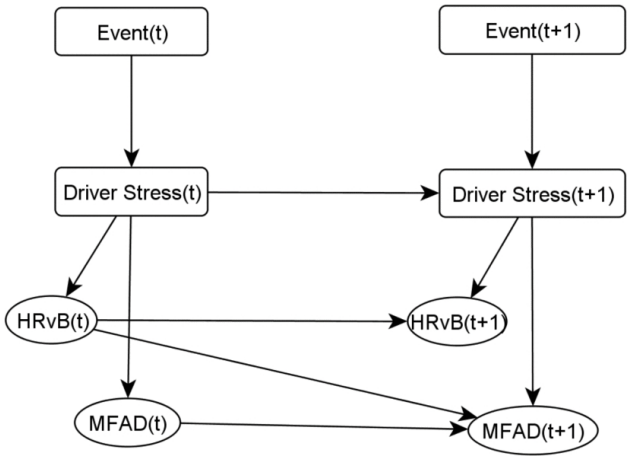


Figure 2.11: The dynamic Bayesian network model for stress detection [31]

Conf. Mat. mean(std)			
	N	L	M-H
Classified as N	119.57(25.50)	44.31(30.03)	18.13(10.31)
Classified as L	24.45(18.88)	78.34(43.56)	59.94(13.35)
Classified as M-H	8.98(6.59)	30.35(7.84)	74.93(6.66)
Overall Accuracy	0.59		
Class	Acc	PPV	
N	0.78	0.34	
L	0.51	0.52	
M-H	0.49	0.34	

Table 2.10: Confusion matrix (conf. mat.), accuracies (acc.) and positive predictive value (PPV) for each class for the classification using naive Bayes classifier of normal(N) vs low stess(L)/medium stress(M)/high stress(H) classes [31]

In another study, Rigas et al. [32] used a Bayesian network (BN) to detect driver's stress events based on physiological features and driving event information in a real unconstrained driving condition. The detection was based on three physiological signals: EKG, EDA, respiration. Features from those signals are extracted using 10s-windows and were: estimation of the trend of HRV (FAD), normalize measures of the differences of skin conductivity (B), the entropy of the spectrum of the respiration (RE). For the event detection (hard braking and overtaking), they used the naive Bayes classifier with the

Conf. Mat. mean(std)			
	N	L	M-H
Classified as N	147.36 (0.94)	4.52(1.85)	2.02(0.06)
Classified as L	0.1 (0.09)	115(13.2)	85.64(14.51)
Classified as M-H	2.54(0.97)	30.48(10.89)	62.34(14.34)
Overall Accuracy	0.72		
Class	Acc	PPV	
N	0.98	0.96	
L	0.77	0.57	
M-H	0.42	0.65	

Table 2.11: Confusion matrix (conf. mat.), accuracies (acc.) and positive predictive value (PPV) for each class for the classification using dynamic Bayesian network of normal(N) vs low stess(L)/medium stress(M)/high stress(H) classes [31]

car information on speed, deceleration, heading change, throttle and RPM of the engine.

A microphone in the car acquired the driver's self-annotations of stress levels caused by events. The driver had to say the event, and next, the perceived stress level. They trained and tested two BN, one with only the physiological features (Figure 2.12a) and one adding to the previous one the driving information (Figure 2.12b).

Comparing the two BN models' results, they observed a significant increase in stress event detection accuracy, adding the driving information. The first model (BN1) had a significant number of false-positive (Table 2.12) due to the increasing of heart rate or skin conductance without drivers experiencing increased stress. The driving events information introduced in the second model (BN2) reduced the false positives increasing the accuracy (Table 2.13).

In the researches described above, it is possible to see that most of the on-road studies on driver comfort are focused on how to assess the stress level of the driver [21] [20] [22] and which are the driving parameters or events that affect the driver stress [7] [31] [32]. Furthermore, it is possible to observe that there are no studies on automated cars. Indeed all the studies are performed in manual driving conditions and, often, are limited to only some maneuvers or events [7] [32] [31]. We overcome these limits with our

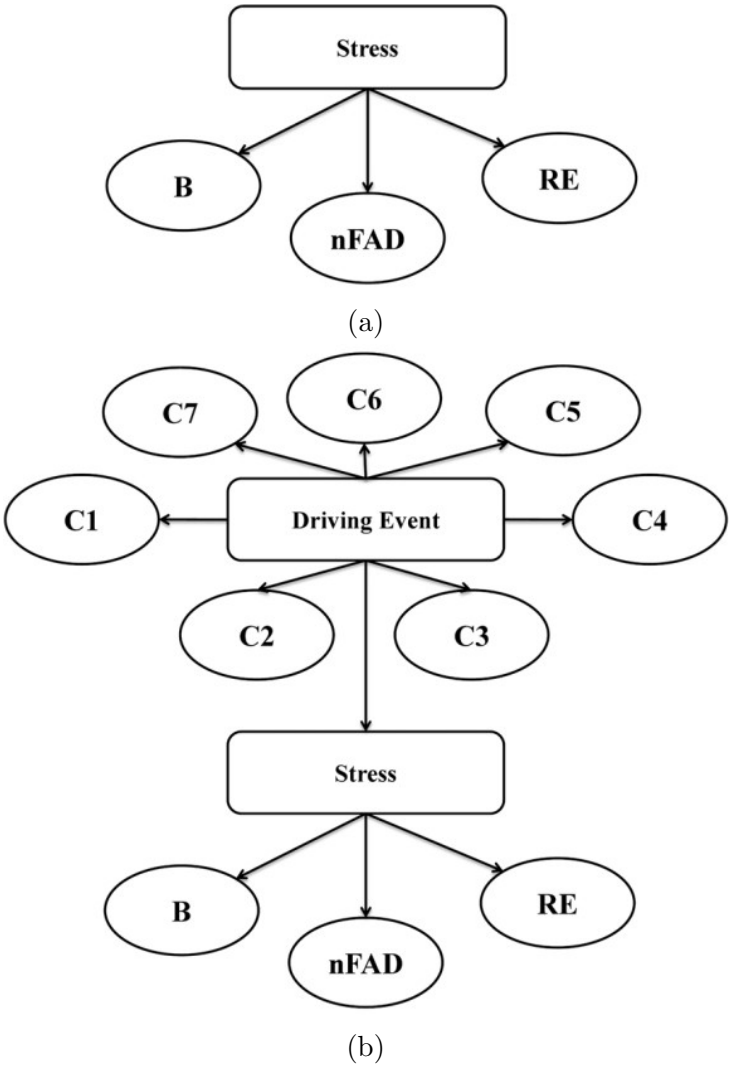


Figure 2.12: (a) BN1 model using only physiological features. (b) BN2 model using also driving information [32]

platform since we can study the drivers' stress level in autonomous driving conditions, and besides, we can compare the stress level with manual driving. We focused on driving parameters that affect the driving stress during the entire driving, not just looking at single maneuvers.

Physiological sensors are used to assess the driver's stress level [21] [20] [22] [31] [32], corroborating our hypothesis that it is possible to assess driver

	Train Subject		Test Subjects	
	Confusion Matrix		Confusion Matrix	
Classified as	**Non S.**	**S.**	**Non S.**	**S.**
Non S.	2352	16	2514	9
S.	327	49	547	76
Specificity	0.99	0.13	0.99	0.13
Sensitivity	0.87	0.75	0.83	0.89
F1	0.93	0.22	0.9	0.21
Acc.	**0.88**		**0.82**	

Table 2.12: Confusion matrix, sensitivity, specificity, and F1 for each class and total accuracy for the classification of stress events for train and test subjects using BN1 model [32]

	Train Subject		Test Subjects	
	Confusion Matrix		Confusion Matrix	
Classified as	**Non S.**	**S.**	**Non S.**	**S.**
Non S.	2590	28	2914	37
S.	89	37	147	48
Specificity	0.99	0.29	0.99	0.24
Sensitivity	0.97	0.57	0.95	0.56
F1	0.98	0.39	0.97	0.34
Acc.	**0.96**		**0.96**	

Table 2.13: Confusion matrix, sensitivity, specificity, and F1 for each class and total accuracy for the classification of stress events for train and test subjects using BN2 model [32]

state information from physiological sensors and giving us information on which are the most promising physiological data to use. Indeed, EKG, SC and respiration are showed to be the most promising signals and the most used in these studies, while EMG signal is showed to be not so reliable [22]. In our studies, we acquired all these promising signals, and these previous studies support our choice.
Questionnaire and interview are used to validate the stress level measured using physiological stress metrics and are not used as the only driver's stress measure.

C H A P T E R

3. I.DRIVE

In our research, we needed to acquire data about vehicle, environment and driver's physiological state. This chapter describes the I.DRIVE hardware, which include the vehicle, the simulator and all equipped sensors.

3.1 I.DRIVE vehicle

The vehicle is a *Tazzari Zero* (Figure 3.1), a two seats electric city car. This car has a range of 140 km given by lithium batteries. The battery of the vehicle powers all the sensors installed in the car, resulting in a reduced, autonomy that, for our studies, is not a problem. The car's maximum speed is 100km/h, and it is not possible to drive on highways since the engine has only 15KW power. The vehicle supports four driving modes:

- **Standard**: this mode is suggested for everyday use in the urban environment
- **Economy**: this mode limits the acceleration and the maximum speed to increase the autonomy of the batteries
- **Rain**: this mode is similar to the standard mode with limits to acceleration and brake assist, in order to increase the grip and safety in rainy conditions
- **Race**: this mode allows maximum performance in terms of acceleration and speed. In this mode, the autonomy decrease

Figure 3.1: The I.DRIVE instrumented Tazzari zero

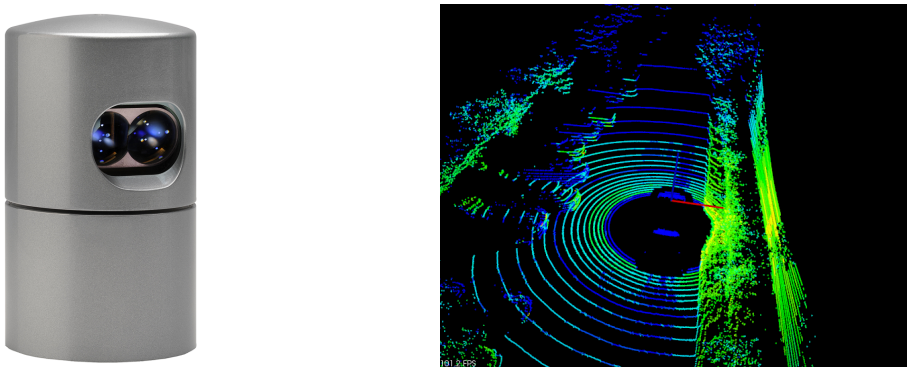
3.1.1 Environmental sensors

Different sensors are equipped in the vehicle to perceive the environment surrounding the car, giving us different information about the type and the distance of the objects around the car (e.g. pedestrians, cars, cyclists)

LIDAR

Light detection and ranging (LIDAR) is a technology that measures the distance by lightening a target with a laser and analyzing the reflected light signals. In our car, we have installed a *Velodyne HDL-32E* (Figure 3.2a) that utilizes 32 lasers vertically aligned from $+10.67°$ to $-30.67°$ to provide an unmatched vertical field of view. Its rotating head design delivers a real-time, 360° horizontal field of view. This sensor generates a point cloud of up to 700,000 points/sec (Figure 3.2b) with a range from 1 meter up to 100 meters with an accuracy of ± 2cm. The rotating head frequency is customizable from 5Hz to 20Hz, but the standard rotation frequency is 10Hz.

LIDAR sensor allows our vehicle to perceive the obstacle around, giving the distance from them. This information is used for obstacle avoidance



(a) (b)

Figure 3.2: (a) The velodyne HDL-32E, (b) an example of pointcloud

in the navigation software. Furthermore, using a neural network [48] [47], we were able to detect the obstacles as well as the type of obstacles (e.g. pedestrians, cars, cyclists).

Cameras

On the car's roof, two *Prosilica GC1020* (Figure 3.3a) have been mounted, one facing the street in front of the car and one looking at the car's back. Another Prosilica Camera is installed inside the car, looking at the face of the driver. The Prosilica GC1020 is a network camera connected to the PC through Ethernet. It can work at a maximum frame rate of 33 fps with a maximum resolution of 1024 x 768. It supports different output format such as:

- Mono format: Mono8, Mono12, Mono12Packed
- Color RGB format: RGB, RGB8Packed, BGR8Packed
- Raw format: BayerRG8, BayerRG12, BayerGR12Packed

We set the Prosilica cameras to work at 30 fps with a resolution of 1024 x 768, and the output format used is BayerRG8. The Prosilica GC1020, looking at the car's front mounts a particular wide-angle lens called *Theia MY125M* (Figure 3.3b). Thanks to the focal length of 1.3mm, this lens allows us to have a horizontal field of view of 125° and a vertical field of view of 109° with very low distortion.

(a) (b)

Figure 3.3: (a) Prosilica GC1020 camera, (b) Theia MY125M lens

The external cameras are used only to have visual feedback of what is happening precisely in front and back of the car. The internal camera is used to see what the driver is doing during the drive.

3.1.2 Vehicle Dynamics Sensors

The vehicle dynamics provide information on the driver's driving style. Furthermore, it is also useful to identify driving events like hard braking or an overtaking. We installed in the vehicle sensors able to acquire this data.

GPS receiver

In our vehicle, we have two different global positioning system (GPS) receivers. The first one is a GPS *Garmin 18LV* (Figure 3.4a) that generates position data according to the standard $GPRMC NMEA at a frequency of 1Hz. The Garmin GPS is connected through a serial port RS-232 to the Velodyne sensor. This GPS is also used to synchronize data generated from the Velodyne sensor with time pulses precision. If Garmin 18LV is not connected to the Velodyne, an internal clock is used that has a drift of about 5 seconds per day. Thus, the Garmin 18LV GPS is essential in our settings because it gives us the car's position and guarantees synchronized data from the Velodyne sensor. Furthermore, with the position information coming from this sensor, we estimate the car's speed.

In our car, we installed another GPS receiver called *Yuan10* (Figure 3.4b. The additional GPS receiver is due to the onboard computer's synchronization problem that we discovered using the Garmin 18LV GPS. The Garmin receiver's synchronization problem is not due to hardware problems but due to this receiver's data acquisition using the UDP protocol.

(a) (b)

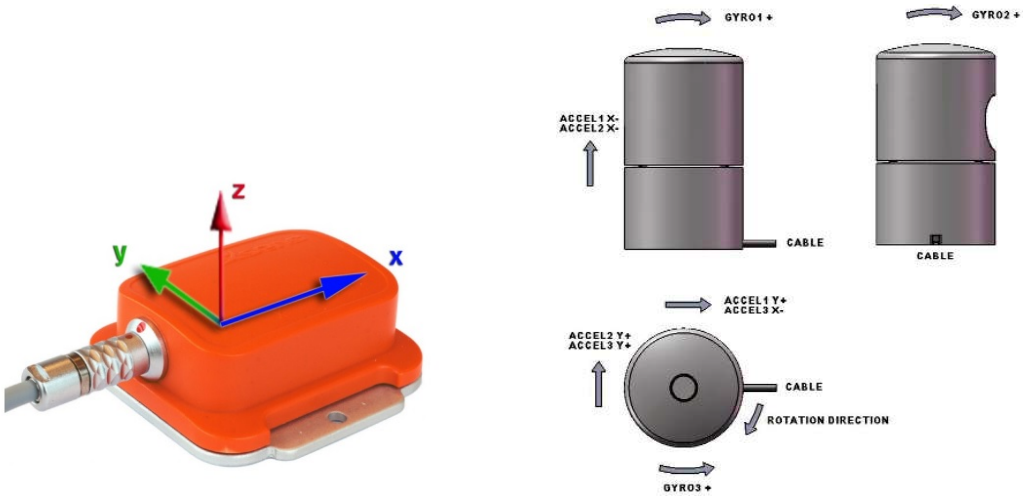
Figure 3.4: (a) The Garmin 18LV GPS receiver, (b) the Yuan10 GPS receiver

Features	**Garmin 18LV**	**Yuan10**
Hot start	1 s	< 1 s
Warm start	38 s	< 29 s
Cold start	45 s	29 s
Reacquisition	< 2 s	1 s
Sensitivity	-155 dBm min	-148 dBm, -161 dBm
Update rate	1 Hz	1, 2, 4, 5, 8, 10 Hz

Table 3.1: Comparison table between Garmin 18LV and Yuan10 GPS receivers

The UDP packets are acquired using different ROS (Robot Operating System) node that introduced a latency that is not acceptable for the onboard computer synchronization. Instead, the Yuan10 is connected to the onboard PC with a USB and, the time information is directly sent to the software that handles the PC clock synchronization (NTP). The Yuan10 receiver has a max frequency of 20Hz, but, in our setup, a frequency of 1Hz is enough. The data generated by this GPS use the NMEA standard.

We selected this GPS receiver for the synchronization since it has a high sensitivity (-148 dBm) that allows it, in a cold start, to have a fast position fix. The high sensitivity during tracking (-161 dBm) allows continuous coverage of the position, also where the satellite coverage is low. In Table 3.1, it is possible to see a comparison table between the used GPS receivers.



(a) (b)

Figure 3.5: (a) Xsens MTi IMU with reference system, (b) the Velodyne IMU with reference system

Inertial measurement unit

The inertial measurement unit (IMU) used in our system is an *Xsens MTi* (Figure 3.5a). This IMU contains a 3D magnetometer that provides information about the earth's magnetic field, a 3d accelerometer that provides information about the linear acceleration and a 3D gyroscope that provides the angular speed. The acquisition frequency of this sensor is 100 Hz.

With the Xsens IMU, we can measure all the acceleration perceived inside the vehicle, and, from this information, we can calculate the jerk. Another useful information retrieved from this sensor is the angular speed that gives us information on how fast the car is turning.

Inside the car, there is another IMU that is embedded in the Velodyne sensor Figure 3.5b. The Velodyne IMU is composed of three 1D gyroscopes and three 2D accelerometers. The information coming from this IMU is less precise than the Xsens data, also because the Velodyne is installed on the roof of the car, and this is not the best position to have precise car's inertial measures.

Figure 3.6: The I.DRIVE fixed-base simulator

3.1.3 On-board computer

In the car, we have installed a *Shuttle slim* computer that, despite its small size, has enough computing power to handle all the ROS acquisition software and has enough storage space to save the data acquired by the sensors. This computer does not perform signals analyses that are done offline on a more powerful computer.

3.2 I.DRIVE driving simulator

The driving simulator consists of a fixed-base physical structure that includes a set of commercial vehicle controls such as the steering wheel with force feedback, the gear shift lever with automatic transmission, and the brake and accelerator pedals (Logitech G920) (Figure 3.6). Like a real car, it is possible to move the seat closer or farther than the steering wheel to find a comfortable driving position. Three 32-inch screens providing 175° field of view show the virtual driving scenario.

The simulation software is crucial since it influences the degree of immersion. In our simulator, we use two different simulation software: a game engine called Unity 3D [42] and IPG CarMaker [16] that is a software developed for driving simulation.

3.3 Physiological sensors

The I.DRIVE vehicle and simulator assess the driver/passenger's state using the same physiological sensors.

3.3.1 Procomp Infiniti

Procomp Infiniti is an 8 channels encoder (Figure 3.7a) used to acquire, in real-time, physiological signals of a person. Two of these eight channels have a sampling rate of 2048 samples/sec, and the other six channels have a sampling rate of 256 sampling/sec. The Procomp Infiniti sample the input signals coming from physiological sensors (Figure 3.7b), encode the samples, and send them to a TT-USB interface. The transmission of the data from the encoder to the TT-USB interface passes through a fiber optic cable to ensure the signals' quality and the best electrical insulation.

The signals that the Procomp Infiniti can acquire are:

- Electromyography (EMG)
- ElectrocardiogramEKG
- Electroencephalogram (EEG)
- Skin conductance SC
- Body temperature
- Respiration
- Blood Volume Pulse (BVP)

These physiological sensors are considered as not invasive.

In our study, we use the respiration sensor that needs to wear a chest band and the EKG and SC sensors that need to apply electrodes on the driver's skin. The EKG is the only signals acquired at 2048 samples/sec, the SC and the respiration are acquired are 256 samples/sec.

(a) (b)

Figure 3.7: (a) Procomp Infiniti encoder, (b) Sensors available for the Procomp Infiniti

3.3.2 Eye tracker

The eye tracker used in our work is a *Pupil core* [24] glasses (Figure 3.8). This sensor has three cameras, one for the world image and two for the eyes. The world camera mounts a wide-angle lens with a field of view of 100° and can acquire images at different resolution and frame rate: 1920x1080 @30fps, 1280x720 @60fps and 640x480 @120fps. Eyes cameras have an infrared sensor that helps find the pupil independently from the iris color and have three possible configurations: 1920x1080 @30fps, 1280x720 @60fps and 640x480 @120fps.

For our experiments, the world camera uses the 1280x720 @60fps configuration, which is a good trade-off between resolution and sampling frequency. The eye camera uses the 640x480 @120fps configuration since, for pupil detection, it is better to have a high sampling rate instead of a high resolution.

The Pupil core needs a calibration every time that a camera is moved and every time that it is taken off and put on. The calibration process uses a marker to correlate the world camera with eye cameras.

With this sensor, we have gaze tracking during driving. We know where the driver/passenger is looking and where the driver/passenger is focusing his attention. Furthermore, by analyzing the world images acquired by the eye tracker with the YOLO neural network [30] (Figure 3.9), we also know what the driver/passenger is looking at (e.g. pedestrians, cars, cyclists).

Figure 3.8: The Pupil core eye tracker

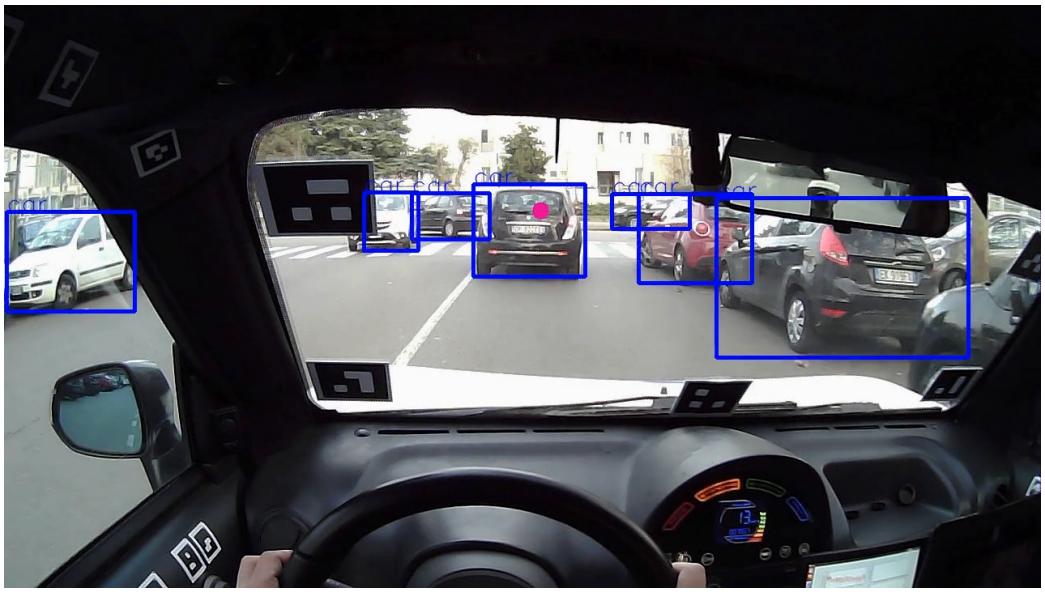


Figure 3.9: An example of the world camera image with the gaze position represented as a purple circle on the followed car and the bounding box generated by the YOLOv3 network

C H A P T E R

4. I.DRIVE software architecture

This chapter describes the software developed to acquire the data of the I.DRIVE vehicle's sensors showed in Chapter 3, and the software developed and used to process and analyze these data. Furthermore, the architecture and the pipeline of the system is shown.

4.1 Vehicle data acquisition

The data acquisition software developed is based on ROS (Robot Operating System) [29] [2], an open-source framework widely use in robotics. One of the main features of ROS is the modularity; indeed, its structure is based on nodes, messages and services. This feature allowed our acquisition software to have a modular design, and we can add or remove sensors driver easily. The only sensor that is not acquired using ROS is the eye tracker since it needs its software to work.

4.1.1 ROS driver and data visualization

Our physical sensors need a ROS driver node that read the signals generated from them and that send the data acquired on a ROS topic. For some of our sensors, ROS already provides the driver nodes, but we need to develop a custom driver node for others.
The Procomp Infinity works only on Windows operating systems. However, our ROS system works on an Ubuntu operating system, which made it necessary to develop a driver on a Windows virtual machine. The Windows

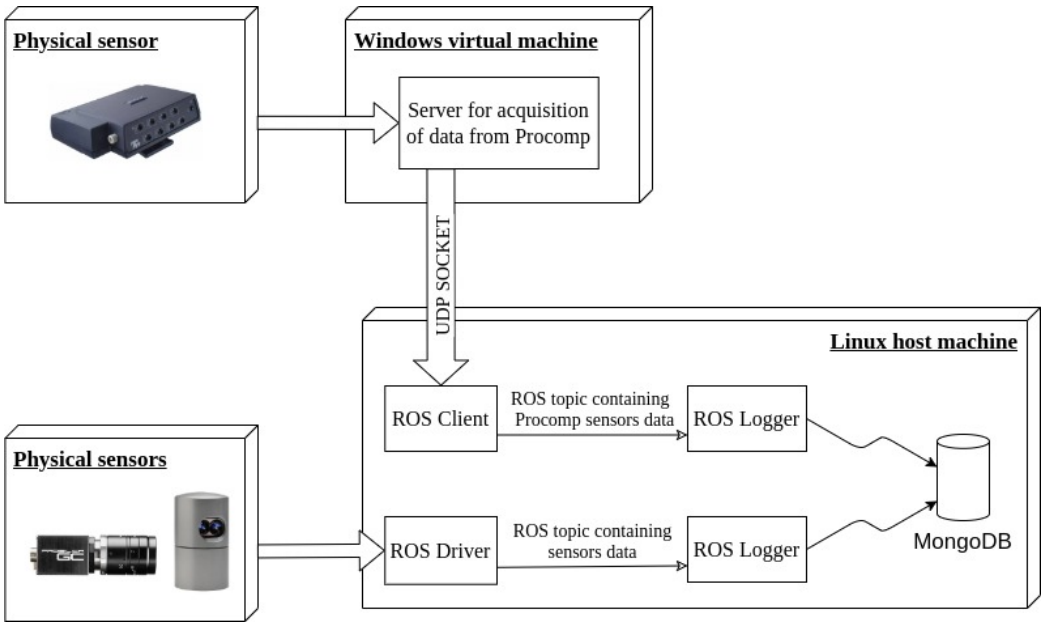


Figure 4.1: Acquisition software architecture from physical sensors to ROS topic

driver read the Procomp data and send them through a UPD socket to a ROS client that convert the data received in a ROS topic.
In Figure 4.1 it is possible to see a general schema of our ROS drivers.

The GPS sensor needs an additional ROS node since the NMEA sentence acquired with the driver is not enough. For extracting the car position and speed from the NMEA sentence, we used a ROS node connected to the GPS driver.

The real-time visualization of the acquired data is essential since it allows to know if the drivers are working correctly. ROS provides a package called RVIZ, a visualizer used to visualize standard ROS topic data. Almost all the topic generated from our sensor drivers use a standard message format, and so it is possible to visualize them on RVIZ (Figure 4.2).
Since the Procomp ROS client use a custom topic and not a standard one, we create a custom ROS node that visualize all the physiological data acquired by the Procomp (Figure 4.3).

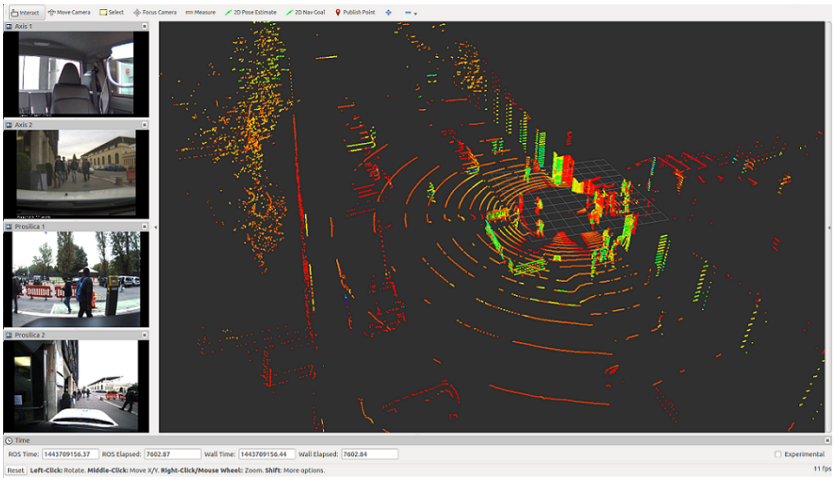


Figure 4.2: RVIZ screen example with the Proslica images and the point cloud generated by the Velodyne

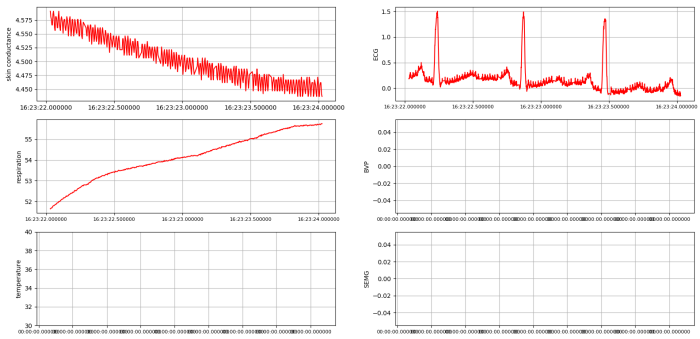


Figure 4.3: Example of the Procomp custom visualizer with the EKG, SC and respiration signals

4.1.2 ROS data storage

The data generated by the sensors and acquired using ROS need to be stored on the PC for the subsequent offline analysis. We use the MongoDB database [1] since it has better performance with respect to other databases. MongoDB is a NoSQL database that uses JSON-like documents called BSON. ROS provide a package that converts the topics into BSON documents and stores them into MongoDB. This node generate a MongoDB collection for

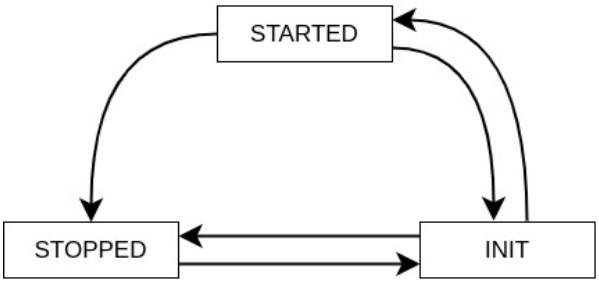


Figure 4.4: The ROS node finite state machine

each of the ROS topic. We modified this package to make it more efficient with the topics that contain a large amount of data. We create a new database for each driver's acquisition so that it is easy to organize and store different acquisitions.

4.1.3 ROS finite state machine

To handle the state of the nodes in our ROS acquisition systems, we created a finite state machine with three possible states:

- INIT: this state indicates that the node is booting and initializing or, for a driver node, that it is waiting that the sensor is connected.
- STARTED: this state indicates that the node is working correctly
- STOPPED: this state indicates that the node is not working or it stopped to work

. The Figure 4.4 show the possible transition between this states.
Each ROS node has its finite state machine and has to communicate, with a ROS service, its state to a ROS node called Heartbeat that manages all node state machines. Every ROS node also has to send an alive signal continuously to the heartbeat node because, if the time between to different alive signals becomes greater than a timeout, the heartbeat node considers that node as stopped and changes the state of that node in STOPPED.
The states of the nodes are shown in a custom visualizer developed by us that display the state of the nodes and some topic statistics such as the delivered messages per second, the dropped messages per second and the


LOGGERS

Sensor	Prosilica 1		Prosilica 2		Xsens IMU	
Sensor node state	stopped		stopped		stopped	
Topic name	/prosilica1/image_raw	/prosilica1/camera_info	/prosilica2/image_raw	/prosilica2/camera_info	/xsens/imu	/xsens/mag
Topic logger state	stopped	stopped	stopped	stopped	stopped	stopped
Delivered msg / sec	0	0	0	0	0	0
Dropped msg / sec	0	0	0	0	0	0
Traffic (Kbytes / sec)	0	0	0	0	0	0

Figure 4.5: An example of ROS node state visualizer for the two Prosilica sensors and the Xsens IMU

traffic generated in Kbytes/sec (Figure 4.5). With this monitoring system we can see in real-time if the entire acquisition system is working or if something is happened during an acquisition.

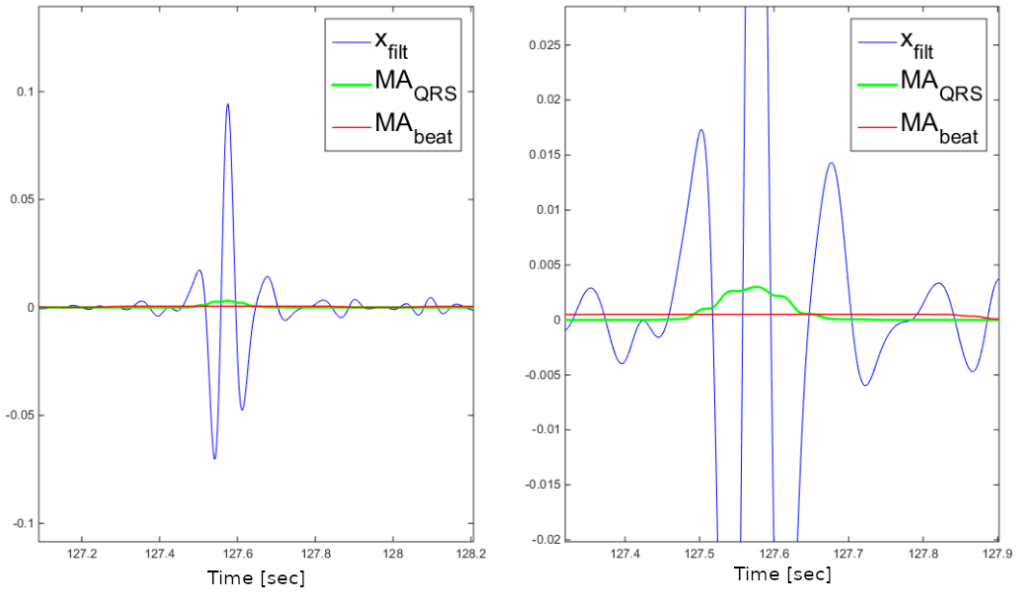
4.2 Features extraction

The data acquired by our ROS system are raw and, for some raw data, we need to prepare and process them to extract further information useful for the subsequent analyzes. The data preparation and processing are done offline using the data saved in the MongoDB databases and the CSV files generated by the eye tracker software.

4.2.1 Stress indexes

EKG

The extraction of the stress index from the physiological signals is an important part of the data processing. To extract the stress indexes we use custom Matlab® software [14]. The EKG raw signal is pre-processed using a band-pass filter to reduce the noise on the signals and extract the signals' components in the interval 8Hz-20Hz, the most interesting EKG frequencies for the QRS complex. The filter output signal (x_{filt}) is squared to improve R peak extraction performance. From this squared signal, we generate two different signals. One signal tells the QRS complex's energy in the entire heartbeat period and is extracted applying a moving average on a window



(a) (b)

Figure 4.6: (a) An example of the QRS signal and the MA_{QRS} and MA_{beat} signals. (b) A zoom on the (a) graph in which it is possible to see that where MA_{QRS} is greater than the MA_{beat} the QRS complex is present

W1=611msec.

$$MA_{beat}[n] = \frac{1}{W1} \sum_{i=-\frac{W1}{2}}^{\frac{W1}{2}} x^2_{filt}[n+i]$$

The other signal helps identify the QRS complex's position using a moving average on a window W2=97msec.

$$MA_{QRS}[n] = \frac{1}{W2} \sum_{i=-\frac{W2}{2}}^{\frac{W2}{2}} x^2_{filt}[n+i]$$

It is possible to identify the position of a QRS complex seeing when the MA_{QRS} are greater than MA_{beat} plus a threshold [13]Figure 4.6. Also using a threshold, this algorithm can identify false QRS complex. We decide to

implement two different threshold, one high and one low,

$$T_{high}[n] = \begin{cases} 1 & \text{if } MA_{QRS}[n] > MA_{beat} + \beta\bar{z} \\ 0 & \text{otherwise} \end{cases}$$

where $\beta = 0.5$ and $\bar{z}$ is the mean value of the squared filter signal x^2_{filt}.

$$T_{low}[n] = \begin{cases} 1 & \text{if } MA_{QRS}[n] > MA_{beat} \\ 0 & \text{otherwise} \end{cases}$$

Afterward, we analyzed the T_{high} and T_{low} checking that the possible QRS complexes identified have a time interval greater or equals to 97ms. When an identified QRS complex on T_{high} pass the control we search in that interval the R peak looking at the highest value of x_{filt}. Furthermore, an AR model is implemented using the measure

$$d(i) = R(i) - R(i-1)$$

where *d(i)* is a temporal series given by the difference of two consecutive positions of the R peak. With the AR model, we can predict the value of d(i+1). We search around the predicted position if there is a valid T_{high} QRS complex. If there is a valid T_{high} QRS complex, a new R peak is extracted in the identified T_{high} interval. Otherwise, we search around the same predicted position if there is a valid T_{low} QRS complex and we use this interval to search the new R peak.

After we have identified all the R peak in the EKG signal and create the temporal series *d(i)*, we can extract stress indexes. One of the stress indexes is the heart rate variability (HRV) spectral power. First, we resample the d series to 4Hz and, then, we analyze a two-minute sliding window extracting the power spectrum with the Welch periodogram [44]. From this spectrum analysis, we extract the frequency power in the interval 0.04Hz - 0.15Hz called Low Frequency (LF) and in the interval 0.15Hz - 0.4Hz called High Frequency (HF). We are interested in the ratio between low frequency (LF) and high frequency (HF) (LF/HF) that is a stress index since the LF is influenced mostly by the Sympathetic Nervous System and the HF is influenced by the Parasympathetic Nervous System. So when a person is stressed, the LF activity increases and the HF activity decreases. Therefore, the more a person is stressed, the higher is the LF/HF ratio.

Another stress metric extracted from the EKG is the heart rate (HR) obtained from the d series.

$$HR(i) = 60s/d(i)$$

The HR is measured as beat per minute and this is why we divide 60s with the time between two R peaks.

Skin conductance

The SC raw is resampled to 4Hz. Afterward, the signal is deconvolved using the Bateman function to obtain the SC driver [10]. This deconvolution has the disadvantage that amplifies error noise to high frequency. To solve this problem, the deconvolution output signal is convolved with a Gaussian filter equivalent to a low-pass filter. The resulting of this process is the SC drive signal $d_{SC}[n]$. This signal is the sum of two components, the driver of the phasic component (d_{ph}) and the driver of the tonic component (d_{to}).

$$d_{SC}[n] = d_{ph}[n] + d_{to}[n]$$

To estimate the tonic component of the SC we performed a peak detection on the $d_{SC}[n]$ function in order to identify impulses. An impulse section is defined by the local minima preceding and succeeding the peak in time. All time sections that are not part of detected impulses are considered to reflect non-overlapped tonic component.

Finally, $d_{to}[n]$ is estimated for a time grid with 10s spacing by averaging the values of available inter-impulse sections within the range of half of the grid spacing before and after the grid points. We interpolate this values based on the grid data to obtain $d_{to}[n]$ (Figure 4.7). Convolving the $d_{to}[n]$ with the Bateman function, we obtained the tonic component (TC) of our SC signal. The phasic component (PC) is retrieved subctracting to the SC signal the tonic component. The phasic component of the skin conductance is a reliable measure of the stress of a person. Furthermore, it is event-related because the response to an event is rapid (1s - 4s).

Another stress metric extracted from the skin conductance is the orienting response [22] [20]. To extract the orienting response features, we detected the onsets and peaks of the orienting responses by first detecting slopes exceeding a critical threshold (0.5) and then finding the local minimum preceding that point (onset) and the local maximum following that point (peak). Afterward,

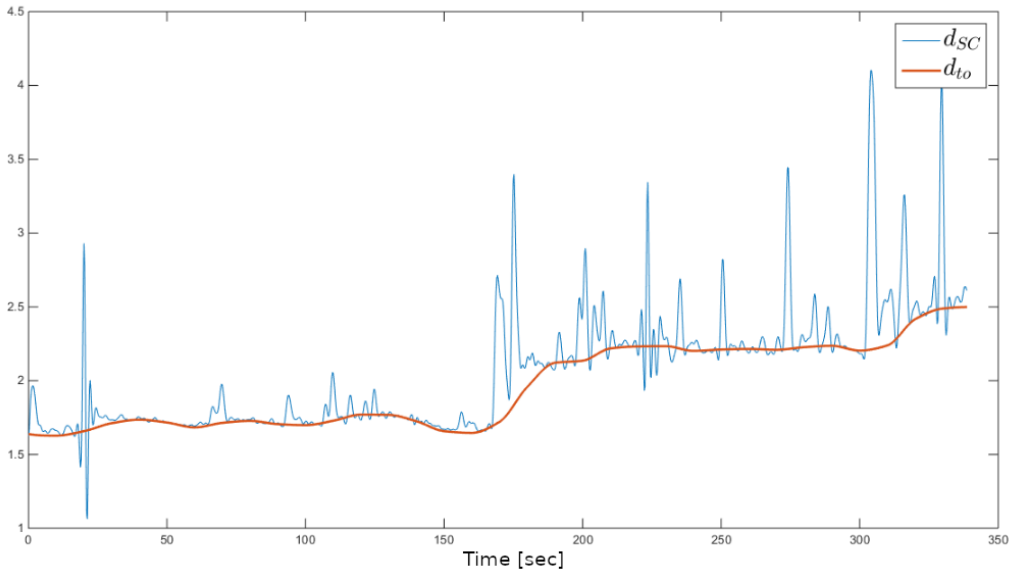


Figure 4.7: The SC driver $d_{SC}[n]$ and the tonic driver $d_{to}[n]$

for each pair of onset and peak, we extract three orienting response features Figure 4.8:

- The magnitude that is the difference between the peak and the onset values

$$O_M = v_{peak} - v_{onset}$$

- The duration that is the difference between the peak and the onset time

$$O_D = t_{peak} - t_{onset}$$

- The energy that is the area of the response

$$O_A = \frac{1}{2} * O_M * O_D$$

Also the frequency of the orienting responses give us stress information. With this stress index, we detected the most stressful events.

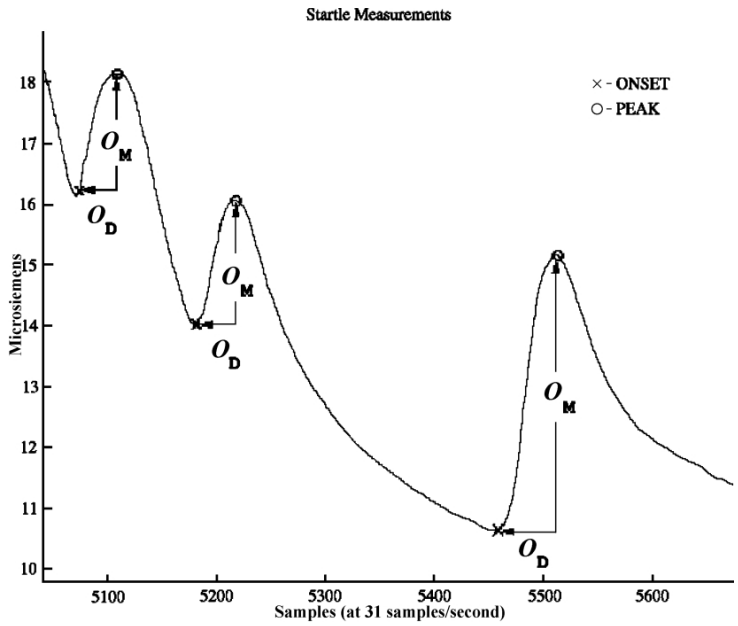


Figure 4.8: An example of orienting response features extracted from the skin conductance (SC) [20] [22]

4.2.2 Gaze position

The Pupil eye tracker needs its software to work. With the Pupil software, we can calibrate the eye tracker before any acquisition using a marker and we can save locally the acquired data. This software also extracts useful information about the gaze and the pupil.

The output of this software is a series of CSV files and the three eye tracker cameras video. In these CSV, there is a timestamped information on the gaze positions in the world camera images with related confidence (Figure 4.10a) and the pupil diameter. Furthermore, we took the world camera images and we processed them with a convolutional neural network called YOLOv3 [30] (Figure 4.9) to perform object detection.

YOLOv3 is trained to identify more than 100 classes of objects. For our purpose, we use it to identify only eight classes that include: car, truck, bus, traffic light, motorbike, bicycle, person. For each identified object, YOLOv3 provides a confidence score that tells how good is the detection. We analyzed only detected objects with a confidence level greater than 25%, since the results were pretty accurate, although the low confidence level used (Fig-

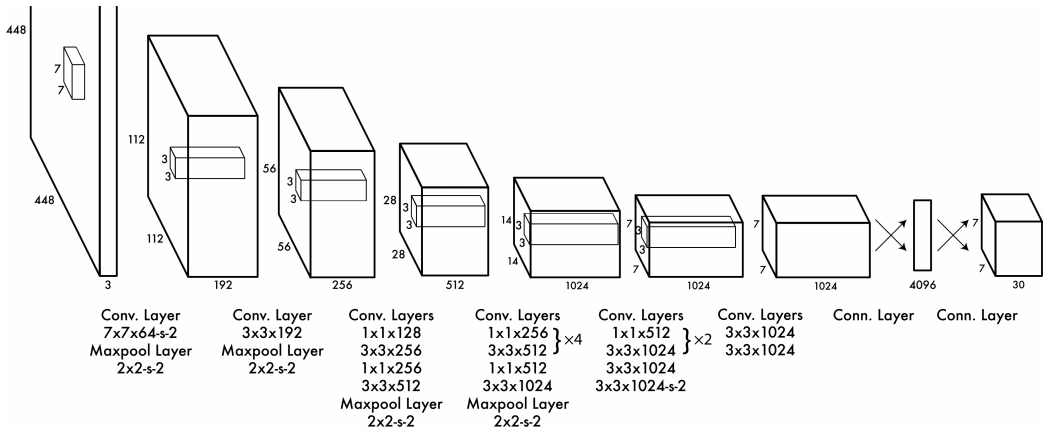


Figure 4.9: The YOLOv3 detection network architecture. The network has 24 convolutional layers followed by 2 fully connected layers [30]

ure 4.10b).

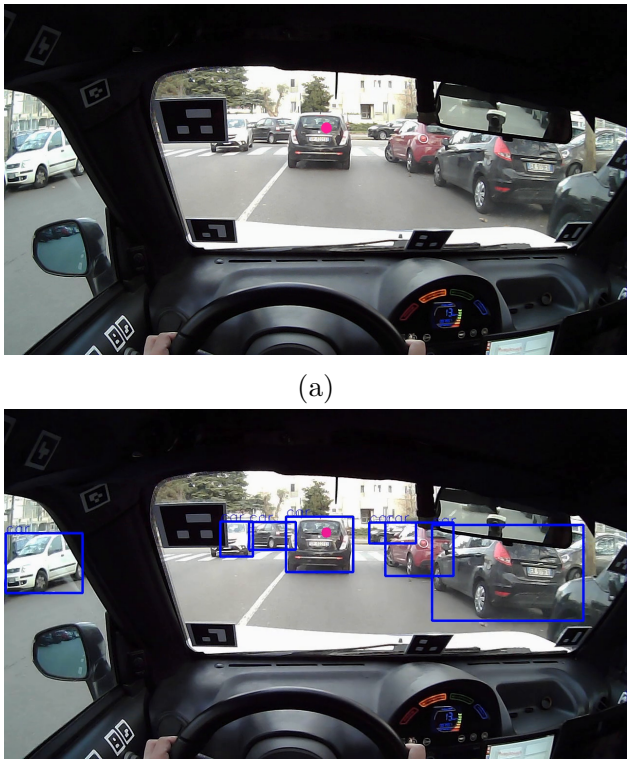
Since the eye tracker world images acquired are taken from inside the car, YOLOv3 detected our cockpit most of the time, classifying it as a car. To overcome this problem, we excluded from the YOLOv3 output all the bounding boxes, which dimension was greater than 60% of the entire frame.

4.2.3 Object detection and identification

We analyze the Velodyne point cloud using e deep neural network named SECOND [48] [47]. SECOND, with his detection network structure (Figure 4.11) is able to identify in the 3D space of a point cloud, different driving objects as cars, cyclist, buses, construction vehicles, motorcycles, pedestrians, traffic cones, trailers, trucks and barriers. It is also able to provide the orientation estimation of the objects detected.

We tried this network on our LIDAR data using the developer's pretrained SECOND network, but the detection results were not good. Our low detection results were due to the dataset used to train the pretrained network called KITTI [15]. The KITTI dataset has a LIDAR with 64 layers that are double our 32 layers. This mismatch causes the low detection result that we obtained.

To overcome this problem, we trained the SECOND network using the NuScene dataset [12], a large-scale public dataset for autonomous driving developed by nuTonomy. NuScenes provide data from the entire sensor suite



(a)

(b)

Figure 4.10: An example of the world camera image with the gaze position represented as a purple circle on the following car after the Pupil software extraction in (a). In (b) the bounding box generated by the YOLOv3 network

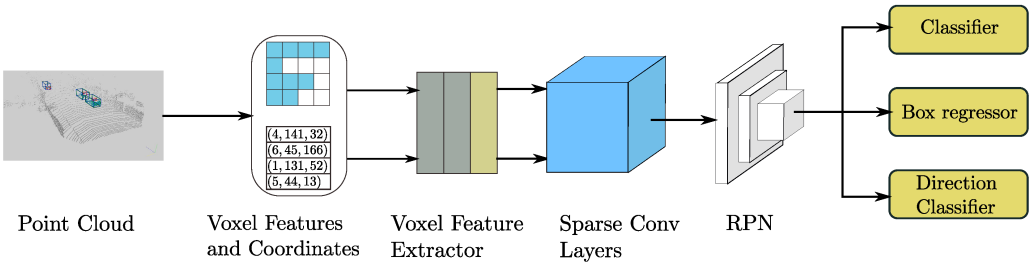


Figure 4.11: The SECOND detection network structure [48]

of an autonomous vehicle: 6 cameras, 5 radars and one lidar, all with a full 360-degree field of view. NuScenes comprises 1000 scenes, each 20s long and

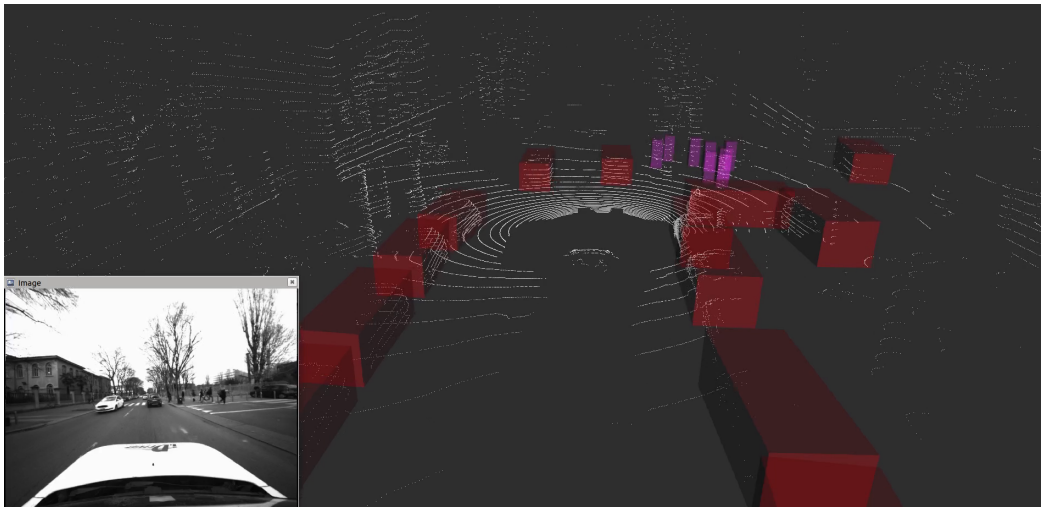


Figure 4.12: An example of the SECOND detection output on our point clouds. The red boxes are cars while the purple ones are pedestrian

fully annotated with 3D bounding boxes for 23 classes and 8 attributes. This dataset uses the same Velodyne HDL-32E sensors that we have on our car, so the LIDAR data provided are the same LIDAR data that we acquired.

To analyze our point clouds with the SECOND network trained with NuScene data, we converted our MongoDB point cloud documents into multiple files with the ".bin" extension using the same data structure of the NuScene dataset.

The output generated by the SECOND network for each point cloud contains (Figure 4.12): the 3D position, orientation and dimension of the bounding boxes detected, the labels of these bounding box (e.g. car, bicycle, pedestrian, etc...) and the confidence of the detection. We have considered only the detection with a score higher than the 60% that avoid false detection, maintaining almost all the correct detection. Some false detections are still present using this threshold, but it is not a problem since they are present in only a small number of frames.

The information coming from this elaboration is crucial since we are able to know what objects are around the car and the distance from these objects.

Since the LIDAR point clouds have a 360° horizontal field of view, we had divided the field of view in 4 interval of 90° (Figure 4.13): front, back, right and left.

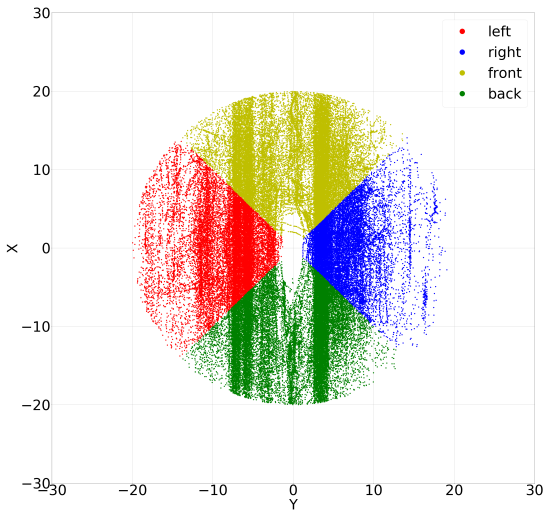


Figure 4.13: An example of identified objects position in an entire acquisition divided in the four interval: front, back, left, right

We extract another driving information from Velodyne that is the time headway, the time difference between the nearest identified object and is calculated as:

$$Time\ headway = \frac{distanceofnearestobjectinfront}{speedofthecar}$$

4.3 Data preparation

After we acquired the sensors and extracted new and useful information, we needed to merge all this data in a data frame using the same sampling frequency for all the time series. Furthermore, the time series sampling frequency had to be fixed for the entire acquisition interval.

We have seen in Chapter 3 that the sensors installed in the vehicle acquire at different sampling frequencies. We have also seen that data processing sometimes changes the sampling frequency. We chose the 100Hz frequency

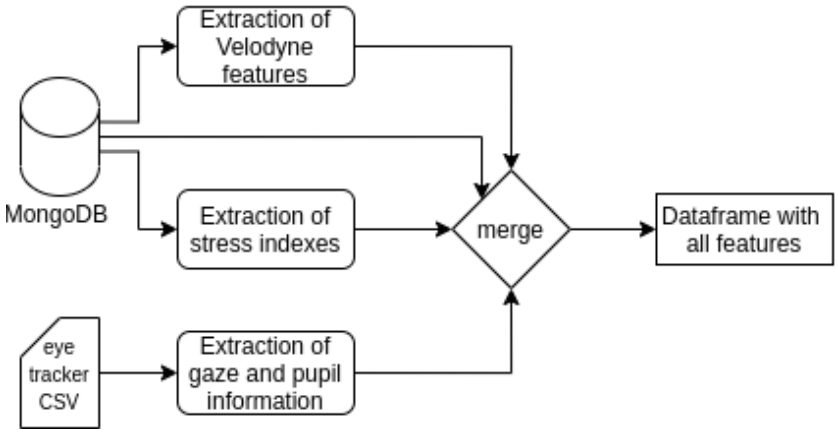


Figure 4.14: Data pipeline from the source of data to the final merged Dataframe used for the correlation analysis

that allows us not to lose any useful information. The frequency of 100 Hz is much higher than needed for some stress indexes, which change with a frequency of 1-4 Hz. Furthermore, 100Hz is the same frequency of both IMU sensors from which we extract essential information on the vehicle dynamics. We interpolate data when there are missing values only if valid data surround the missing data. After interpolation, the missing information will be considered as NaN that will be handled by our correlation analysis software.

In Figure 4.14 it is possible to visualize the data pipeline from the data storage to the final data-frame. We retrieve the data from the MongoDB database and the CSV generated by the Pupil software, extract new features with the algorithms showed in section 4.2. After that, we merged all these data downsampling and upsampling them. The merged data are used subsequently for the correlation analysis.

4.4 Data analysis

The first data analysis we performed answered the question of which driving feature affects the driver's stress level and which values of a single feature affect and how the stress level. This analysis is performed using the K-NN algorithm [26] [35], an algorithm able to finds the correlation between a time series and an events sequence.

This algorithm is applied for each driving session (manual driving, au-

tonomous driving and manual driving as passenger) of each participant.

We considered for this analysis, 4 physiological stress indexes:

- Phasic component (PC) and tonic component (TC) of skin conductance
- Ratio between Low Frequency (LF) and High Frequency (HF) (LF/HF) of EKG
- Heart rate (HR)

Furthermore, we considered 20 driving features:

- Angular velocities on xy axes obtained from the Xsens IMU
- Linear accelerations on xy axes obtained from the Xsens IMU
- Jerks on xy axes extracted from the Xsens IMU
- Angular velocity on xy axes obtained from the Velodyne IMU
- Linear acceleration on xy axes obtained from the Velodyne IMU
- Speed extracted from the Velodyne GPS
- Number of identified objects for the back, front, left and right position extracted from the Velodyne point cloud
- Distance of the nearest identified object for the back, front, left and right position extracted from the Velodyne point cloud
- Time headway extracted from the Velodyne point cloud

We analyze each physiological stress index, comparing them with each driving feature obtaining 4 x 20 levels of correlations.

4.4.1 Event sequence identification

We needed, first of all, to identify the event sequence in environmental and vehicle dynamics data. In a time series, the event sequence was identified by searching when the time series values are above/below a threshold. Since we did not know which thresholds to use to define the events, we used percentiles concerning all single acquisition. The percentiles gave us the possibility to test more easily different events threshold for a time series, identifying,

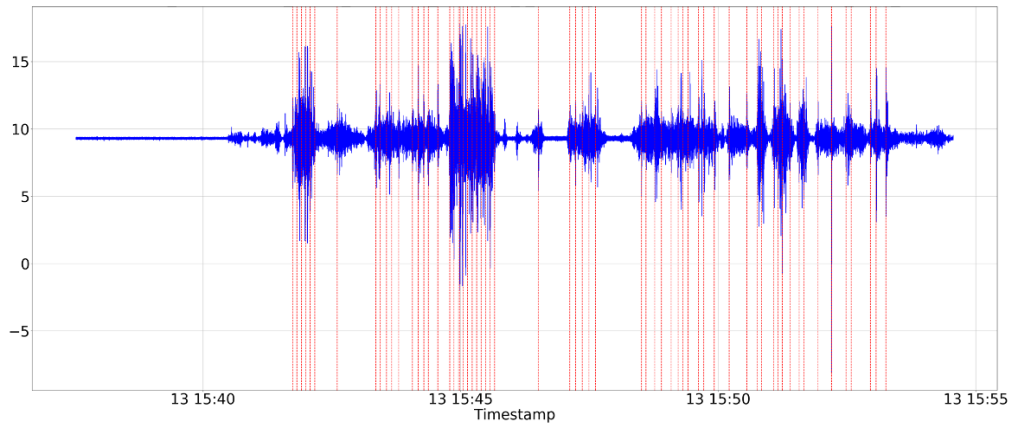


Figure 4.15: An example of the events sequence (the red lines) extracted from the Xsens linear acceleration on the x-axis (blue line) using the 99th percentile that correspond to an acceleration of 11.17 m/s^2

subsequently, the thresholds that maximize/minimize the stress level of the driver/passenger. The algorithm used to obtain the event sequence from our time series for a given percentile is defined in Algorithm 1. The events returned by the Algorithm 1 had a minimum time between them since, without this minimum time, we could have many events in a small time interval that did not give us useful information. In Figure 4.15 an example of the events sequence extracted from the Xsens linear acceleration on the x-axis using the 99th percentile that correspond to an acceleration of 11.17 m/s^2.

Suppose the events sequence and a stress index time series have a correlation relationship. Every time an event happened, it is possible to see a corresponding change of the time series. To study the potential temporal relationships between an event's occurrence and its corresponding change of the time series, we compared the sub-series before and after the event occurred.

The physiological features do not respond instantaneously to a stressful event but have a delay or an advance in the response time. To take into account the delay or the advance response time of the stress indexes, we used some temporal parameter to select the rear and front sub-series:

- **Latency**: defined as the time between the end of the event and the start of the effect on the physiological measure, if 0 it means that the

Algorithm 1 Transformation of time series in event sequence

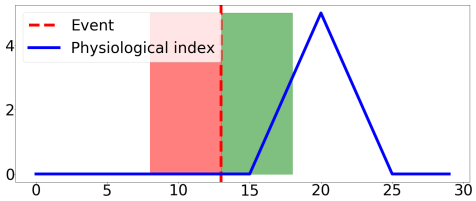
require time-series (TS), Percentile, PercentileStep, minimum time between two events from 0 to T (min)
return Events

```
define Events = [ ]
if Percentile > 50 then
    CurrentPercentile = 100
    while CurrentPercentile > Percentile do
        for i in 0...T do
            if TS[i] >= CurrentPercentile of TS then
                Events ← i
                i+ = min
            end if
        end for
        CurrentPercentile − = PercentileStep
    end while
else
    CurrentPercentile = 0
    while CurrentPercentile < Percentile do
        for i in 0...T do
            if TS[i] <= CurrentPercentile of TS then
                Events ← i
                i+ = min
            end if
        end for
        CurrentPercentile + = PercentileStep
    end while
end if
Drop Duplicates of Events
```

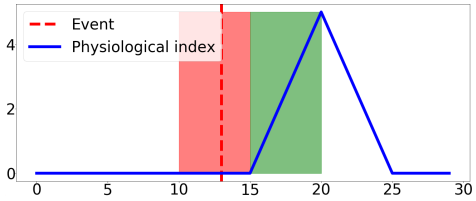
change in the physiological index is considered immediately when the event occurs.

- **Duration**: defined as the time window in which a physiological index continue to change. The rear and the front sub-series must have the same duration.

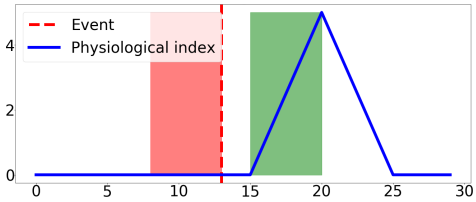
- **Shift**:defined as a temporal shift of the event occurrence. It can be ei-



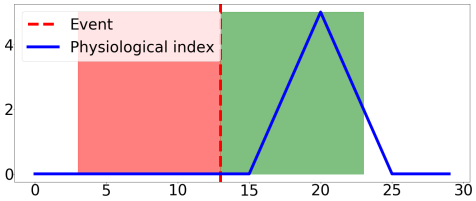
(a) 0s latency, 0s shift, 5s duration



(b) 0s latency, +1s shift, 5s duration



(c) 1s latency, 0s shift, 5s duration



(d) 0s latency, 0s shift, 10s duration

Figure 4.16: Example of temporal parameters. Front sub-series is defined by a red color and rear sub-series by green color

ther negative which means that there is an anticipatory response in the physiological indexes (for example before breaking) or positive which menas that the physiological indexes changes after a certain amount of time.

In Figure 4.16 is shown an example of latency, duration and shift use.

We used the Algorithm 2 to obtain the correlation relationship between an events sequence and a time series using different percentile, latency, duration and shift. For each event we applied different parameter to extract different front and rear sub-series and, using another algorithm, we search if there is a change between the front and the rear sub-series due to the event occurred. We applied the same procedure for different percentile.

4.4.2 Correlation algorithms

The algorithm used for correlation analysis had to detect a change in the physiological stress time series without knowing precisely what this change means. Changes in our problem affect the whole system since, after a change, the system does not return to the initial state but to a changed one. Indeed, if we considered T_0 as the driver's initial state of body stress, when a stressful event occurs, due to short-term stress, the state change in T_1. After a certain

Algorithm 2 Extraction of correlation relationship between an events sequence and a time series using different percentile, latency, duration and shift

require physiological time-series (TS), environmental or vehicle time-series (TS_D)
result Correlations

define Arrays of percentiles[0...P],latencies[0...L],shifts[0...S] and durations[0...D], Correlations=[ ]
for percentile in percentiles **do**
  $Events = Algorithm\ 1(TS\_D, percentile)$
  for latency in latencies **do**
    for shift in shifts **do**
      for duration in durations **do**
        $FrontStart = -duration + shift$
        $FrontEnd = +shift$
        $RearStart = +shift + latency$
        $RearEnd = +duration + shift + latency$
        for event in Events **do**
          $Front = TS[event + FrontStart : event + FrontEnd]$
          $Rear = TS[event + RearStart : event + RearEnd]$
          $Correlations \leftarrow Algorithm(Front, Rear)$
        end for
      end for
    end for
  end for
end for

amount of time, depending on the person and which physiological index we are considering, the influence of short-term stress decreases, but longer-term stress could remain, returning to a stress state $T_{0'}$ that is different from the initial one. However, in our study, we considered only the short-term stress and not the influence of longer-term stress. For this reason, we defined the time between events to be equal or greater to the duration of short-term stress.

Furthermore, in our problem, we extracted only the direction of change (increase or decrease of stress) and not the magnitude of change, since it depends on the person. So our correlation algorithm needed to consider each

occurred event as a single event, searching for an increase or a decrease of stress.

For the correlation study between rear and front sub-series, we compared the performance in terms of accuracy and computational complexity of four different algorithms: Pearson Correlation Coefficient (PCC), Auto-Regressive Moving Average (ARMA), Nearest Neighbor Statistic Based Method (NNSBM) and energy test.

Pearson Correlation Coefficient

The Pearson Correlation Coefficient is a method that measures the correlation between two time series X and Y. The PCC dentoted as ρ, is calculated as follows:

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}$$

where cov is the covariance, μ_X and σ_X are the mean and the standard deviation of X, μ_Y and σ_Y are the mean and the standard deviation of Y and $E[*]$ denotes the expected value.

The Pearson Correlation Coefficient ranges from -1 to 1. A value of $\rho = 1$ implies a total linear correlation between X and Y, while a value of $\rho = -1$, implies a total negative linear correlation. A value of $\rho = 0$, implies no linear correlation. So the more the value of ρ is near to 0, the weaker is the correlation between X and Y.

PCC is weak to outliers and might not be efficient if the population is not normal.

Auto-Regressive Moving Average

The Auto-Regressive Moving Average (ARMA) model are widely used in forecasting of time series. In our study, we compare prediction based on front sub-series data with the rear sub-series data.

The AR (Auto-regressive) part involves regressing the variable on its own past values. The MA (Moving Average) part involves modeling the error term as a linear combination of error terms occurring contemporaneously and at various times in the past. Given a time series X_t an ARMA(p,q) model is defined as:

$$X_t - \alpha_1 X_{t-1} - \ldots - \alpha_{p'} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_p \varepsilon_{t-q}$$

To compare the prediction based on the front sub-series with the actual rear sub-series we used the Root Mean Square Error (RMSE) defined as:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(Xpred_i - Xrear_i)^2}{n}}$$

where $Xpred$ is the prediction value, $Xrear$ the actual value of the rear sub-series and n is the length of the sub-series. To define the parameters of the ARMA model, we compared results using different parameters.

Nearest Neighbor Statistic Based Method

With the Nearest Neighbor Statistic Based Method (NNSBM) [35] [26] our correlation analysis problem can be transformed to a multivariate two-sample hypothesis-testing problem. Let $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ be independent random samples of $\Re^d$ from unknown distributions $F(x)$ and $G(x)$, respectively, with corresponding densities $f(x)$ and $g(x)$. The hypotheses of the two-sample problem can be stated as follows:

$$\begin{cases} H_0 : & F(x) = G(x) \\ H_1 : & F(x) \neq G(x) \end{cases}$$

If the H_1 is true, that means that the probabilistic distribution $F(x)$ is statistically different from the $G(x)$ distribution.
Given $\Omega_1 = \{1, \ldots, n_1\}$ and $\Omega_2 = \{n_1 + 1, \ldots, n\}$ and label the pooled samples as $Z_1, \ldots, Z_n$ with $n = n_1 + n_2$ where

$$Z_i = \begin{cases} X_i & i \in \Omega_1 \\ Y_{i-n_1} & i \in \Omega_2 \end{cases}$$

With the $NN_i(r)$ that represent the r-th nearest neighbour to the sample point Z_i we define:

$$I_i(r) = \begin{cases} 1 & \text{if } NN_i(r)\text{belongs to same sample as } Z_i \\ 0 & \text{otherwise} \end{cases}$$

In order to test our hypothesis, we considered the statistic:

$$T_{k,n} = \frac{1}{nk} \sum_{i=1}^{n} \sum_{r=1}^{k} I_i(r)$$

which is the proportion of all k nearest neighbor comparisons in which a point and its neighbor are members of the same sample. The $T_{k,n}$ is small under the null hypothesis H_0 and large when the two sub-series are different.

Energy test

The energy test [40] is a statistical distance between two observations. The energy test is based on Newton's gravitational potential energy. The idea of energy statistic is to consider statistical observations as bodies governed by a statistical potential energy, which is zero if and only if null hypothesis is true. In our study, we used the two sample energy statistic corresponding to the energy distance $\varepsilon(X, Y)$ for independent random samples $X = X_1, \ldots, X_{n_1}$ and $Y = Y_1, \ldots, Y_{n_2}$:

$$\varepsilon_{n_1,n_2}(X,Y) = \frac{2}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{m=1}^{n_2}|X_i - Y_m| - \frac{1}{n_1^2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_1}|X_i - X_j| - \frac{1}{n_2^2}\sum_{l=1}^{n_2}\sum_{m=1}^{n_2}|X_l - X_m|$$

The statistic:

$$T_{n_1,n_2} = (n_1 n_2/(n_1 + n_2))\varepsilon_{n_1,n_2}$$

is applied for testing homogeneity (i.e. the equality of distributions X and Y). The H_0 null hypothesis is rejected for large values of T_{n_1,n_2}.

4.4.3 Correlation algorithms comparison

For the comparison of the correlation algorithm, we used sub-series extracted from our test acquisition with and without abrupt changes. We compare the accuracy and complexity of each algorithm.

To evaluate the algorithms' accuracy, we applied them to different sub-series in which we knew if there was an abrupt change or not. We extracted the ROC curve, measuring the true positive rate (TPR) and the false positive rate (FPR) of the algorithm output. We calculated the TPR and FPR as:

$$TPR = TP/(TP + FN)$$

$$FPR = FP/(FP + TN)$$

AUC range	Accuracy classification
$0.9 < AUC < 1.0$	Excellent
$0.8 < AUC < 0.9$	Good
$0.7 < AUC < 0.8$	Worthless
$0.6 < AUC < 0.7$	Not Good

Table 4.1: Accuracy classification using the AUC value

p\q	**0**	**1**	**2**	**3**	**4**	**5**
0	-	0.75	0.56	0.52	0.56	0.77
1	0.78	0.79	0.80	0.75	0.70	0.64
2	0.79	0.79	0.81	0.69	0.63	0.56
3	0.84	0.85	0.87	0.82	0.79	0.61
4	0.79	0.83	**0.90**	0.77	0.70	0.59
5	0.77	0.85	0.82	0.74	0.72	0.61

Table 4.2: AUC results for the ARMA(p,q) model

where TP represents the true positive, FN represents the false negative, FP represents the false positive and TN represent the true negative.
To generate a summary statistic from the ROC curve, we calculated the area under the curve called AUC as:

$$AUC = \int_0^1 ROC(t)dt$$

The closer to 1 is the AUC, the more accurate the algorithm is. A value of 0.5, instead, provides an accuracy equal to a random test. In Table 4.1 it is possible to see the accuracy classification for each AUC range.

For the ARMA(p,q) algorithm, we calculate the AUC value for different p and q parameter (from 0 to 5) to identify the best parameters value to use in our study. In Table 4.2 it is possible to see the AUC value for each pair of p and q parameters. The best results had been obtained with the ARMA(4,2) model with an AUC of 0.90.
Also for the NNSBM we calculated the AUC value for different values of the k parameter (from 1 to 10). In Figure 4.17 is possible to see how AUC value change with the different k parameters. The best result was obtained using $k = 4$ with a AUC of 0.889.
For the Pearson Correlation Coefficient we obtained a AUC value of 0.77

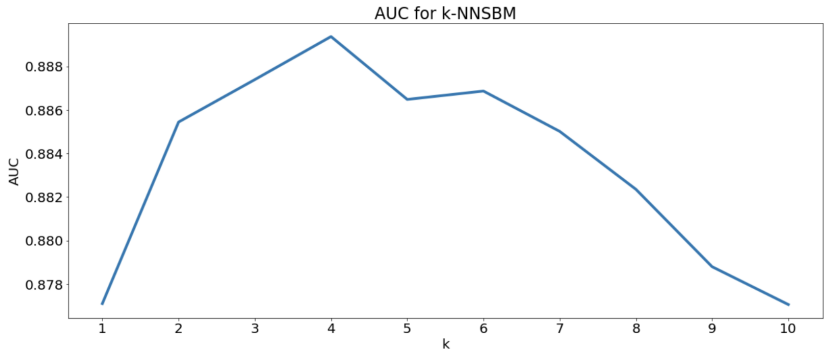


Figure 4.17: AUC value trend for different k values of the NNSBM algorithm

which result in a worthless accuracy level.
For the energy test we obtained a AUC value of 0.91 that is an excellent result.
In Figure 4.18 it is possible to see the ROC curves with the relative AUC values of all the algorithm analyzed. The 4-NNSBM, ARMA(4,2) and energy test algorithms obtained similar accuracy results while PCC obtained a very low accuracy.

Furthermore, we considered the complexity of the algorithms based on a series of length n. The complexity of the algorithms are;

- **PCC**: $O(n)$
- **ARMA(p,q)**: the complexity of this model depends on the value of the parameters (p,q) and it is $O((p+q)n^2)$
- **k-NNSBM**: the complexity of the algorithm depends on parameter k and it is $O(k^2 n)$
- **Energy test**: $O(n^2)$

Despite the lowest complexity, the PCC algorithm had a too low accuracy to be used. The other algorithms showed a similar level of accuracy. However, the NNSBM algorithms turned out to be the most efficient, so we decided to use this algorithm in our analysis.

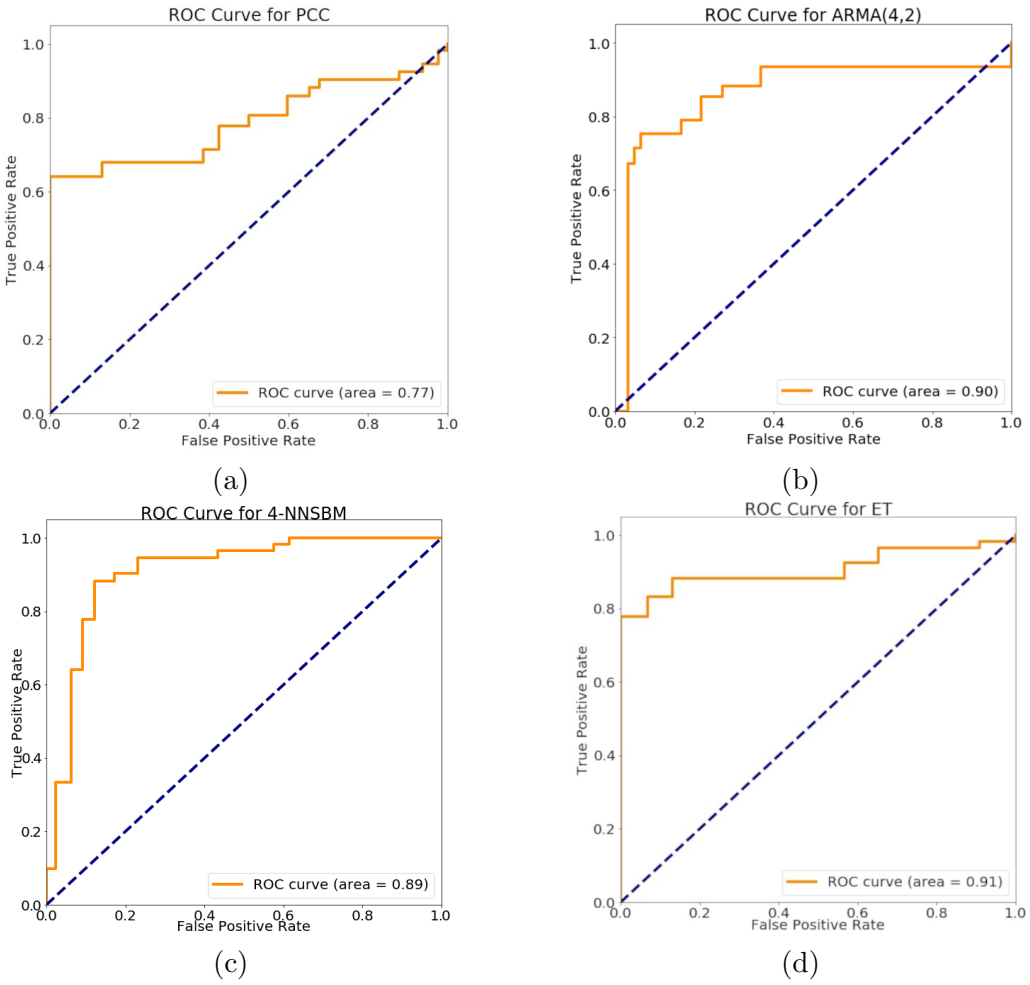


Figure 4.18: ROC curves and AUC values for the algorithm: (a) PCC, (b) ARMA(4,2), (c) 4-NNSBM, (d) energy test

4.4.4 Application of correlation algorithms

Applying the 4-NNSBM algorithm only to the rear and front sub-series might not be enough to understand if there is a change in the time series due to the event occurred [26]. To better understand if there was a change in the time series due to an event, we need to find the average correlation for every considered time-series using it as a reference threshold.

To compute the average correlation for a time-series, we randomly sam-

Algorithm 3 Computation of the average level of correlation for a time series

require TS = time-serie with timestamps from 0 to T, latency, shift, duration, maxIterations, minIterations, maxConfidence
result *AverageCorrelation*

define Correlations=[ ]
$FrontStart = -duration + shift$
$FrontEnd = +shift$
$RearStart = +shift + delay$
$RearEnd = +duration + shift + delay$
for $iterations\ in[0 \dots maxIterations]$ **do**
    $r = Random\ i \in [0 \dots T]$
    $Front = TS[r + FrontStart : r + FrontEnd$
    $Rear = TS[r + RearStart : r + RearEnd]$
    $Correlations \leftarrow NNSBM\_Algorithm(Front, Rear)$
    if $iterations > minIterations$ **then**
        if $Confidence(Correlations) \leq maxConfidence$ **then**
            break loop
        end if
    end if
end for
$AverageCorrelation = average(Correlations)$

pled the considered time series and, applying the 4-NNSBM algorithm to each random front and rear sub-series, we obtained correlation values that were used to compute an average level of correlation. This algorithm is showed in Algorithm 3, where it is also possible to see that this algorithm have a minimum number of iteration (30). Algorithm 3 stops when a maximum number of iteration is reached (5% of all population) or when the confidence interval is sufficiently small (Figure 4.19). In Figure 4.20 there is an example of the correlation average computed using our algorithm on different time series.

Once we had computed the average correlation of random samples of a time series, we needed to understand if, on average, the time-series increased or decreased.

We applied the Algorithm 4 that compare the average correlation computed using the Algorithm 3 with a correlation of new random samples. If the correlation of the new random samples is below the average correlation,

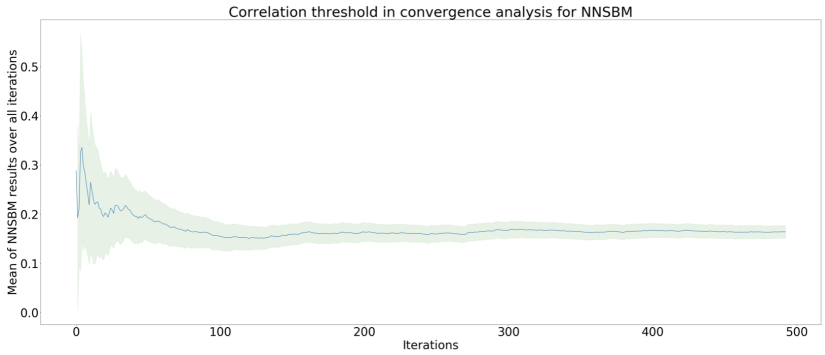


Figure 4.19: Example of average correlation's trend on different iterations with the confidence interval

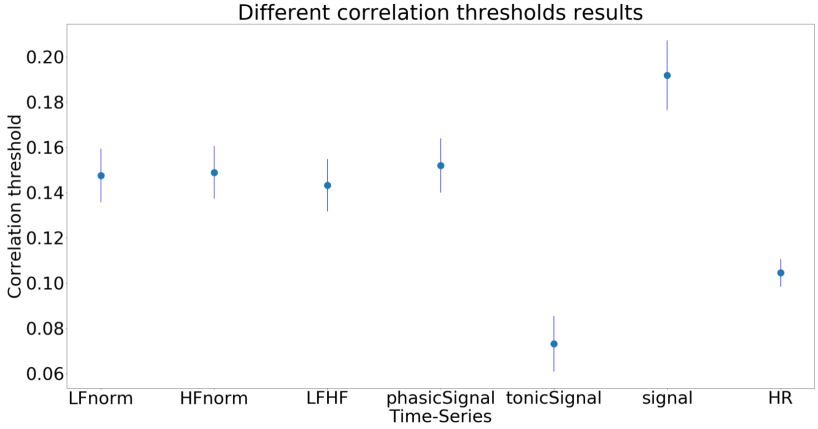


Figure 4.20: Average level of correlation and confidence interval for different time-series computed with Algorithm 3

it means that there is a change in the values between the front and rear sub-series. To identify whether there is a significant value increase or decrease from front to rear sub-series, we used the t-test. The t_{score} between front and rear is computed by the following equation:

$$t_{score} = \frac{\mu_{front} - \mu_{rear}}{\sqrt{\frac{\sigma^2_{front} + \sigma^2_{rear}}{n}}}$$

where μ_{front} and μ_{rear} are the mean values of front and rear sub-series. σ_{front} and σ_{rear} are their variance values. If $t_{score} < -\alpha$ we have a posi-

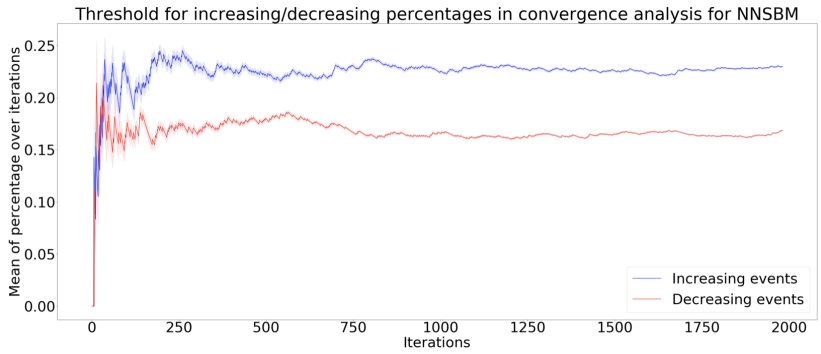


Figure 4.21: Example of average increasing and decreasing probabilities trend on different iterations with the confidence interval

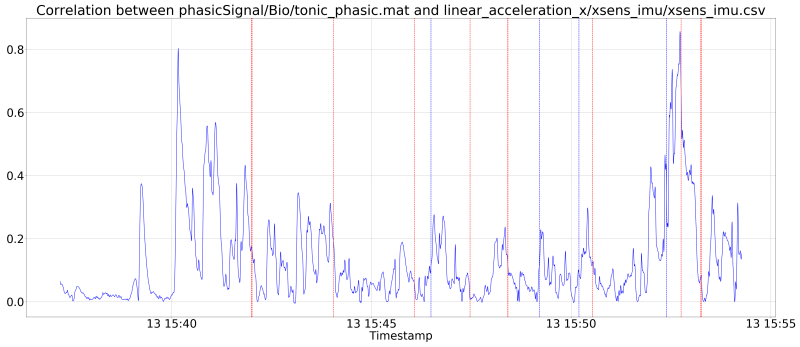


Figure 4.22: Example of events extracted from the linear acceleration on x axis with the information on whether each event has increased (blue vertical lines) or decreased (red vertical lines) the phasic signal time series

tive monotonic effect and, if $t_{score} > \alpha$ we have a negative monotonic effect. If $t_{score} < |\alpha|$ we can not determine the effect type with a high level of confidence.
Algorithm 4 stops when a maximum number of iteration is reached or when the confidence interval is sufficient small (Figure 4.21).

Afterward, we applied the Algorithm 4 but, instead of using random samples, we used the events sequence extracted using the Algorithm 2. In Figure 4.22 there is an example of events sequence extracted using Algorithm 2 with the information retrieved using Algorithm 4 on whether each event has increased or decreased the time series.

In order to compare the average correlation of increasing and decreasing

Algorithm 4 Extraction of average probability that a time series increase or decrease

require TS = time-series with timestamps from 0 to T, latency, shift, duration, α, maxIterations, minIterations, maxConfidence
result $AverageIncreasing, AverageDecreasing$

define Increasing = [ ], Decreasing = [ ]
$AverageCorrelation = Algorithm\ 3(TS)$
$FrontStart = -duration + shift\ FrontEnd = +shift$
$RearStart = +shift + delay\ RearEnd = +duration + shift + delay$
for $iterations\ in [0 \dots maxIterations]$ **do**
    $r = Random\ i \in [0 \dots T]$
    $Front = TS[r + FrontStart : r + FrontEnd]$
    $Rear = TS[r + RearStart : r + RearEnd]$
    $Correlation = Algorithm(Front, Rear)$
    if $Correlation < AverageCorrelation$ **then**
        compute t_{score}
        if $t_{score} > \alpha$ **then**
            $Increasing \leftarrow 1\ Decreasing \leftarrow 0$
        else
            if $t_{score} < -\alpha$ **then**
                $Increasing \leftarrow 0\ Decreasing \leftarrow 1$
            else
                $Increasing \leftarrow 0\ Decreasing \leftarrow 0$
            end if
        end if
    else
        $Increasing \leftarrow 0\ Decreasing \leftarrow 0$
    end if
    if $iterations > minIterations$ **then**
        if $Confidence(Increasing) \leq maxConfidence$ **then**
            if $Confidence(Decreasing) \leq maxConfidence$ **then**
                break loop
            end if
        end if
    end if
end for
$AverageIncreasing = average(Increasing)$
$AverageDecreasing = average(Decreasing)$

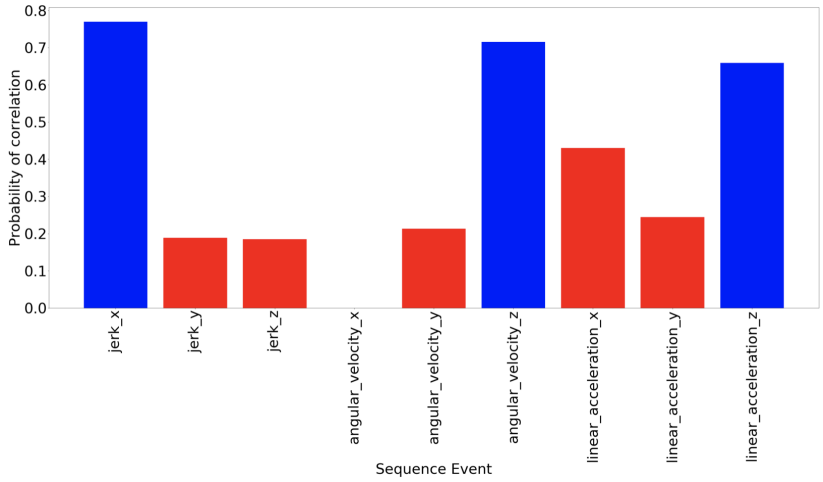


Figure 4.23: Example of probability that the average correlation of increasing event sequence is higher than the average correlation of increasing random samples

events sequence with the average correlation of increasing and decreasing random samples, we used the Welch's t-test [43] since our samples have unequal size and unequal variance. The Welch's t-test formula is:

$$t = \frac{\mu_X - \mu_Y}{\sqrt{\frac{\sigma_X^2}{n_1} + \frac{\sigma_Y^2}{n_2}}} \qquad df = \frac{(\frac{\sigma_X^2}{n_1} + \frac{\sigma_Y^2}{n_2})^2}{\frac{\sigma_X^4}{n_1^2(n_1-1)} + \frac{\sigma_Y^4}{n_2^2(n_2-1)}}$$

where μ_X and μ_Y are the mean of the samples, σ_X σ_Y are the standard deviation of the samples, n_1 and n_2 are the samples size, t is the t statistic and df is the degree of freedom.
Using Welch's t-test, we have the probability that the average correlation of increasing and decreasing events sequence is higher than the average correlation of increasing and decreasing random samples. In figure Figure 4.23, it is possible to see an example of the output of Welch's t-test for increasing events sequence.
Figure 4.24 showed a diagram of all the data analysis steps that we performed from the time series to the probability of correlation.

Afterward, we studied how the results obtained from the whole procedure with respect to percentiles since we had to understand how different

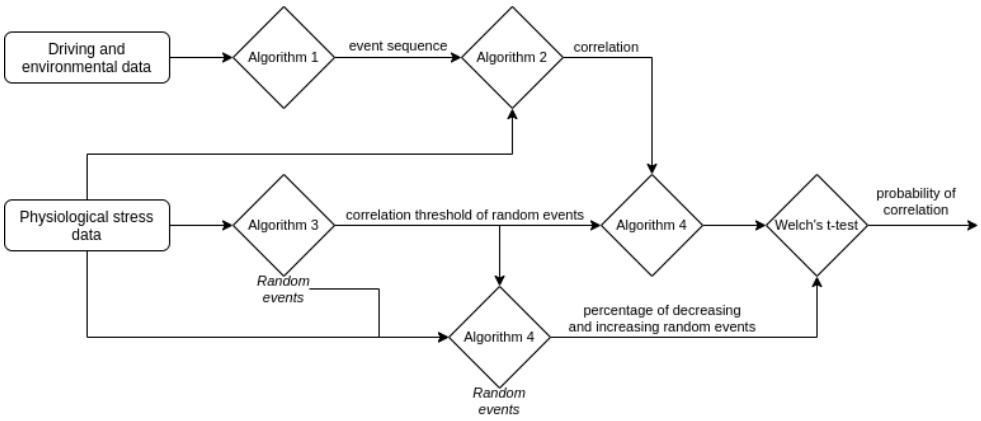


Figure 4.24: Diagram of the data analysis steps from the time-series to the correlation probability

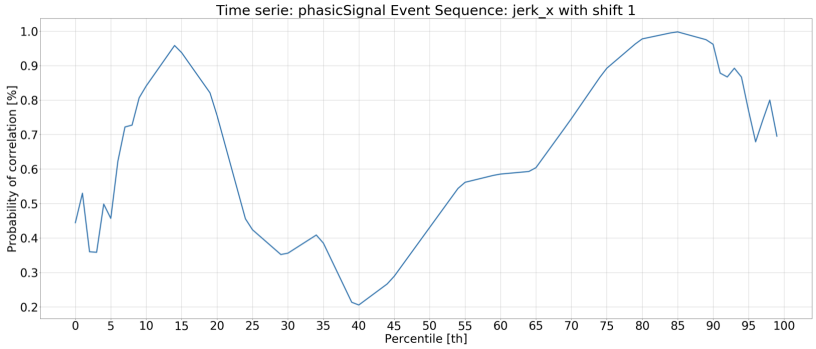


Figure 4.25: Example of probability of correlation trend with respect to percentile for the events sequence of jerk on x axis applied to the phasic component of the skin conductance

percentiles influence the probability obtained. In Figure 4.25 is showed an example of correlation probability with respect to percentiles.

Finally, to understand if a driving style feature impacts driver's stress, we compared the probability of correlation curves with respect to percentiles (Figure 4.25) with two curves. One curve identifies a stress-relevant feature. It starts from a correlation probability equal to 1 at the 100th and the 0th percentile and decreases to 0.5 at the 50th percentile. The no-stress-relevant curve, instead, is a flat line with a probability value of 0.5. We use the root mean square error to understand whether a curve is nearer to the stress-

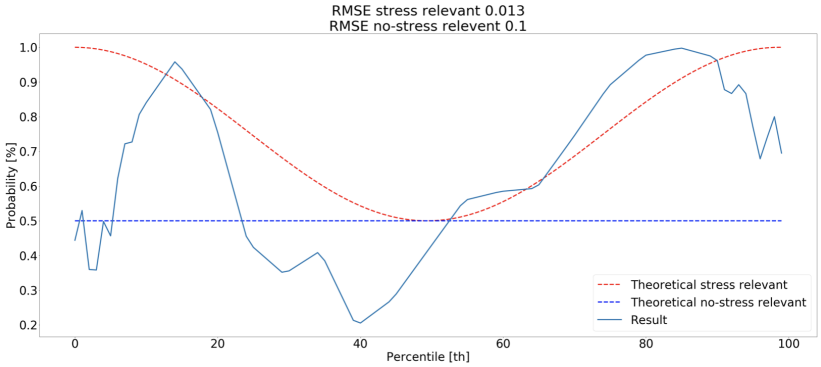


Figure 4.26: An example of probability of correlation trend compared to the stress-relevant (RMSE=0.013) and no-stress-relevant (RMSE=0.1) reference curves

relevant or the no-stress-relevant reference curves. In Figure 4.26 is it possible to see the probability curve of Figure 4.25 compared with the stress/no-stress reference curves. We use the ratio between RMSE as following:

$$stress\_ratio = \frac{RMSE(no\text{-}stress\text{-}relevant)}{RMSE(stress\text{-}relevant)}$$

This ratio gave us information about how much the correlation probability curve is near the stress-relevant curves with respect to the no-stress-relevant curves. A good ratio must be at least 2 (higher is better) to have reasonable confidence that the curve is much closer to stress relevant than no-stress relevant reference.

C H A P T E R

5. Experiments

We tested all the hardware and software developed on field acquisitions. These acquisitions aimed to validate the entire system pipeline using questionnaires and looking at the results obtained from the correlation analysis. This chapter describes the preliminary test conducted to see that the platform developed worked properly, showing preliminary results,and a field campaign showing the final setup used, the protocol applied for the acquisition, and the results. Furthermore, we discuss the results obtained.

5.1 Preliminary test

Before starting with the acquisition campaign, we made an on-road acquisition to check that all our hardware and software worked properly. This acquisition was performed on a single driver in a manual drive settings on the route in Figure 5.1. We acquired all I.DRIVE vehicle sensors except for the respiration sensor that was added after this test. After we checked that all the sensors had been acquired properly without loosing any data, we use the acquisition data to test the analysis software.

Firstly we tuned the duration, latency and shift parameters (Subsection 4.4.1) for each physiological index that we used. To compare and select the best parameters to use in the data analysis, we searched the parameters that maximize the *stress_ratio*. We tested duration values from 1s to 20s, latency values between 0s and 10s, and shift values from -5s to 5s. All the following *stress_ratio* graph are relative to the event sequence extracted from the jerk on x-axis since, in literature, the longitudinal jerk was reported to be a feature that has a high impact on the driver's stress level[7].

As we can notice from Figure 5.2 we have the *stress_ratio* values of ratio

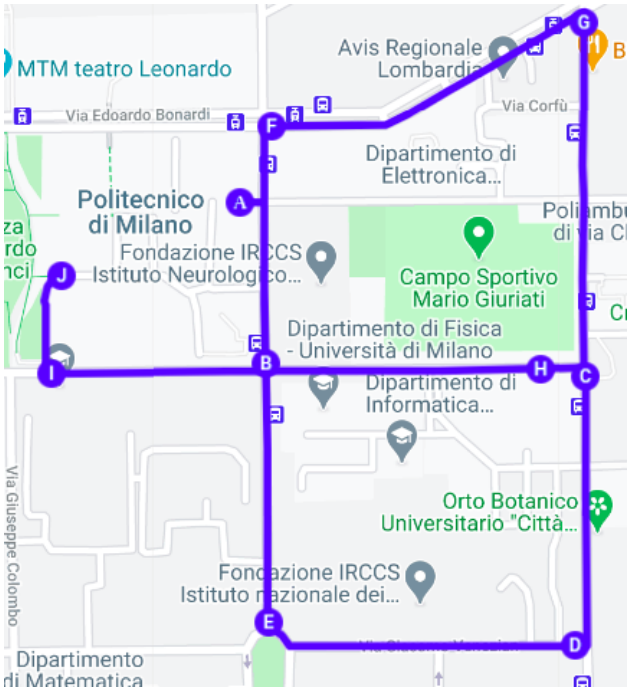


Figure 5.1: The test track route

between low frequency (LF) and high frequency (HF) (LF/HF) for different duration. We have a peak for $duration = 4s$ with a *stress_ratio* below 1. Since a good *stress_ratio* must be at least 2 to have a good correlation probability, the LF/HF index is not considered as useful for our correlation study. For this reason, we did not analyze the shift and latency parameter for this stress index.

In Figure 5.3 we can see that the heart rate have a *stress_ratio* peak at 10s with a value above 2. Using the duration value of 10s, we analyzed the *stress_ratio* for the latency. In Figure 5.4 it is possible to see that we have a peak with a latency value of 2s. Finally, we studied the heart rate shift. The best *stress_ratio* value for the shift parameters is obtained for a value of 1s (Figure 5.5). For correlation analysis of the heart rate, we then used a duration value of 10s, a latency value of 2s, and a shift value of 1s.

In Figure 5.6 we report the *stress_ratio* values for the tonic component of the skin conductance according to duration values. We have 3 different peaks but none of them exceeds a *stress_ratio* value of 2. For this reason, also for this index, we did not go further on the analysis.

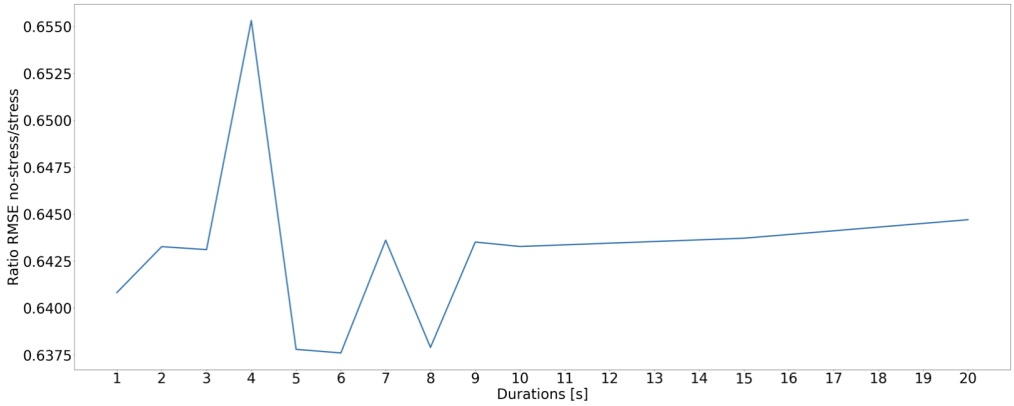


Figure 5.2: *Stress_ratio* of LF/HF for different duration values

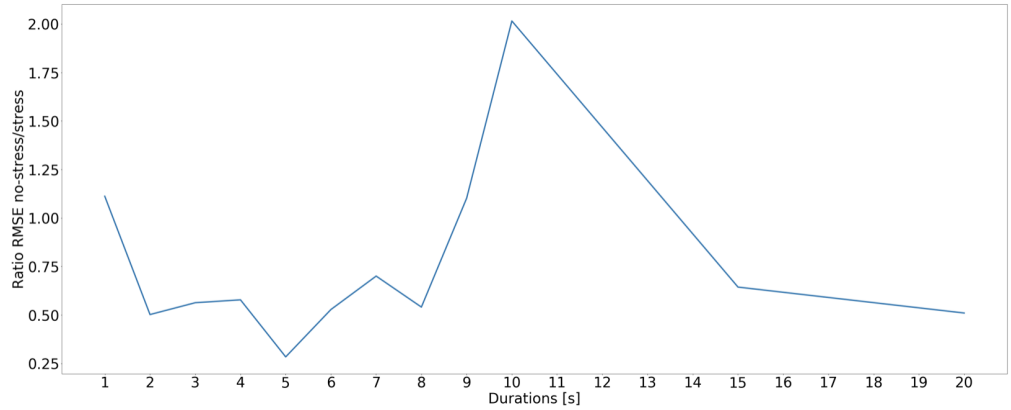


Figure 5.3: *Stress_ratio* of heart rate for different duration values

The skin conductance's phasic component had a peak at 5s with a *stress_ratio* above 3 (Figure 5.7). This *stress_ratio* value provides the information that, with a duration value of 5s this stress feature is useful to distinguish if an event is stressful. We studied the skin conductance's phasic component *stress_ratio* for latency values and we discovered that we had a peak with a latency of 3s (Figure 5.8). Finally, for the phasic component we studied the *stress_ratio* for different shift values (Figure 5.9). We found that the best shift value is 0s since, for this value, we have a *stress_ratio* peak. For correlation analysis of the skin conductance's phasic component we used a duration of 5s, a latency of 3s and a shift of 0s.

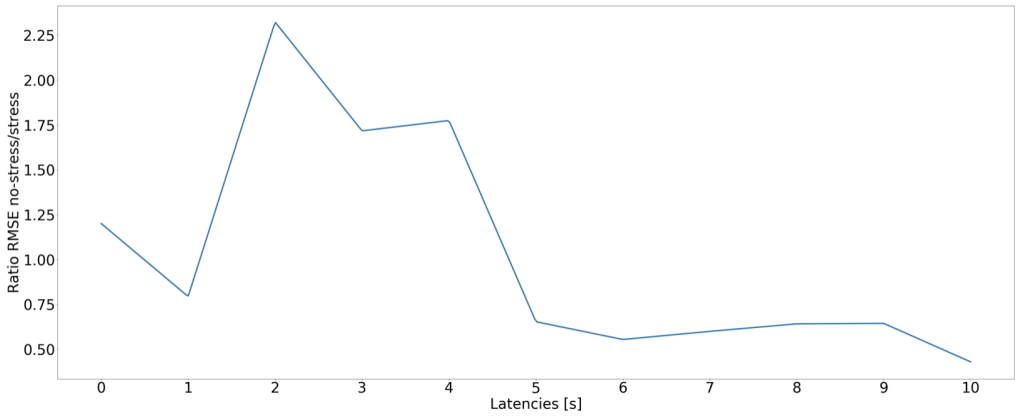


Figure 5.4: *Stress_ratio* of heart rate for different latency values

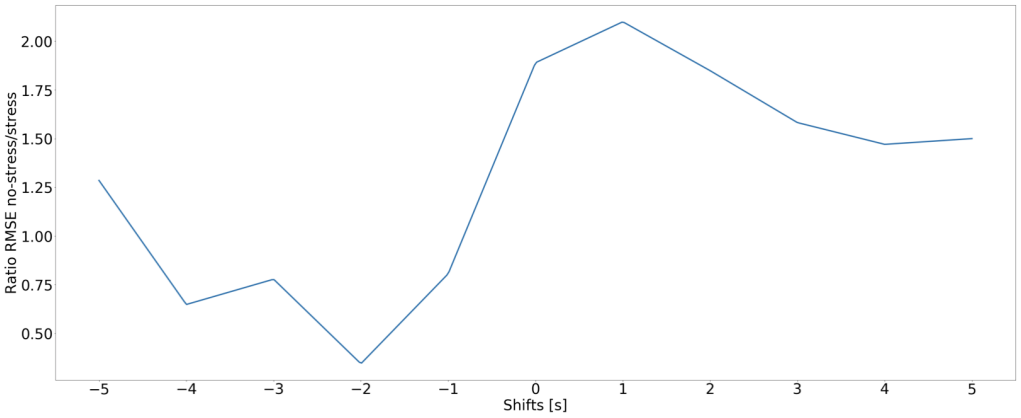


Figure 5.5: *Stress_ratio* of heart rate for different shift values

This time parameters analysis confirmed that the phasic component is one of the best physiological stress indexes as reported in the literature studies. We discovered that also the heart rate was a good estimator of the driver's physiological stress.

We proved that the ratio between low frequency (LF) and high frequency (HF) and the tonic component are not useful in our study for understanding the driver's stress. We expected a low *stress_ratio* for the skin conductance's tonic component. Indeed, although several studies have shown that changes in tonic component can be related to general changes of emotional state and level of stress, it changes slowly, and the measurement intervals have

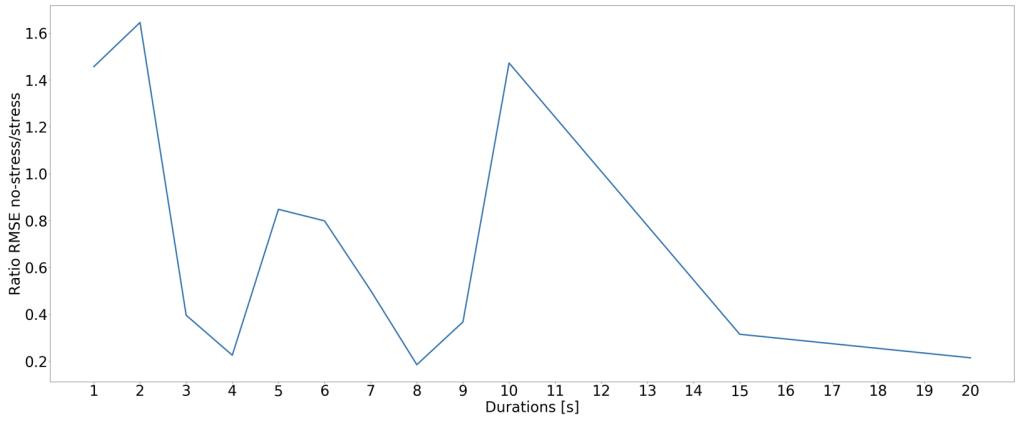


Figure 5.6: *Stress_ratio* of skin conductance's tonic component for different duration values

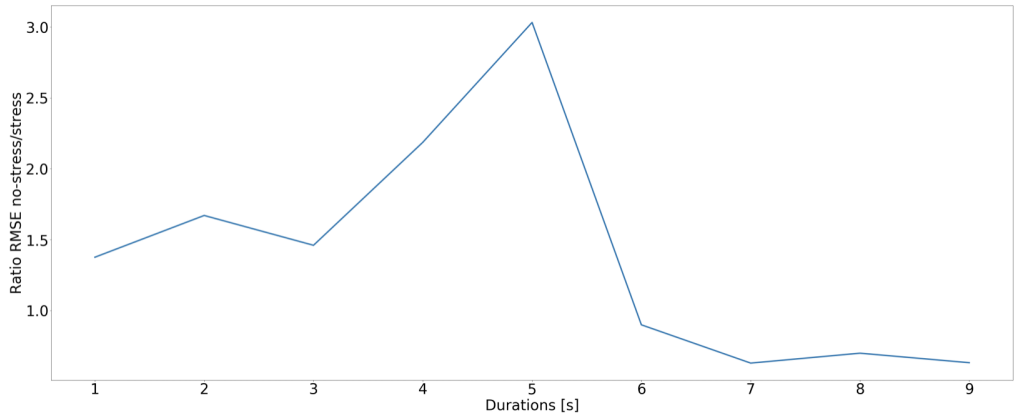


Figure 5.7: *Stress_ratio* of skin conductance's phasic component for different duration values

to be long, from 10s to minutes. The low *stress_ratio* value for the ratio between low frequency (LF) and high frequency (HF) could happen since it has computed on time windows of 2 minutes, and thus, it is not event-related. Besides, the frequency analysis could be corrupt by noise on the EKG signals that are not a problem for the heart rate computation.

Since in the analysis of physiological data we discovered that the only reliable stress indexes in our setup were the skin conductance's phasic component and the heart rate, we reported below the results obtained using only

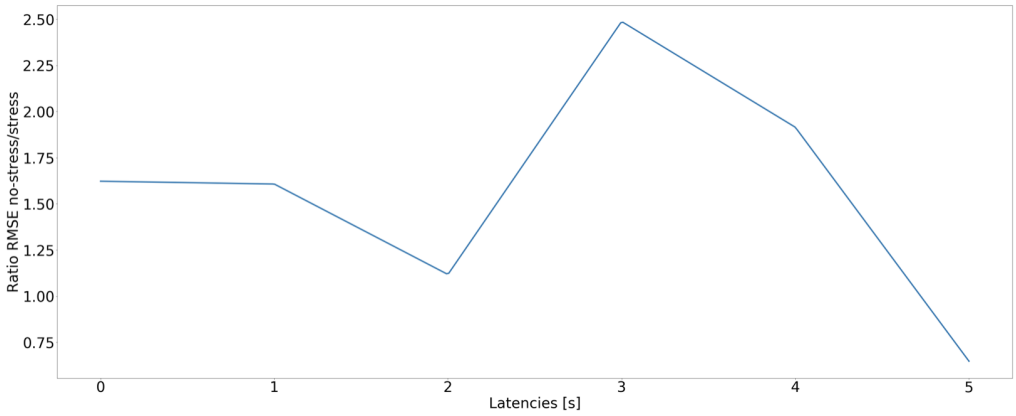


Figure 5.8: *Stress_ratio* of skin conductance's phasic component for different latency values

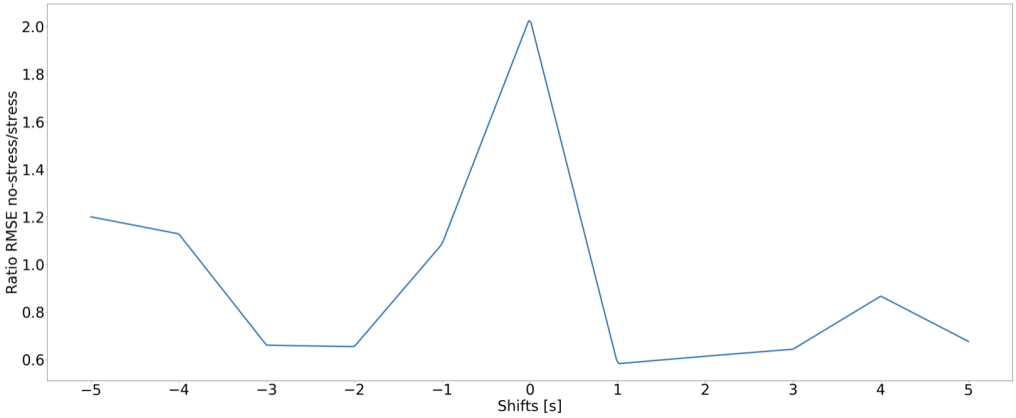


Figure 5.9: *Stress_ratio* of skin conductance's phasic component for different shift values

these stress features.

We computed, for each driving feature, the correlation probability according to different percentiles using the algorithm described in Section 4.4. We computed the *stress_ratio* separately for high percentiles (50th-100th) and low percentiles (0th-50th) since the correlation probability curves had different trends with respect to the stress-relevant and no-stress-relevant curves in the two percentile intervals. The different trend was due to the different meaning that high percentile and low percentile had, depending on the

Jerk X	**2.75**
Jerk Y	1.70
Jerk Z	1.59
Speed	1.27
Time headway pedestrians	—
Time headway cars	1.55
Angular velocity X	0.48
Angular velocity Y	1.47
Angular velocity Z	1.38
Linear acceleration X	1.64
Linear acceleration Y	0.54
Linear acceleration Z	1.63
Linear velocity X	**2.88**
Linear velocity Y	1.30

Table 5.1: *Stress_ratio* of driving features in correlation with skin conductance's phasic component at high percentiles

driving feature considered. For example, for longitudinal acceleration, low percentiles mean that the car is braking, and high percentiles mean that it is accelerating.

In Table 5.1 we report the *stress_ratio* at high percentile for each driving feature with respect to the phasic component. The only driving features that shows a *stress_ratio* above 2 are the jerk and linear velocity on x-axis. Time headway for the pedestrian is not available due to the few events that occurred during this acquisition. A high value of jerk on x-axis denotes a sudden increase in acceleration. A high jerk is frequent when the car starts from a standstill. High linear velocity, instead, is frequent on the straight road.

In Figure 5.10 and Figure 5.11 it is possible to see the correlation probability with respect the stress-relevant and no-stress-relevant curves for jerk and linear velocity at high percentiles. Seeing at the figures, it is possible to understand why the two curves have a high *stress_ratio*.

In Table 5.2 we report the *stress_ratio* at low percentile. As it is possible to see, the driving features that showed a *stress_ratio* value above 2 are different with respect to high percentiles. In this case, the only driving feature that has a high *stress_ratio* is the angular velocity on y-axis. The

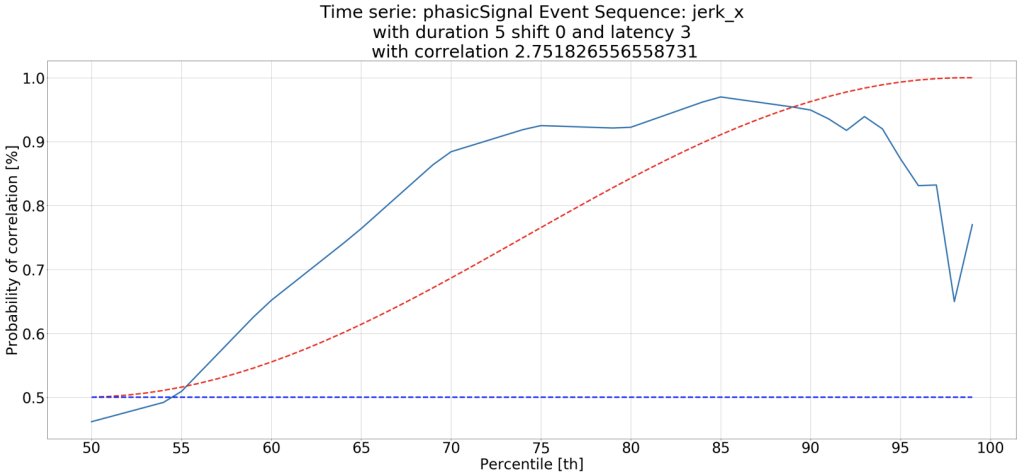


Figure 5.10: Probability of correlation for Jerk on x in correlation with skin conductance's phasic component at high percentiles (blue curve) with respect to stress-relevant (red dashed curve) and no-stess-relevant (blue dashed curve)

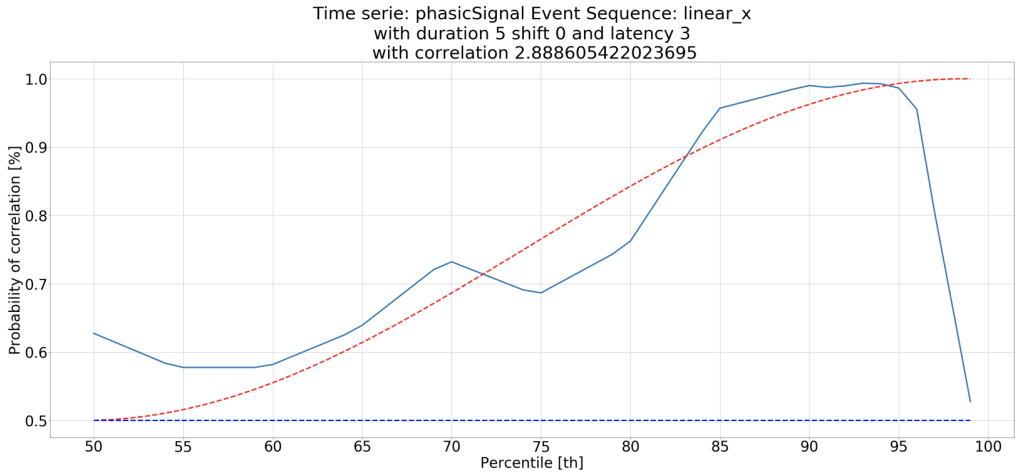


Figure 5.11: Probability of correlation for linear velocity on x in correlation with skin conductance's phasic component at high percentiles (blue curve) with respect to stress-relevant (red dashed curve) and no-stess-relevant (blue dashed curve)

Jerk X	1.34
Jerk Y	1.72
Jerk Z	0.75
Speed	0.46
Time headway pedestrians	—
Time headway cars	1.68
Angular velocity X	0.64
Angular velocity Y	**2.73**
Angular velocity Z	1.59
Linear acceleration X	1.03
Linear acceleration Y	0.67
Linear acceleration Z	1.71
Linear velocity X	1.71
Linear velocity Y	1.82

Table 5.2: Ratio of driving style's features in correlation with PC index at low percentiles

angular velocity on y-axis provides information on how fast the pitch angle changes. The pitch angle could change due to the road (downhill or uphill) or the suspension movement when the car is braking or accelerating. Since our test drives were performed on a flat road, the angular velocity on y-axis, in our study, change only due to the car's braking or accelerating. Figure 5.12 showed the correlation probability of the angular velocity on y-axis.

Table 5.3 showed the *stress_ratio* at high percentile for each driving feature with respect to the heart rate. The heart rate confirmed what we had found with the phasic component features. Indeed, considering the heart rate, the driving features that show a high impact on the stress are jerk on x-axis and linear velocity on x-axis. In Figure 5.13 and Figure 5.14 it is possible to see the correlation probability curves for jerk and linear velocity at high percentiles.

For low percentiles, the heart rate confirmed again what we discovered analyzing the phasic component. Indeed, the HR considered the angular velocity on y-axis as a stress factor (Table 5.4). Furthermore, the heart rate considered as a stress factor also Jerk on z-axis. The jerk on z-axis denotes a change in the vertical acceleration (perpendicular to the floor). Figure 5.15 and Figure 5.16 showed the correlation probability curve for the angular

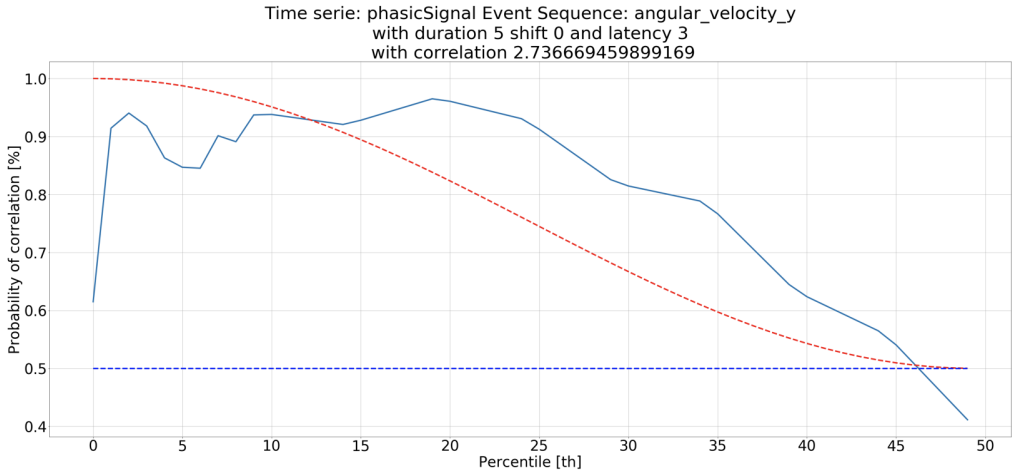


Figure 5.12: Probability of correlation for angular velocity on y in correlation with skin conductance's phasic component at low percentiles (blue curve) with respect to stress-relevant (red dashed curve) and no-stess-relevant (blue dashed curve)

Jerk X	**2.35**
Jerk Y	1.50
Jerk Z	1.95
Speed	1.09
Timeheadway pedestrians	—
Timeheadway cars	1.62
Angular velocity X	0.91
Angular velocity Y	0.48
Angular velocity Z	0.40
Linear acceleration X	0.45
Linear acceleration Y	0.54
Linear acceleration Z	0.35
Linear velocity X	**2.47**
Linear velocity Y	1.51

Table 5.3: Ratio of driving style's features in correlation with heart rate index at high percentiles

velocity on y-axis and jerk on z-axis with the heart rate at low percentiles.

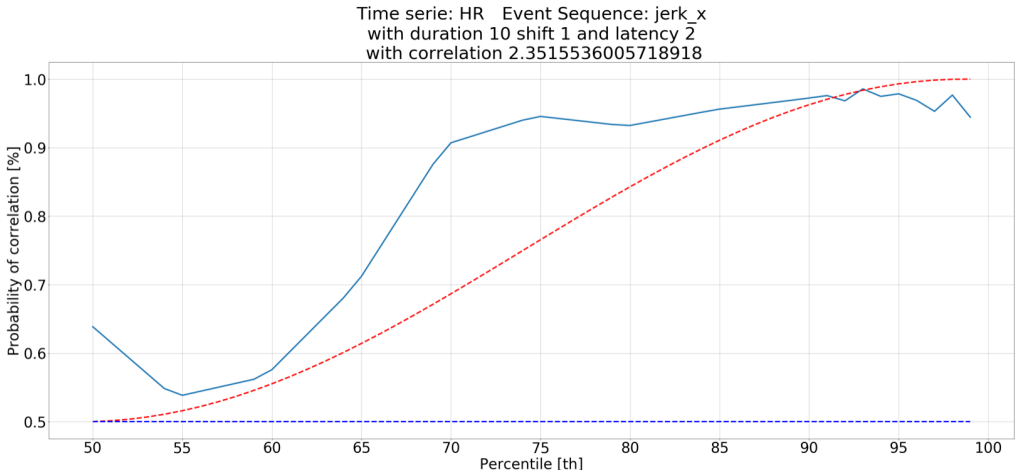


Figure 5.13: Probability of correlation for Jerk on x in correlation with the heart rate at high percentiles (blue curve) with respect to stress-relevant (red dashed curve) and no-stess-relevant (blue dashed curve)

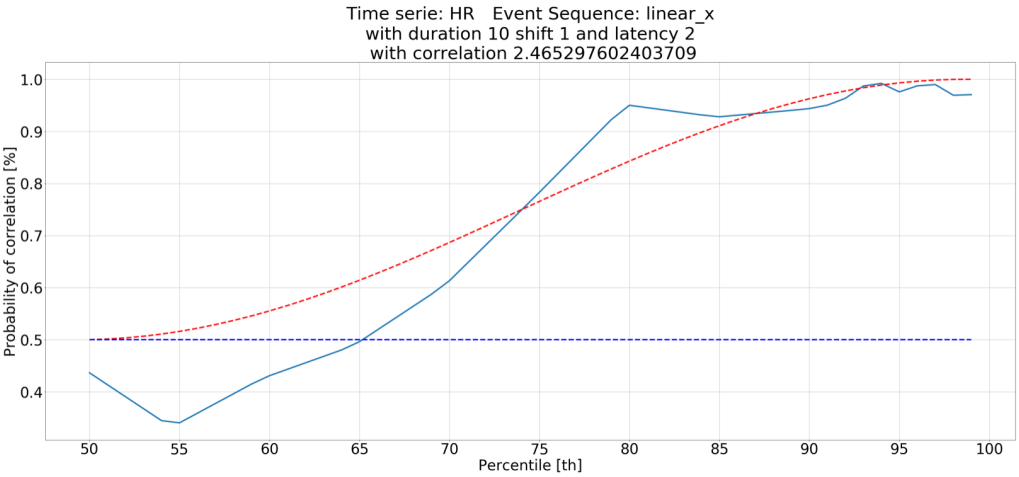


Figure 5.14: Probability of correlation for linear velocity on x in correlation with the heart rate at high percentiles (blue curve) with respect to stress-relevant (red dashed curve) and no-stess-relevant (blue dashed curve)

In this preliminary analysis, results demonstrate that not all the stress indexes computed in our study were reliable for detecting the driver's stress in the wild. We showed that only the heart rate extracted from EKG and

Jerk X	1.57
Jerk Y	1.44
Jerk Z	**2.36**
Speed	0.83
Timeheadway pedestrians	—
Timeheadway cars	1.12
Angular velocity X	0.57
Angular velocity Y	**2.16**
Angular velocity Z	0.42
Linear acceleration X	1.16
Linear acceleration Y	1.84
Linear acceleration Z	0.58
Linear velocity X	1.56
Linear velocity Y	1.59

Table 5.4: Ratio of driving style's features in correlation with heart rate index at low percentiles

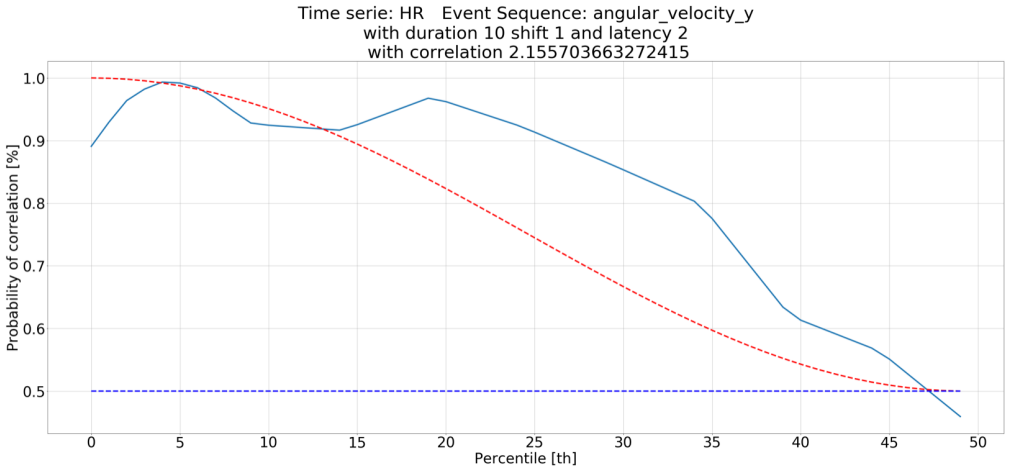


Figure 5.15: Probability of correlation for angular velocity on y in correlation with the heart rate at low percentiles (blue curve) with respect to stress-relevant (red dashed curve) and no-stess-relevant (blue dashed curve)

phasic component extracted from skin conductance are highly correlated with driving features. Instead, the LF/HF and the skin conductance's tonic com-

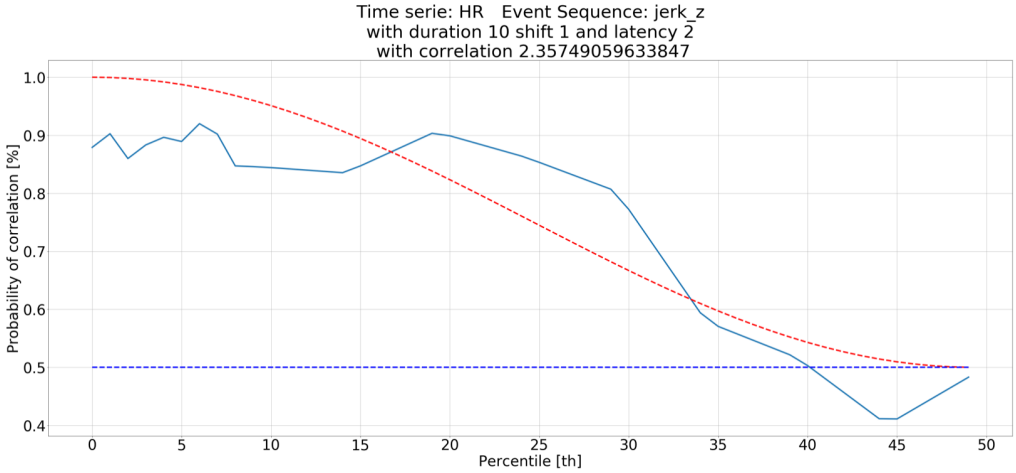


Figure 5.16: Probability of correlation for jerk on z in correlation with the heart rate at low percentiles (blue curve) with respect to stress-relevant (red dashed curve) and no-stess-relevant (blue dashed curve)

ponent measures returned a low correlation with driving events. We expected a low correlation for the skin conductance's tonic component but not for the LF/HF. Indeed, for the skin conductance's tonic component is it known that it changes slowly and, for this reason, the measurement intervals have to be long (from 10s to minutes). Since we focused on the event-related stress in our analysis, our measurement interval is too short (max 10s) to perceive a change in the tonic component. For the LF/HF, instead, several studies have shown that it is a valid stress index, but this study does not confirm it. The EKG signal may have too much noise due to the driver's movement or the driver's speaking. This noise could affect the frequency analysis and so the LF/HF quality. Furthermore, the LF/HF is computed on a window of 2 minutes and, for this reason, it could be not so event-related as we expected. Nevertheless, the LF/HF should be investigated further.

Because of the result obtained from this preliminary test, we decided to study in the acquisition campaign only the skin conductance's phasic component, the heart rate and the LF/HF.

5.2 Acquisition campaign

In the acquisition campaign, we tested our research questions using our platform with several participants in three different scenarios: a common manual drive, an autonomous drive, and a manual drive but in the passenger seat.

5.2.1 Participants and experimental setup

A total of 6 men aged between 24-32 years participated in our experiments. Only PhD students and professors of Politecnico di Milano took part in this study due to insurance issues. All participants held a valid driver's license, and four of them had a previous autonomous driving experience, while two had never experienced autonomous driving. Data were collected anonymously, and informed consent was obtained after explaining the task to the participants. No rewards were given for participation in the experiments. All experimental procedures were conducted in accordance with the permission of the Politecnico di Milano ethics commission granted on 2th December 2020.

The experiment took place on a 270mt route created in the Politecnico di Milano campus (Figure 5.17) using the I.DRIVE instrumented vehicle described in Chapter 3. This car was equipped with additional hardware and software that enable the car to drive autonomously in a known environment. Indeed, to drive the car autonomously, we created a test track's map driving the car on the predefined route and acquiring the lidar and GPS data.

5.2.2 Protocol

The driving study is divided into three different sessions: a common manual drive, an autonomous drive, and a manual drive but in the passenger seat.

Before starting with the experiments, the participant had to read an informative about the experiment's aims, how the experiment was performed, what the participant had to do, the possible risks connected to the experiments and how we had handled the data respecting European privacy law called GDPR. Afterward, the participant had to sign an informed consent and a consent to data processing. Furthermore, the participant had to fill a pre-drive questionnaire (Appendix A) with general questions about his driving experience and stress level before the experiment.

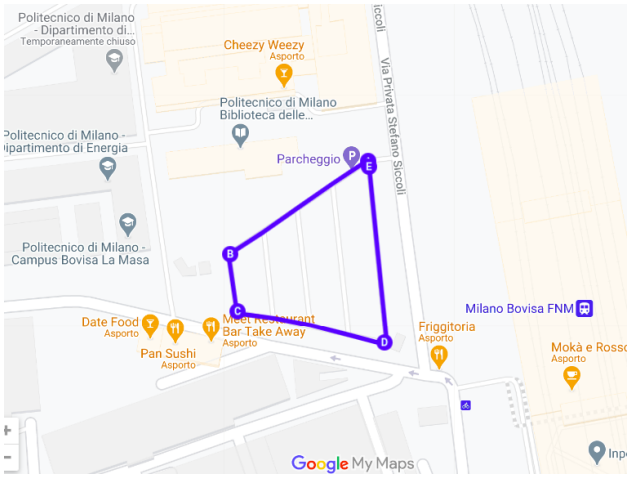


Figure 5.17: The track route

We had asked the participant to remove all the electronic devices he had (i.e. smartphone, watches) since they might interfere with the Procomp sensors. Afterward, we started to place on the participants the 3 electrodes to measure the electrocardiogram (EKG) sensor, the 2 electrodes on the fingers to measure the skin conductance (SC) and the chest band to measure the respiration.

The participant wore the eye-tracking and performed the calibration process. In the calibration procedure, the participant had to look at the center of the marker (Figure 5.18) and slowly move his head in a spiral motion while continuing to gaze at the center of the marker. The data collected during the calibration period is used afterward from the pupil software to correlate the world camera with the eye cameras.

Afterward, the participant entered the car sitting on the driver's seat. We connected all the physiological sensors to the Procomp. After a brief check that all the sensors are working correctly, we asked the participants to speak only if it is necessary. Indeed, speaking during an acquisition changes the respiration rate and, consequently, the heart rate. At this point, we started the acquisition with a 3 minutes baseline for the physiological sensors. During the baseline, the participant had to relax with the eyes closed to avoid distractions. When the 3-minute baseline has passed, the participant drives manually on the track at the speed he prefers for 5 laps. After driving, the participant had to fill a questionnaire (Appendix A) on the driving just

Figure 5.18: The marker used for the eye tracker calibration

experienced.

Subsequently, we started again acquiring the 3 minutes baseline, but, in this session, the participant experienced an autonomous drive for 5 laps. The autonomous drive was performed at about 10-11 km/h to ensure maximum safety. After this drive, we asked the participants to fill a questionnaire on the autonomous drive just experienced.

Finally, we asked the participant to move to the passenger seat and we acquired the participant baseline and the data on a manual drive but as a passenger. The participant had to fill the last questionnaire on the manual driving experienced as a passenger.

During the three baselines and the three driving sessions, we acquired the vehicle, the environmental, and the physiological sensors installed in the I.DRIVE car.

5.2.3 Results of the acquisition campaign

Time parameters analysis

In the first part of the analysis, we used each participant data acquired in the three scenarios (autonomous, manual and passenger) to extract the duration, latency and shift parameters for each stress index that we used (skin conductance's phasic component, heart rate, LF/HF and respiration rate). To select the best parameters to use in the data analysis, we searched the parameters that maximize the *stress_ratio* as we did in the preliminary test. All the following *stress_ratio* are relative to the event sequence extracted

from the jerk on x-axis. In this case, low percentiles of the jerk on x-axis indicate an abrupt acceleration, while high percentiles of the jerk on x-axis indicate an abrupt deceleration.

For the skin conductance's phasic component we extracted each participant's *stress_ratio* values for different durations in the range 1s - 10s. We computed the best duration values on both high percentiles and low percentiles. For high percentiles, we found that only participant 5 had a *stress_ratio* value above 2 for the duration of 5s (Figure 5.19a). We can observe a peak above 2 also at a duration value of 1s for participant 5, but the value is lower than the peak at 5s. For low percentiles, we discovered that participant 1,5,6 had a high correlation value with a duration value of 5s while participant 4 had a high *stress_ratio* values for a duration of 6s (Figure 5.19b). Participant 5 presented again 2 peaks above 2 and we selected the one with a higher value.

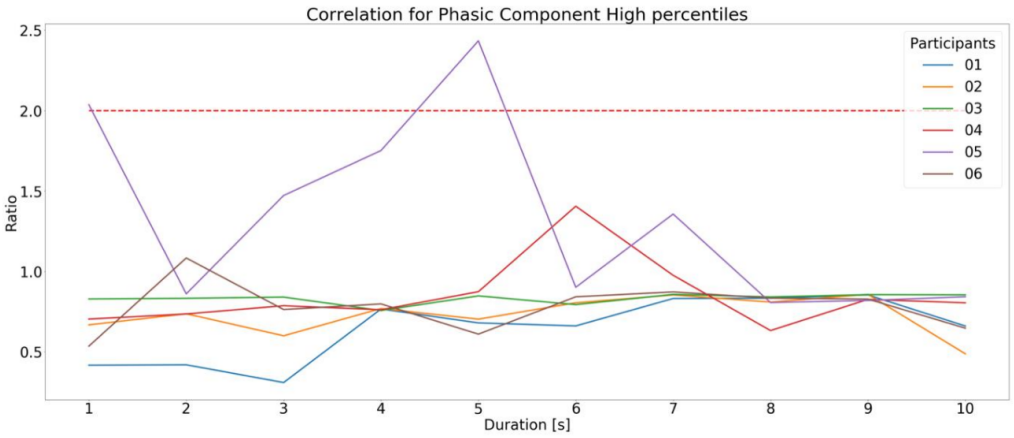
For the analysis of latencies and shift values of the skin conductance's phasic component we considered only low percentiles since they had shown more correlation than high percentile. Furthermore, in these analyses, we considered only the participants who showed a high correlation value in the duration parameter analysis.

Figure 5.20 shows, for each participants, the *stress_ratio* for different latency values (0s - 5s). We found that participant 1 and 4 had a high *stress_ratio* for a latency value of 0s, participant 5 had a high correlation for a value of 3s and participant 6 had a peak with a latency of 1s. Figure 5.21 showed, for each participants, the *stress_ratio* for different shifts values (-2s - 2s). We found that all the participants have a *stress_ratio* value above 2 for a shift value of 0s.

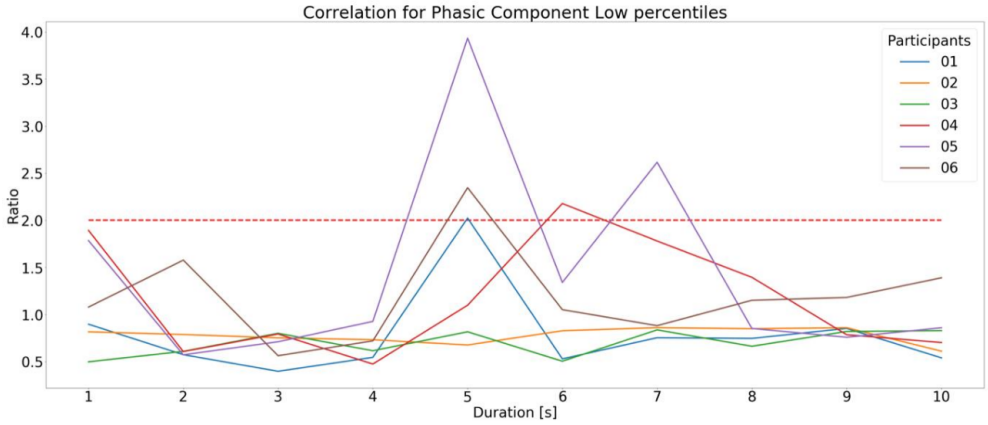
In Table 5.5 are reported the duration, latency and shift parameter of the skin conductance's phasic component that we used in the subsequent correlation analysis for each participant. For participants 2 and 3 we decided to use a duration value of 5s, a latency value of 1s and a shift value of 0s since they showed a low *Stress_ratio* and no valid time parameters were extracted for them.

For the heart rate we extracted each participant's *stress_ratio* values for different durations in the range 2s - 20s. For both low percentiles (Figure 5.22a) and high percentiles (Figure 5.22b), we did not find a *stress_ratio* values above 2. This means that no participant had a high correlation with the heart rate.

Since no participants showed a high correlation value, we decided to use



(a) High percentiles



(b) Low percentiles

Figure 5.19: *Stress_ratio* values of skin conductance's phasic component for different duration values. The red dotted horizontal line is the *Stress_ratio* threshold

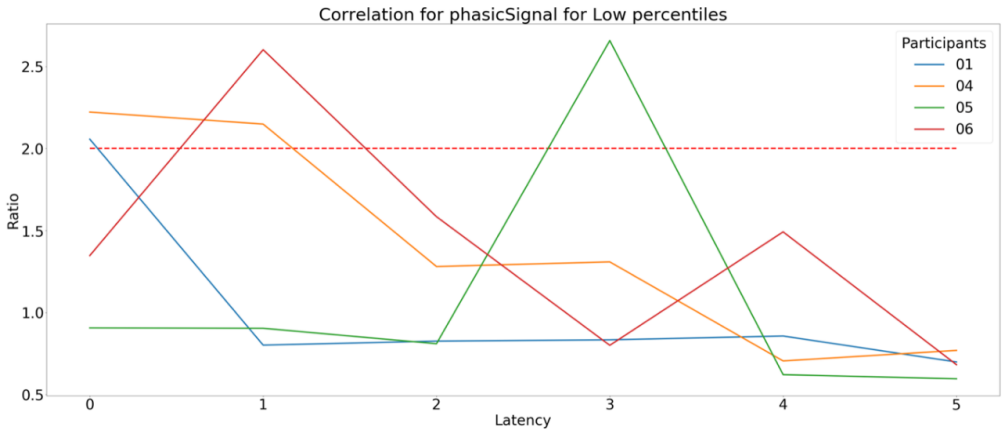


Figure 5.20: *Stress_ratio* values of skin conductance's phasic component at low percentiles for different latency values. The red dotted horizontal line is the *Stress_ratio* threshold

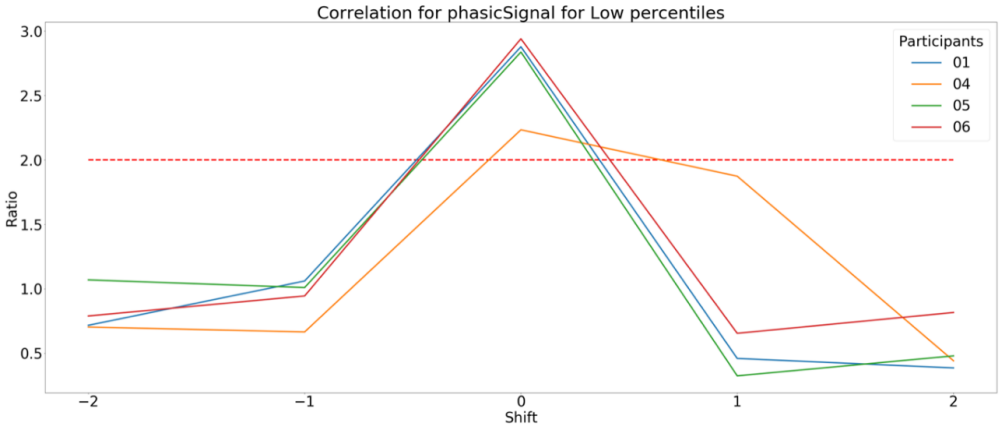
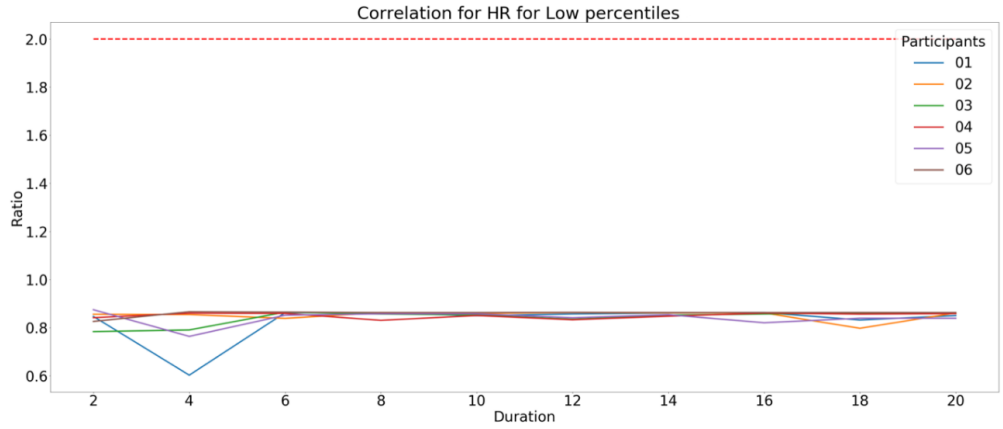
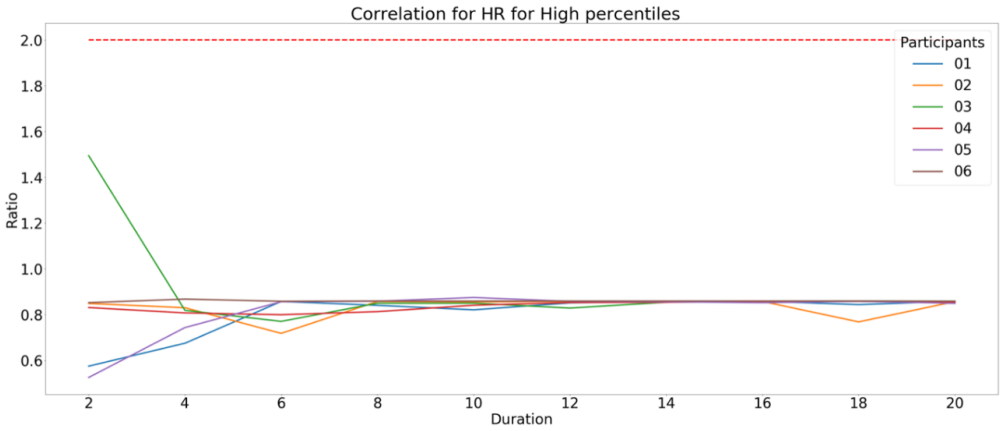


Figure 5.21: *Stress_ratio* values of skin conductance's phasic component at low percentiles for different shift values. The red dotted horizontal line is the *Stress_ratio* threshold



(a) Low percentiles



(b) High percentiles

Figure 5.22: *Stress_ratio* values of heart rate for different duration values. The red dotted horizontal line is the *Stress_ratio* threshold

Participant	Duration [s]	Latency [s]	shift [s]
1	5	0	0
2	-	-	-
3	-	-	-
4	6	0	0
5	5	3	0
6	5	1	0

Table 5.5: Values of duration, latency and shift used in the skin conductance's phasic component correlation analysis for each participant

for the subsequent correlation analysis of the heart rate, the time parameters founded in the preliminary test: a duration of 10s, a latency of 2s and a shift of 1s.

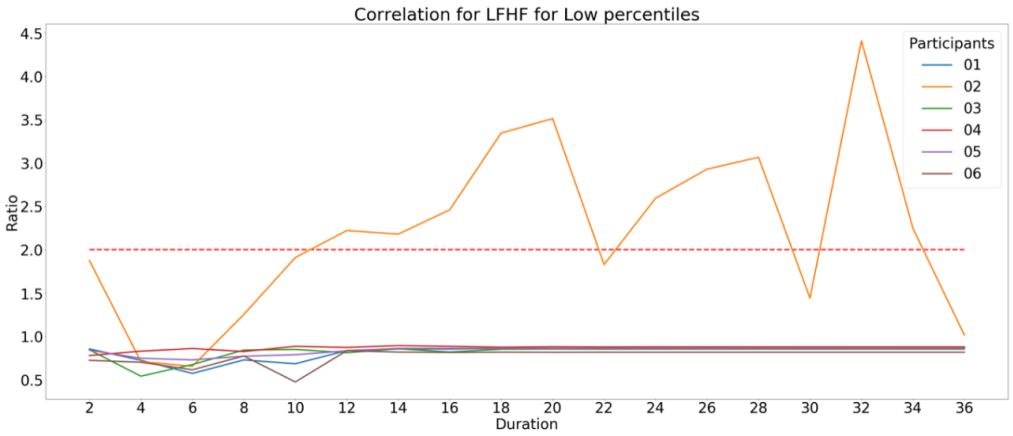
For the ratio between low frequency (LF) and high frequency (HF) (LF/HF) we extracted each participant's *stress_ratio* values for different durations in the interval 2s - 36s. For both low and high percentiles, we found that only participant 2 had different *stress_ratio* peaks above 2 (Figure 5.23a)(Figure 5.23b). The highest peak for both low and high percentiles have a duration value of 32s.

We analyzed the latency and shift values only for participants 2 since he was the only one that showed a high correlation in the duration study. Figure 5.24a and Figure 5.24b show the *stress_ratio* trend for different latency values in the interval 0s - 5s. We discovered that for both low and high percentiles, the best participant 2 latency value was 2s.

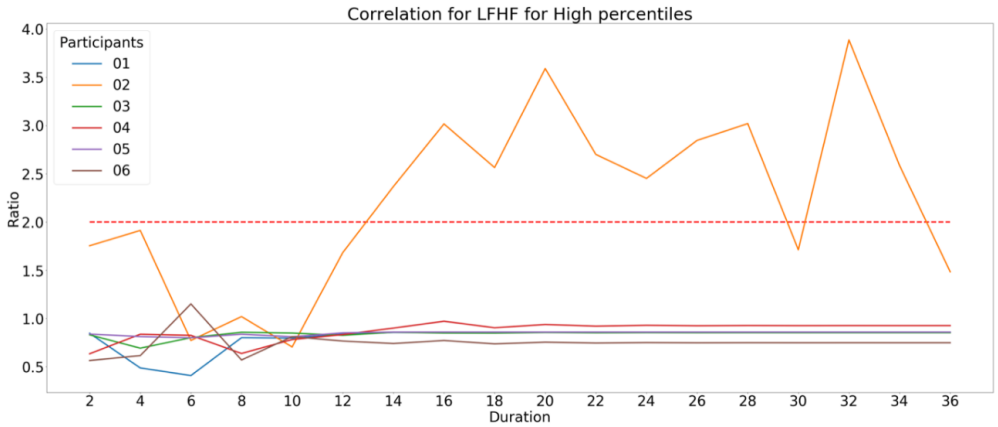
Finally, we computed the *stress_ratio* for different shift values. Figure 5.25a shows that the best shift value was 1s for low percentiles while Figure 5.25b shows that the best shift value for high percentiles was 2s.

For all participants that did not show a high correlation value, we used the same parameters discovered for participant 2, and so, we used a duration value of 10s, a latency value of 2s and a shift value of 1s.

For the respiration rate, we extracted each participant's *Stress_ratio* values for different duration in the interval 1s - 10s. For high percentiles (Figure 5.26a), we discovered that only participant 5 have a high *Stress_ratio* with a duration value of 2s while, for low percentiles (Figure 5.26b), participant 1 and 3 show a high correlation ratio at 2s while participant 5 showed a high *Stress_ratio* at 4s.

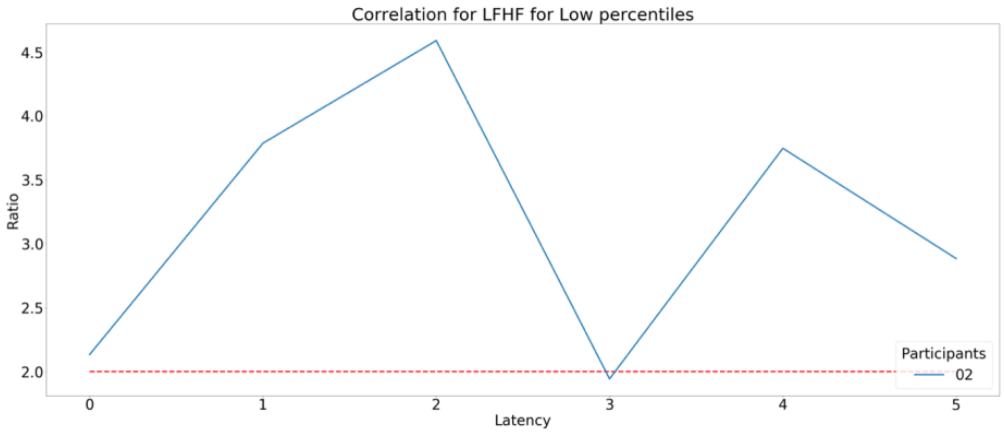


(a) Low percentiles

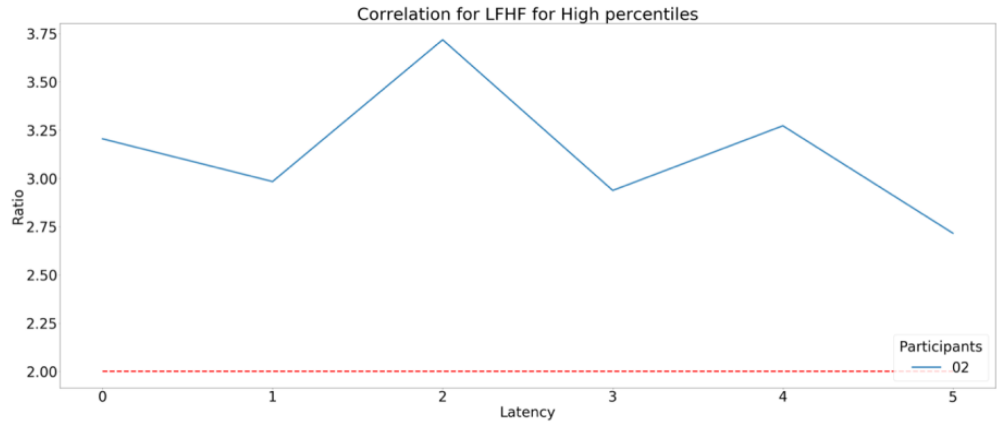


(b) High percentiles

Figure 5.23: *Stress_ratio* values of LF/HF for different duration values. The red dotted horizontal line is the *Stress_ratio* threshold

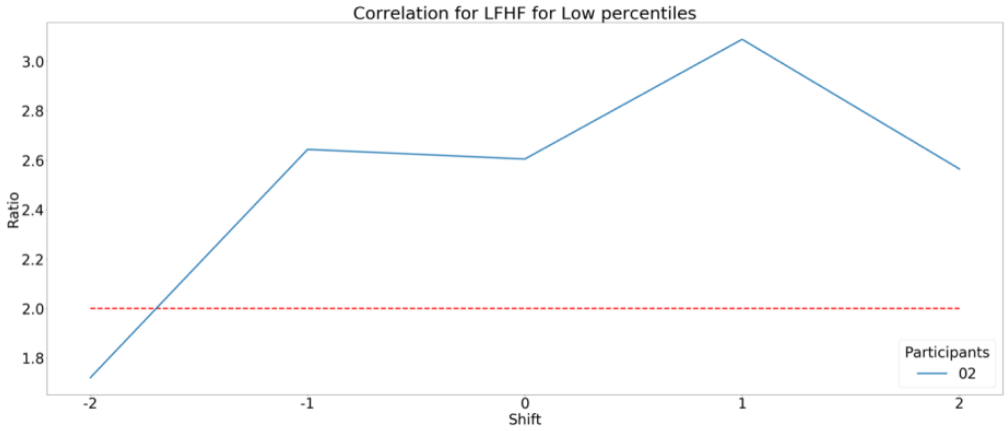


(a) Low percentiles

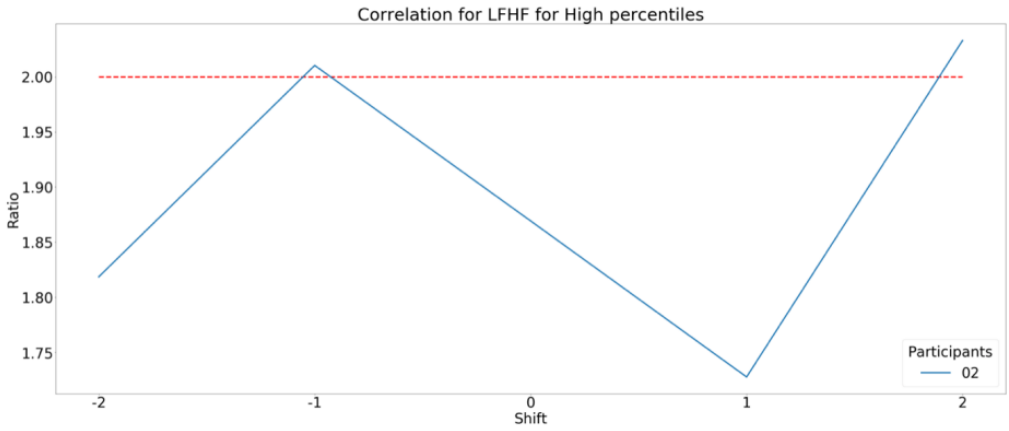


(b) High percentiles

Figure 5.24: *Stress_ratio* values of LF/HF for different latency values. The red dotted horizontal line is the *Stress_ratio* threshold

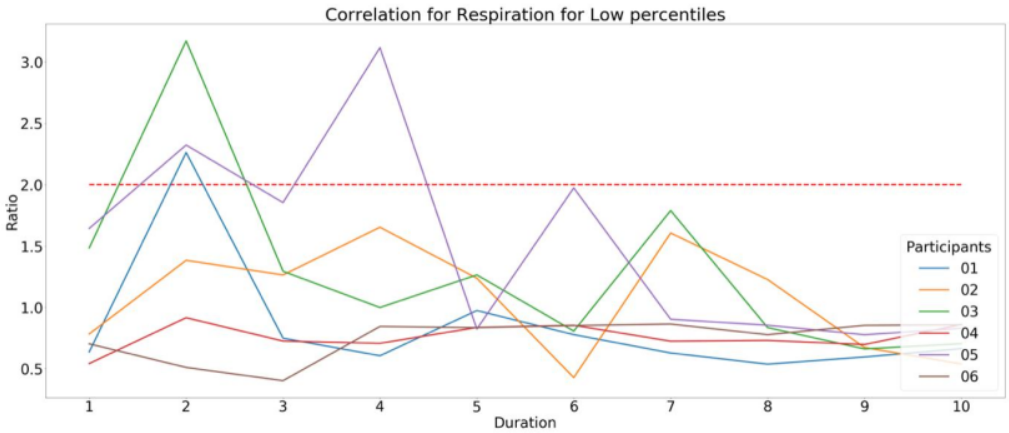


(a) Low percentiles

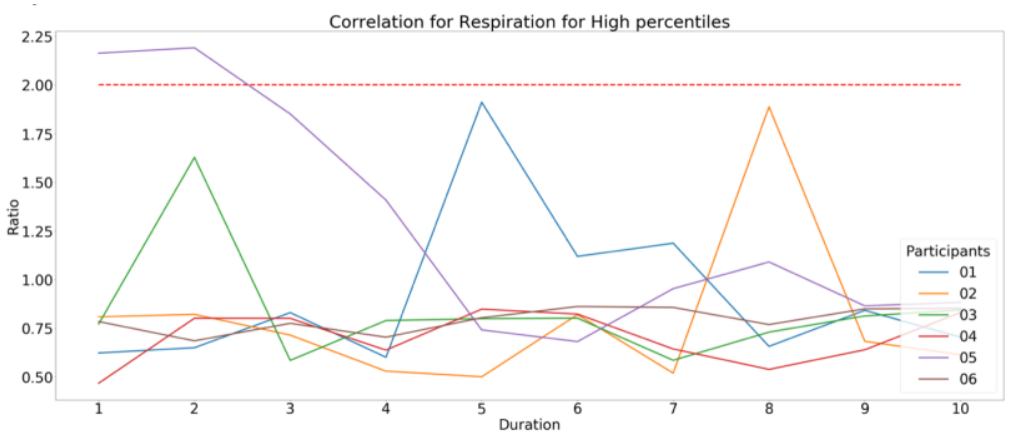


(b) High percentiles

Figure 5.25: *Stress_ratio* values of LF/HF for different shift values. The red dotted horizontal line is the *Stress_ratio* threshold



(a) Low percentiles



(b) High percentiles

Figure 5.26: *Stress_ratio* values of respiration rate for different duration values. The red dotted horizontal line is the *Stress_ratio* threshold

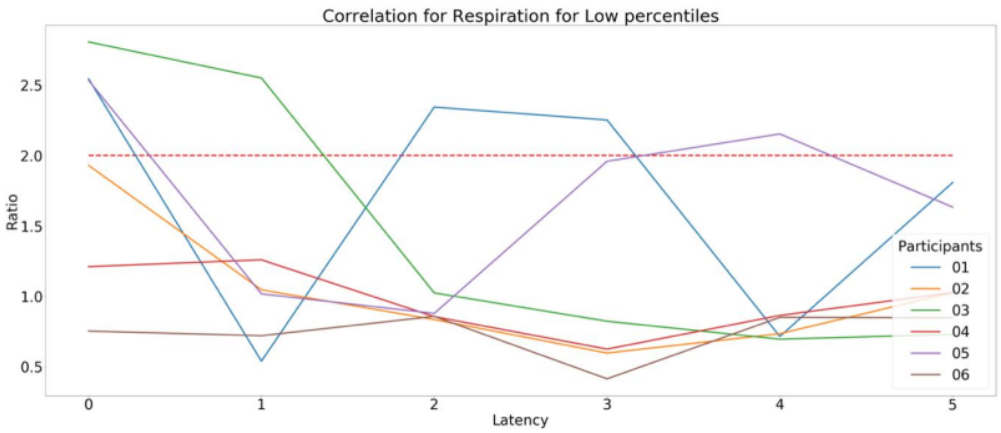


Figure 5.27: *Stress_ratio* values of respiration rate at low percentiles for different latency values. The red dotted horizontal line is the *Stress_ratio* threshold

We analyzed the latency and shift values only for low percentiles since, in the duration parameter analysis, we discovered that only one participant showed a significant correlation at high percentiles. In the latency analysis (Figure 5.27) we discovered that participant 1, 3 and 5 have a significant *Stress_ratio* value for a latency of 0s. For the shift parameter (Figure 5.28), we found a *Stress_ratio* value above 2 for participant 1 and 3 with a shift value of 0s and, for participant 5 with a shift value of -1s.

In Table 5.6 are reported the duration, latency and shift parameter of the resiration rate that we used in the subsequent correlation analysis for each participant. For participants 2, 4 and 5 we decided to use a duration value of 2s, a latency value of 0s and a shift value of 0s.

Correlation analysis

After the analysis of physiological data time parameters, we computed, for each participant and each driving feature, the correlation ratio according to different percentiles using the algorithm described in Section 4.4.

From Figure 5.29 to Figure 5.34 is possible to see the *stress_ratio* related to the skin conductance's phasic component of each participant for every driving features that we had considered. In Table 5.7 is possible to see the frequency of the first 5 driving features that showed highest stress correlation

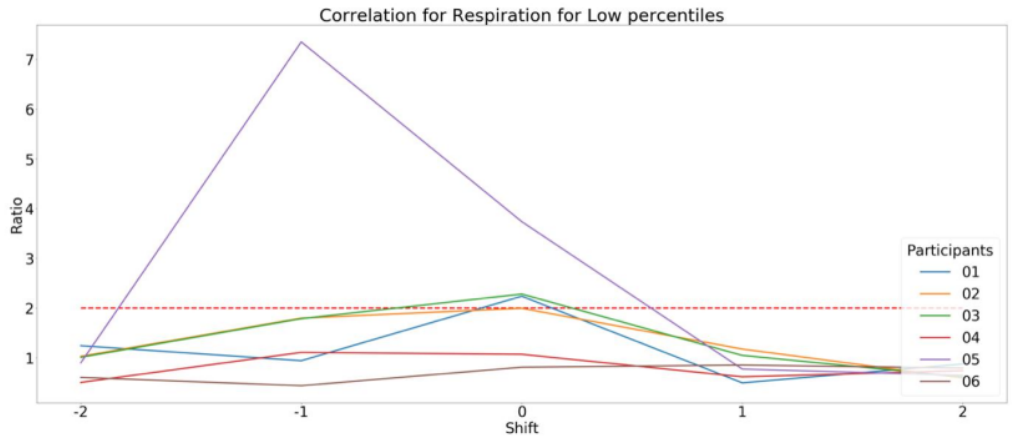


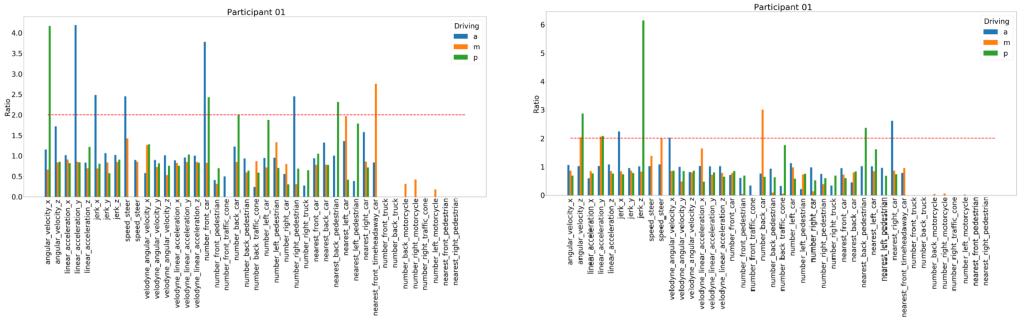
Figure 5.28: *Stress_ratio* values of respiration rate at low percentiles for different shift values. The red dotted horizontal line is the *Stress_ratio* threshold

with the skin conductance's phasic component.

The jerk on x-axis demonstrated that it is highly correlated with the stress in both high and low percentiles. This means that the abrupt changes in the acceleration in both acceleration and braking phase as an impact on the stress observed with the skin conductance's phasic component. The angular velocity on x-axis, the linear acceleration on y-axis, and the angular velocity on z-axis have a higher stress correlation. All these driving features identify, in different ways, that the car is turning. The angular velocity on x identifies a change in the roll angle due to the car suspension's lateral movement. The linear acceleration on y measures the lateral acceleration applied to the car. The angular velocity on z measures a change in the yaw angle of the car.

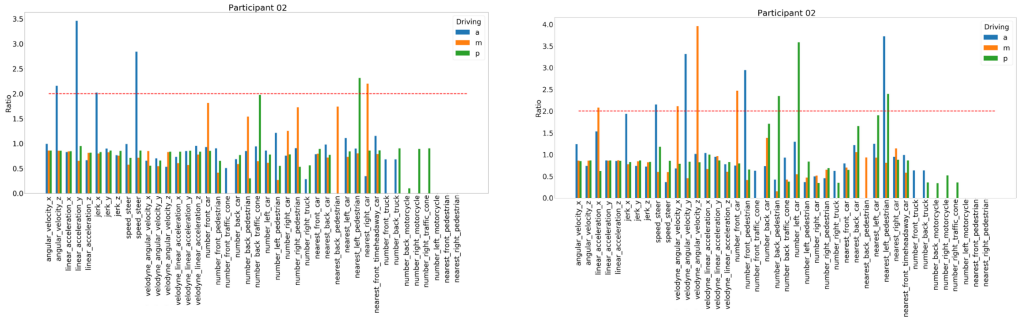
In Table 5.7 is possible to notice that almost all the stressful driving features found have a higher frequency in the autonomous drive scenario. This is confirmed from Figure 5.35 that shows the *stress_ratio* mean over all participants in the autonomous, manual and passenger scenario. Indeed, in Figure 5.35 is possible to see that the *stress_ratio* using the skin conductance's phasic component is higher in autonomous condition with respect to the other two scenarios. The mean of the *stress_ratio* in the manual scenario is lower than the passenger scenario but not significantly, especially for low percentiles.

We computed the *stress_ratio* related to the respiration rate of each par-



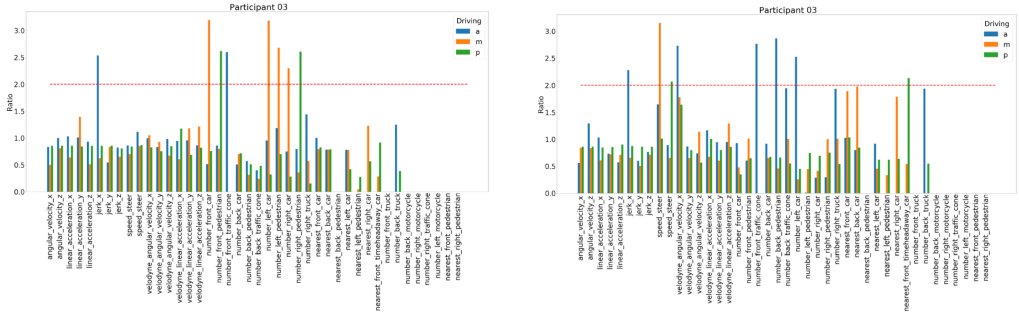
(a) High percentiles (b) Low percentiles

Figure 5.29: Participant 1 stress ratio of the skin conductance's phasic component for each driving features in the three driving scenarios (autonomous, manual, passenger)



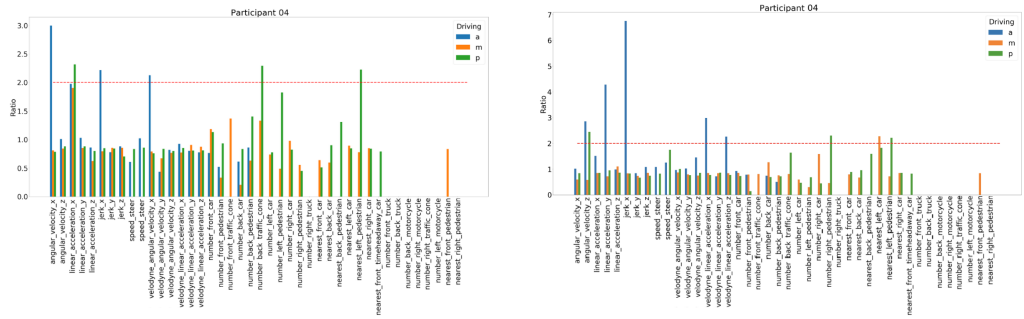
(a) High percentiles (b) Low percentiles

Figure 5.30: Participant 2 stress ratio of the skin conductance's phasic component for each features in the three driving scenarios (autonomous, manual, passenger)



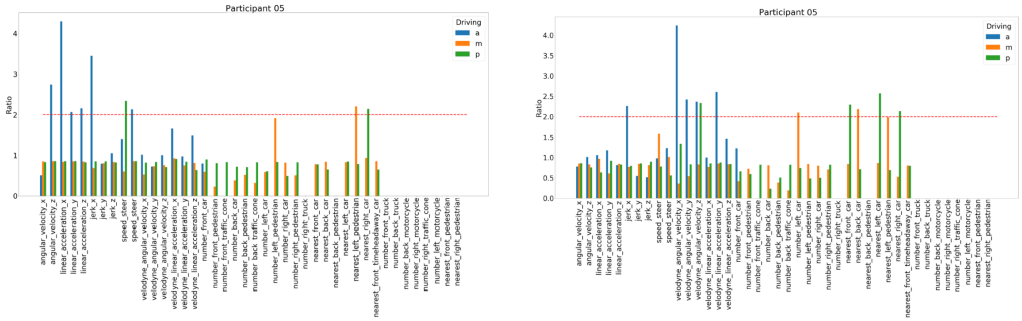
(a) High percentiles (b) Low percentiles

Figure 5.31: Participant 3 stress ratio of the skin conductance's phasic component for each features in the three driving scenarios (autonomous, manual, passenger)



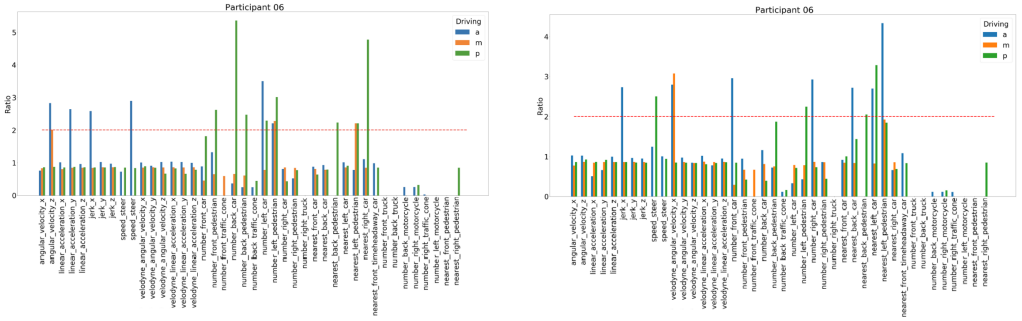
(a) High percentiles (b) Low percentiles

Figure 5.32: Participant 4 stress ratio of the skin conductance's phasic component for each features in the three driving scenarios (autonomous, manual, passenger)



(a) High percentiles (b) Low percentiles

Figure 5.33: Participant 5 stress ratio of the skin conductance's phasic component for each features in the three driving scenarios (autonomous, manual, passenger)



(a) High percentiles (b) Low percentiles

Figure 5.34: Participant 6 stress ratio of the skin conductance's phasic component for each features in the three driving scenarios (autonomous, manual, passenger)

Participant	Duration [s]	Latency [s]	shift [s]
1	2	0	0
2	-	-	-
3	2	0	0
4	-	-	-
5	4	0	-1
6	-	-	-

Table 5.6: Values of duration, latency and shift used in the respiration rate correlation analysis for each participant

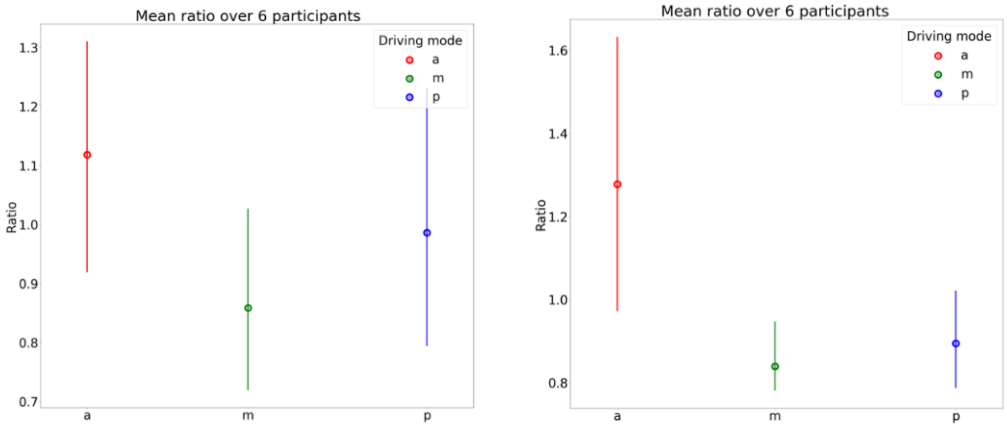
Driving features	A	M	P
High jerk on x	6	0	0
Low jerk on x	5	0	0
Low angular velocity on x	4	2	0
High linear acceleration on y	4	0	0
High angular velocity on z	3	1	0

Table 5.7: Frequency distribution of the most stressful driving features observed between all participants in the three driving scenario (Autonomous, Manual, Passenger) for the skin conductance's phasic component

ticipant for every driving features. From Figure 5.36 to Figure 5.41 is possible to see the *stress_ratio* related to the respiration rate of each participant for every driving features.

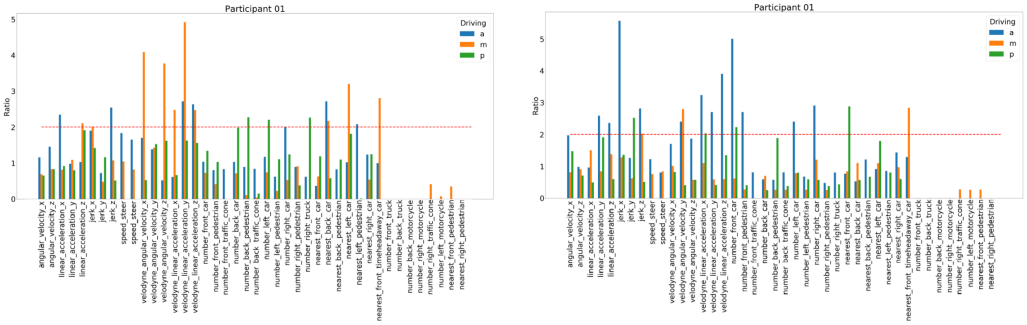
In Table 5.8 is possible to see the frequency of the first 5 driving features that showed the highest stress correlation with the respiration rate. The result found with the respiration rate confirmed what we discovered with the skin conductance's phasic component. Indeed, for the respiration rate, the jerk on x was still one of the most stressful driving features, as also the linear acceleration on y and the angular velocity on x. The only difference between the two analysis was the angular velocity on z that, for the respiration rate, did not show a high-stress correlation as for the skin conductance's phasic component.

Figure 5.42 confirmed, for the respiration rate, that the autonomous scenario was more stressful than the manual scenario. For high percentiles (Fig-



(a) High percentiles (b) Low percentiles

Figure 5.35: Mean *Stress_ratio* values of the skin conductance's phasic component over all participants in the three driving scenario (Autonomous, Manual, Passenger)

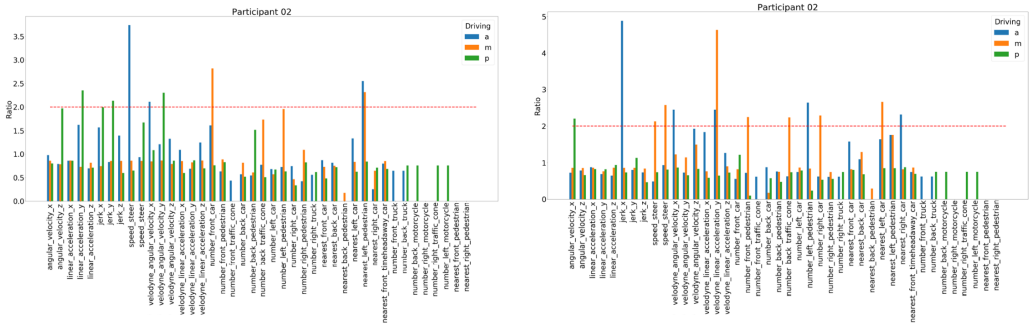


(a) High percentiles (b) Low percentiles

Figure 5.36: Participant 1 stress ratio of the respiration rate for each driving features in the three driving scenarios (autonomous, manual, passenger)

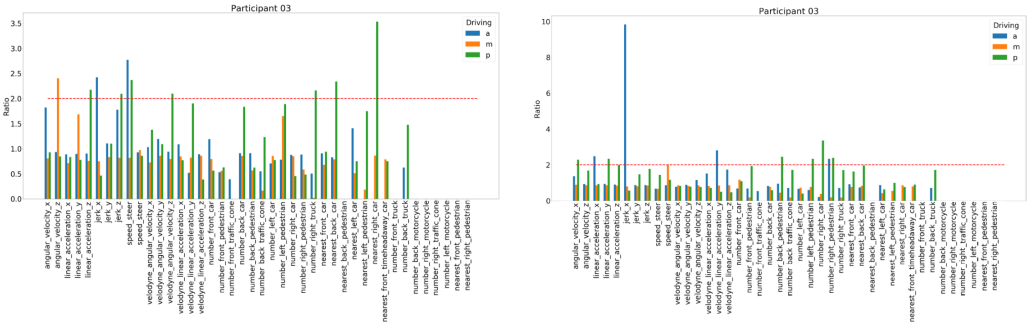
ure 5.42a), the passenger scenario was the most stressful, and the difference between the autonomous and the manual scenarios are not high. Instead, for low percentiles, the passenger scenario had a similar *stress_ratio* with respect to the manual scenario.

We computed the *stress_ratio* related to the heart rate of each participant for every driving features. From Figure 5.43 to Figure 5.48 is possible to see



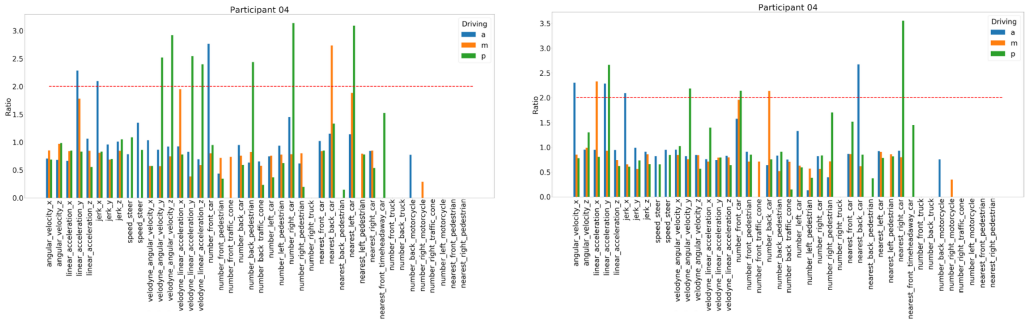
(a) High percentiles (b) Low percentiles

Figure 5.37: Participant 2 stress ratio of the respiration rate for each features in the three driving scenarios (autonomous, manual, passenger)



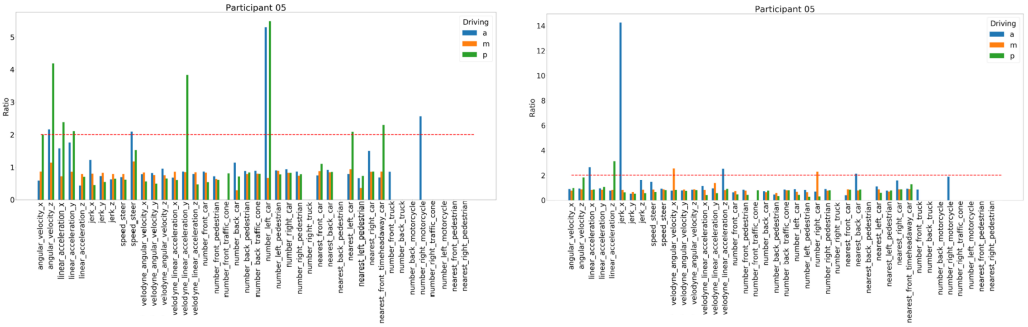
(a) High percentiles (b) Low percentiles

Figure 5.38: Participant 3 stress ratio of the respiration rate for each features in the three driving scenarios (autonomous, manual, passenger)



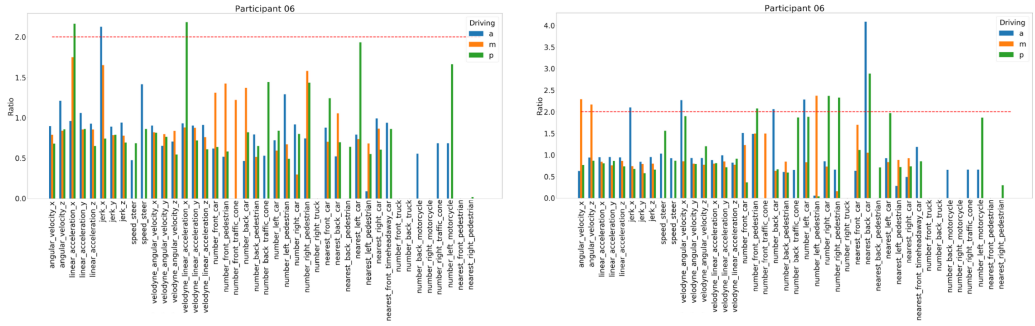
(a) High percentiles (b) Low percentiles

Figure 5.39: Participant 4 stress ratio of the respiration rate for each features in the three driving scenarios (autonomous, manual, passenger)



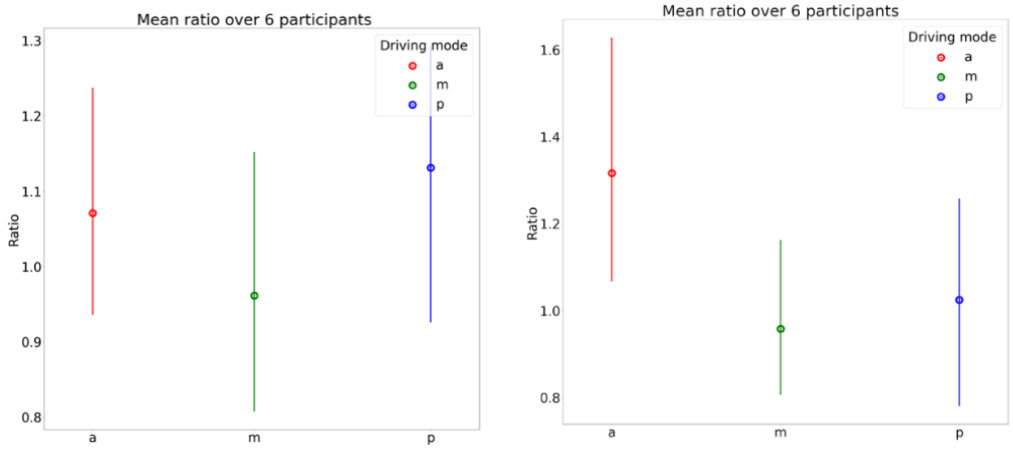
(a) High percentiles (b) Low percentiles

Figure 5.40: Participant 5 stress ratio of the respiration rate for each features in the three driving scenarios (autonomous, manual, passenger)



(a) High percentiles (b) Low percentiles

Figure 5.41: Participant 6 stress ratio of the respiration rate for each features in the three driving scenarios (autonomous, manual, passenger)

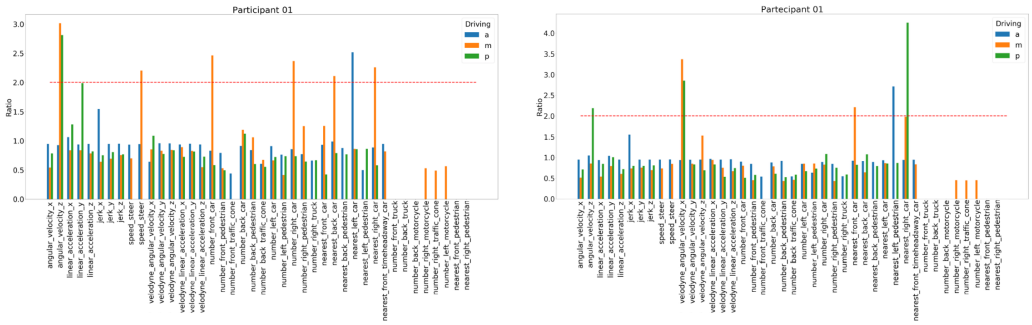


(a) High percentiles (b) Low percentiles

Figure 5.42: Mean *Stress_ratio* values of the respiration rate over all participants in the three driving scenario (Autonomous, Manual, Passenger)

Driving features	A	M	P
Low jerk on x	6	0	0
Low linear acceleration on y	3	1	0
High jerk on x	3	1	0
Low nearest car on the back	3	0	1
Low angular velocity on x	2	1	0

Table 5.8: Frequency distribution of the most stressful driving features observed between all participants in the three driving scenario (Autonomous, Manual, Passenger) for the respiration rate



(a) High percentiles

(b) Low percentiles

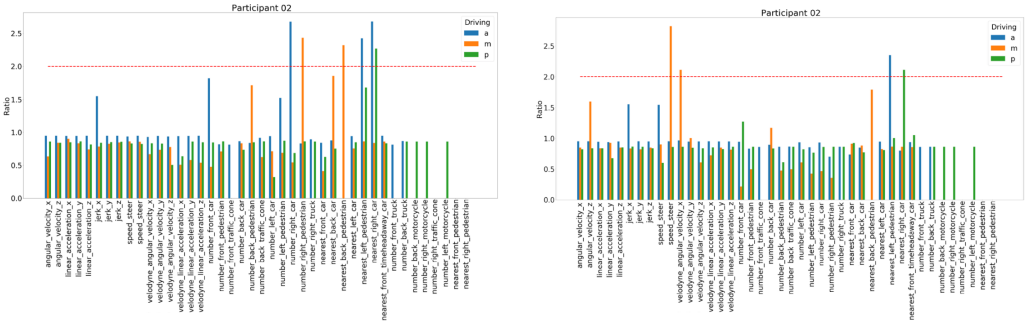
Figure 5.43: Participant 1 stress ratio of the heart rate for each features in the three driving scenarios (autonomous, manual, passenger)

the *stress_ratio* for each participant for low percentiles and high percentiles.

We decided to analyze the correlation of the heart rate despite, in the time parameters analysis, we discovered that the stress correlation of the heart rate was very low. The correlation analysis confirmed the low correlation value that we found in the time parameters analysis, and no driving features showed a significant frequency over all the participants.

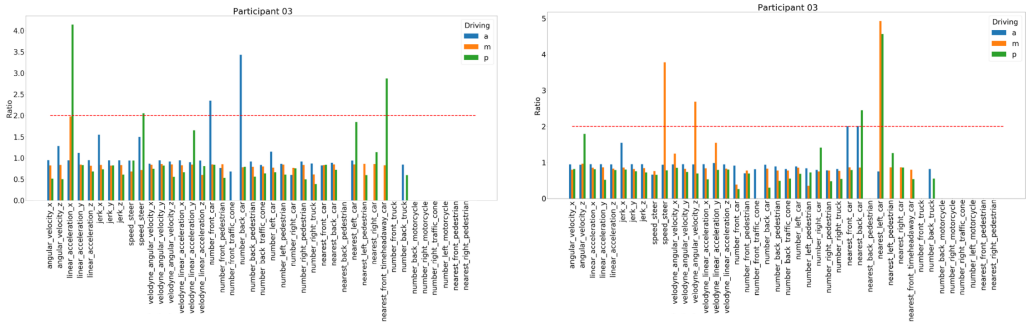
Despite the low *stress_ratio* showed from the heart rate, in Figure 5.49 is possible to see that also for the heart rate the *stress_ratio* mean in the autonomous scenario is higher than the manual and passenger scenario. However, in this case, the passenger scenario showed a slightly lower stress correlation with respect to the manual drive.

Finally, we computed the *stress_ratio* related to the heart rate variability



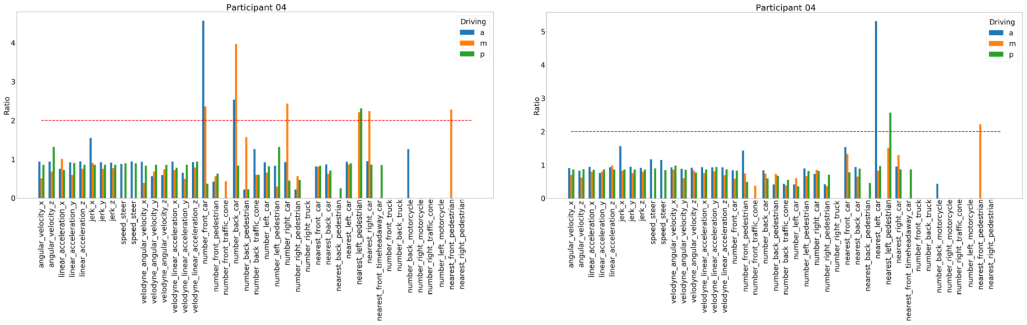
(a) High percentiles (b) Low percentiles

Figure 5.44: Participant 2 stress ratio of the heart rate for each features in the three driving scenarios (autonomous, manual, passenger)



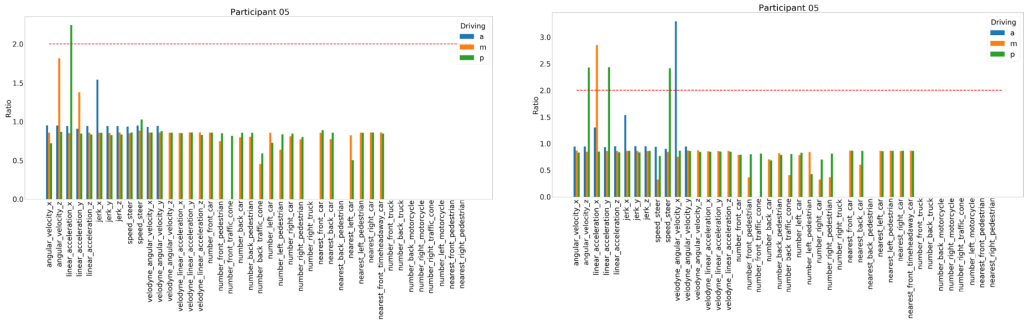
(a) High percentiles (b) Low percentiles

Figure 5.45: Participant 3 stress ratio of the heart rate for each features in the three driving scenarios (autonomous, manual, passenger)



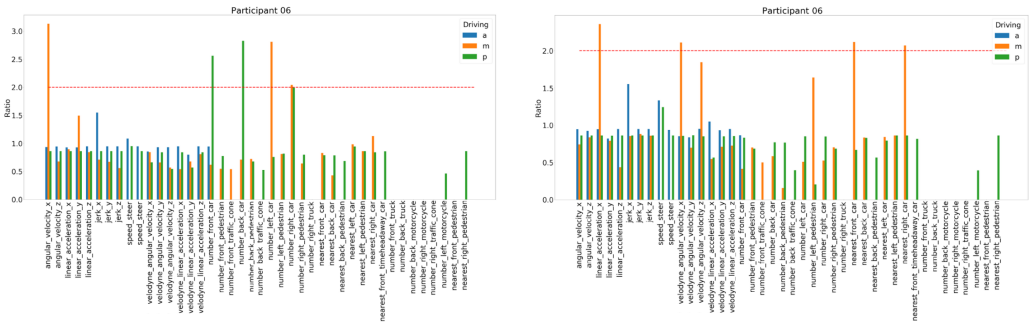
(a) High percentiles (b) Low percentiles

Figure 5.46: Participant 4 stress ratio of the heart rate for each features in the three driving scenarios (autonomous, manual, passenger)



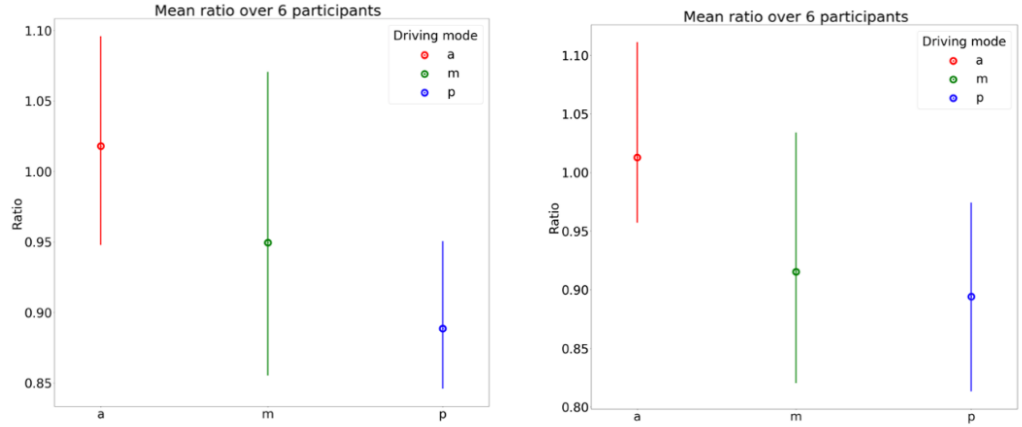
(a) High percentiles (b) Low percentiles

Figure 5.47: Participant 5 stress ratio of the heart rate for each features in the three driving scenarios (autonomous, manual, passenger)



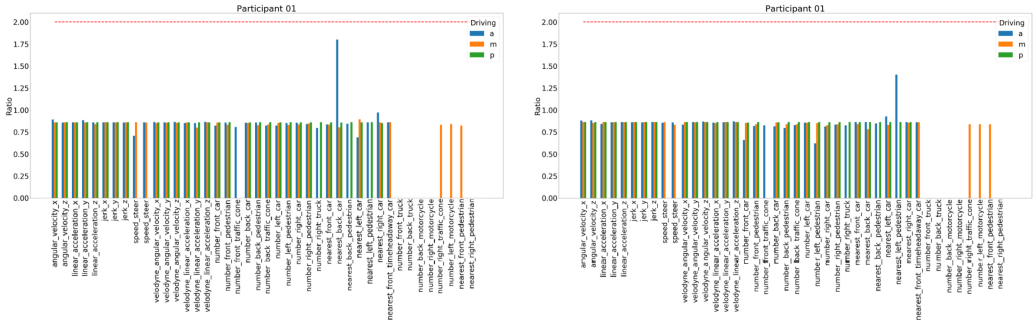
(a) High percentiles (b) Low percentiles

Figure 5.48: Participant 6 stress ratio of the heart rate for each features in the three driving scenarios (autonomous, manual, passenger)



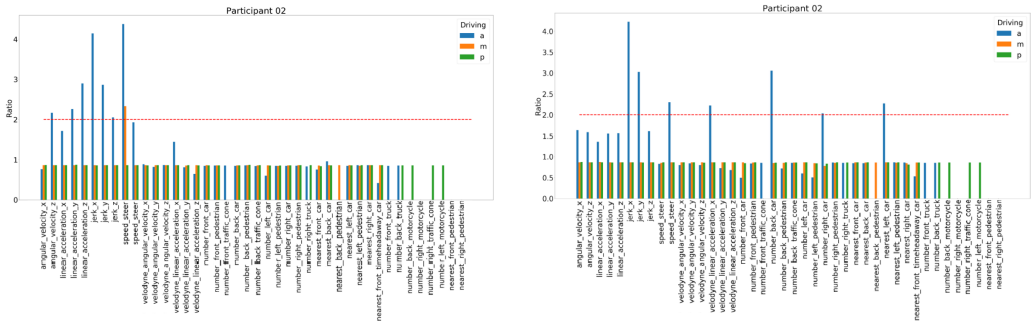
(a) High percentiles (b) Low percentiles

Figure 5.49: Mean *Stress_ratio* values of the heart rate over all participants in the three driving scenario (Autonomous, Manual, Passenger) for the heart rate



(a) High percentiles (b) Low percentiles

Figure 5.50: Participant 1 stress ratio of the heart rate variability for each features in the three driving scenarios (autonomous, manual, passenger)

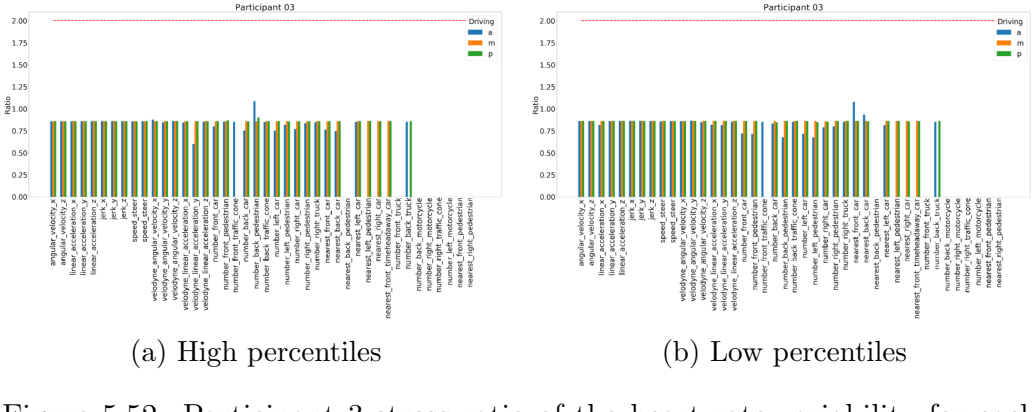


(a) High percentiles (b) Low percentiles

Figure 5.51: Participant 2 stress ratio of the heart rate variability for each features in the three driving scenarios (autonomous, manual, passenger)

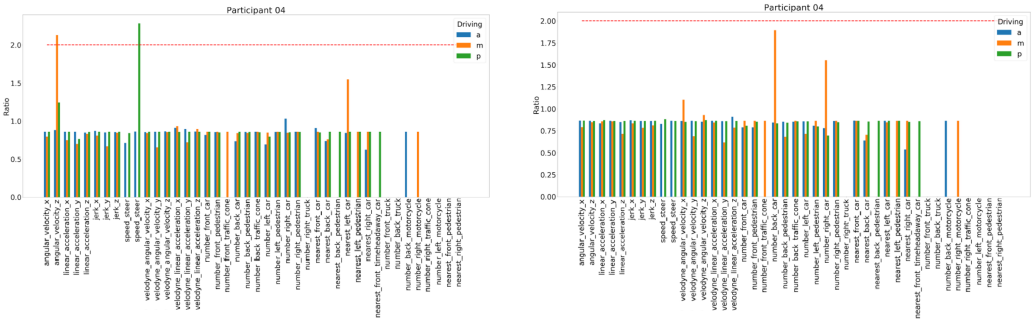
of each participant for every driving features. From Figure 5.50 to Figure 5.55 is possible to see the *stress_ratio* for each participant for low percentiles and high percentiles.

The correlation analysis confirmed the low correlation value that we found in the time parameters analysis, and no driving features showed a significant frequency over all the participants. However, it is possible to see in Figure 5.56 that also for the heart rate variability, the *stress_ratio* mean in the autonomous scenario is higher than the manual and passenger scenario. Furthermore, the passenger scenario and the manual scenario show a similar stress correlation.



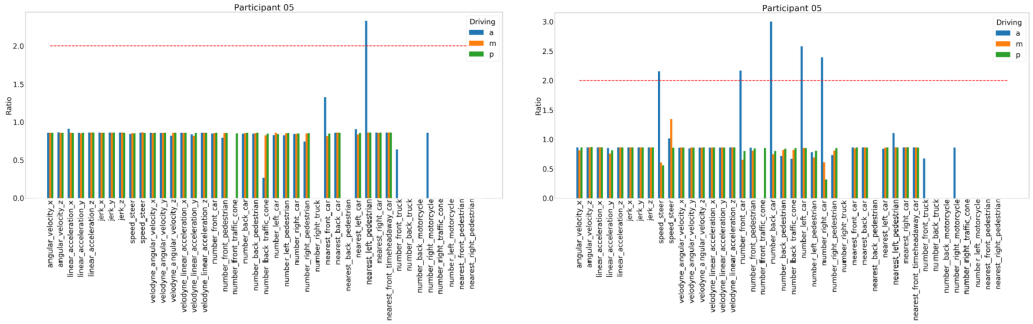
(a) High percentiles (b) Low percentiles

Figure 5.52: Participant 3 stress ratio of the heart rate variability for each features in the three driving scenarios (autonomous, manual, passenger)



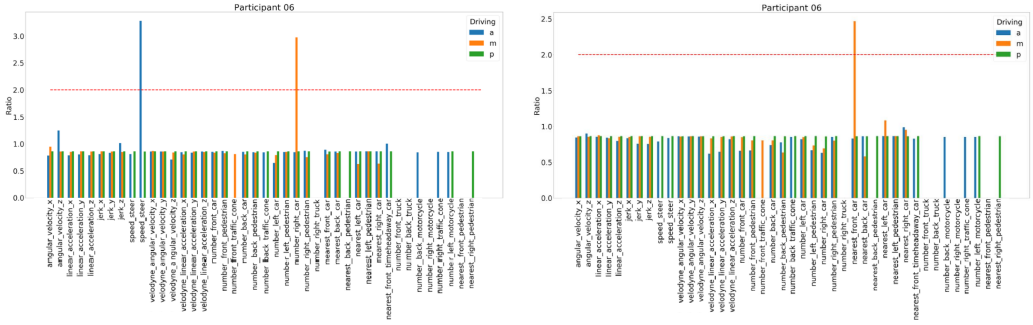
(a) High percentiles (b) Low percentiles

Figure 5.53: Participant 4 stress ratio of the heart rate variability for each features in the three driving scenarios (autonomous, manual, passenger)



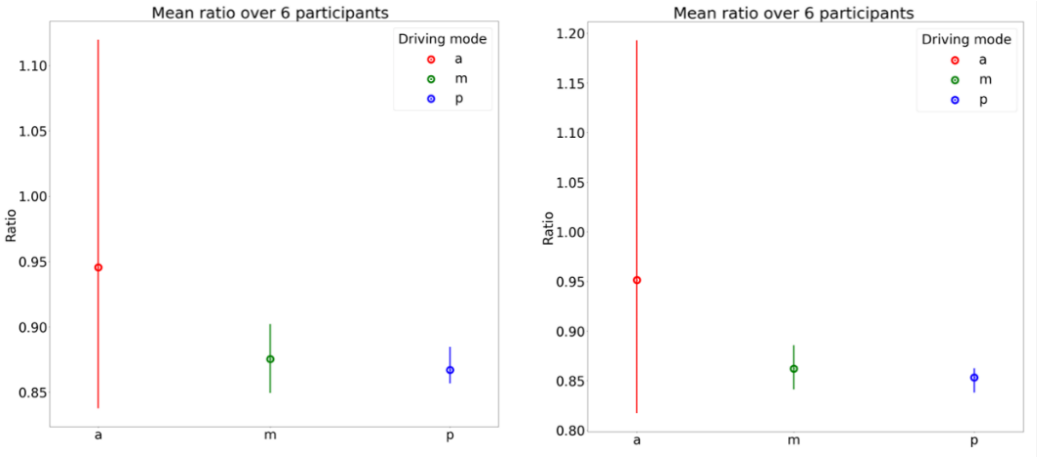
(a) High percentiles (b) Low percentiles

Figure 5.54: Participant 5 stress ratio of the heart rate variability for each features in the three driving scenarios (autonomous, manual, passenger)



(a) High percentiles (b) Low percentiles

Figure 5.55: Participant 6 stress ratio of the heart rate variability for each features in the three driving scenarios (autonomous, manual, passenger)



(a) High percentiles (b) Low percentiles

Figure 5.56: Mean *Stress_ratio* values of the heart rate variability over all participants in the three driving scenario (Autonomous, Manual, Passenger)

Questionnaire analysis

We analyzed the post-driving questionnaires (Appendix A) that participants filled after each scenario to compare the questionnaire results with the correlation analysis results.

In the questionnaire, we asked the participants to evaluate their agitation during the driving scenario. In Figure 5.57, it is possible to see that the participants evaluated the autonomous scenario as the scenario where they felt most agitated. This confirmed what we discovered in the correlation analysis. Contrary to what we discovered in the correlation analysis, the participants evaluated the passenger scenario as the calmest scenario.

In our questionnaire, we also asked the participants to rate the stress they felt using multiple statements. We calculate the average over all the statements for each participant. After, we compute the mean of all the participants. Figure 5.58 shows again that the autonomous scenario was the most stressful, while the other two scenarios had a similar degree of stress. If we compared Figure 5.58 with Figure 5.57 it is possible to observe that the results were similar but, all the values in Figure 5.58 were smaller than the values in the Figure 5.57.

We asked to the participants to evaluate the degree of control they felt during the driving scenario. Figure 5.59 demonstrates that, in the au-

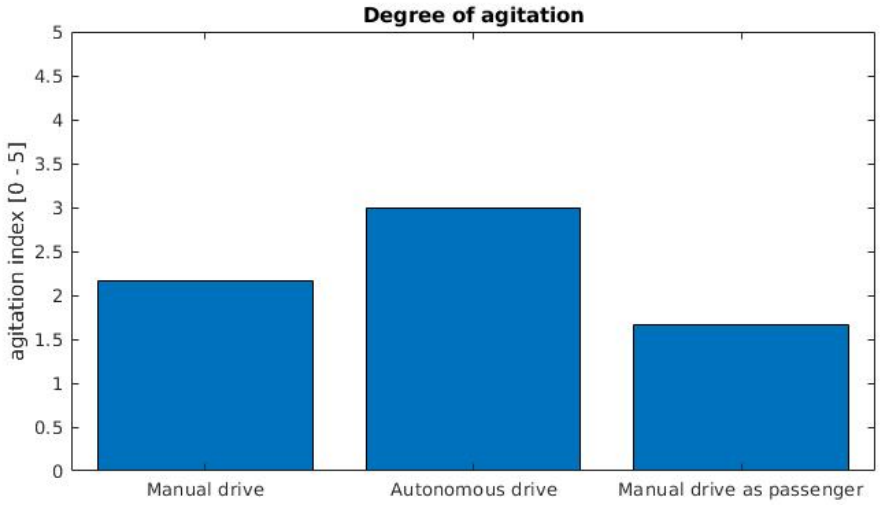


Figure 5.57: Mean of the degree of agitation over all participant in the three scenarios on a scale from 0 (very calm) to 5 (very agitated)

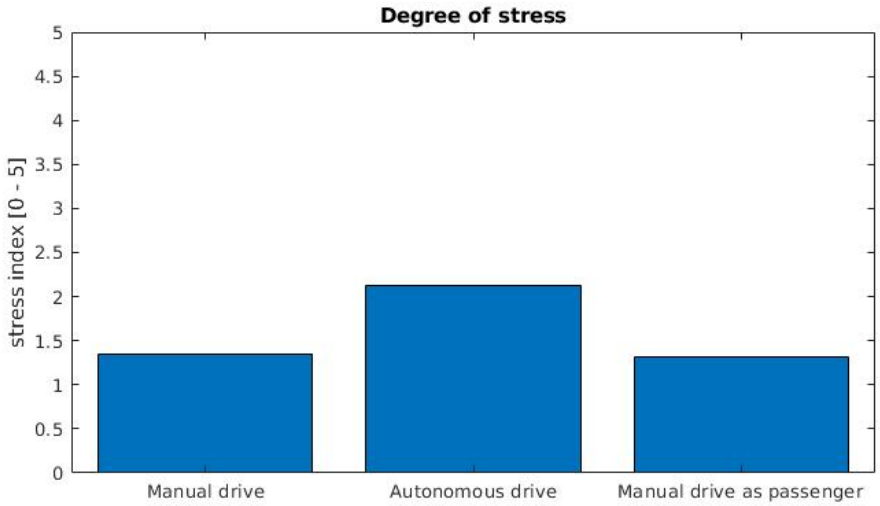


Figure 5.58: Mean of the degree of stress over all participant in the three scenarios on a scale from 0 (no stress) to 5 (stressed)

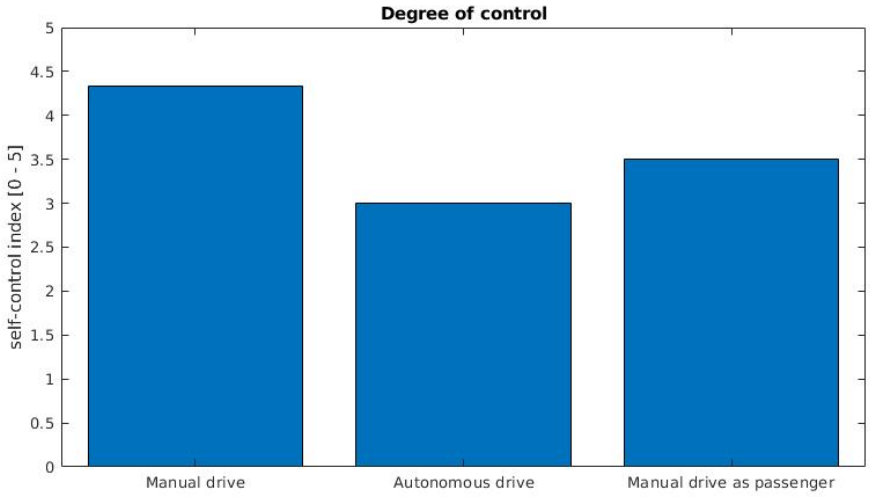


Figure 5.59: Mean of the degree of control over all participant in the three scenarios on a scale from 0 (no control) to 5 (full control)

tonomous scenario, the participants perceived a lack of of control. The lack of control was perceived from the participants also in the passengers scenario.

5.2.4 Discussion

The time parameters results demonstrated that not all the stress indexes computed in our study were reliable for detecting the driver's stress in the wild. We showed that only the phasic component extracted from skin conductance and the respiration rate were correlated with driving features. Instead, the LF/HF and the heart rate measures returned a low correlation with driving events. We expected a low correlation for the LF/HF since the results we obtained in the preliminary test, but we did not expect a low correlation for the heart rate.

The skin conductance's phasic component demonstrated to be a reliable stress index in our experiments since, in every acquisition, it showed good correlation values with the driving features. Furthermore, the results obtained with this stress index confirmed the questionnaire results.

The HR need further investigation since, in the preliminary test, it showed an excellent correlation, and it confirmed the skin conductance's phasic com-

ponent results. In contrast, in the acquisition campaign, it showed to have a very low correlation. A cause of the low correlation should be due to the noise in the EKG signal. Indeed, in the acquisition campaign, we faced with different body types and, when the participant sat in the car, sometimes happened that the EKG became really noisy due to the electrode on the iliac crest that, sometimes, in the sitting position, did not provide the best contact with the skin.

We demonstrated that the physiological indexes' time parameters changed from index to index and were not the same for all participants. We found that some time parameters had the same value over all the participants, like the shift for the skin conductance's phasic component or the latency for the respiration rate. However, we need more data to demonstrate that these fixed parameters are valid for every possible participant. Furthermore, the time parameters could also change in time when getting more comfortable with autonomous driving. Constant monitoring in time for each participant could lead to the detection of changes in the time parameters, but we not investigated the time parameters evolution.

The correlation analysis showed that the jerk on x-axis, the linear acceleration on y-axis and the angular velocity on x-axis are the driving features that mostly affect the participants' stress. The jerk on x showed a high correlation on both acceleration and braking phases since it showed a high correlation for both low and high percentiles. The angular velocity on x does not have a strong correlation for high values; this means that the stress is detected only when the car is turning on the left.

Both the skin conductance's phasic component and the respiration rate showed the same correlation results with the driving features, except for angular velocity on z, which was only detected for the phasic component and the nearest car on the back that was only detected for the respiration rate. Since we found a correlation with the angular velocity on x, the correlation with the angular velocity on z could make sense since it is another feature that provides information on how fast the car is turning. Instead, the correlation with the nearest car on the back might be a false positive detection since we had very few cars following us during the test. However, we cannot say precisely why we have this correlation difference between the two physiological indexes with our analysis, and a further investigation is needed.

In both the skin conductance's phasic component and the respiration rate analysis, the stressful driving features founded had a higher frequency in the autonomous drive scenario. It should happen due to the higher stress

level that we detected in the autonomous scenario. Indeed, in the manual and passenger scenarios, we found a lower stress level than the autonomous one. Consequently, we found a shallow frequency for the stressful driving features. It means that the driving features discovered in the three scenarios are not the same due to the difference in the stress level detected in the three scenarios.

Since in the autonomous scenario, we limited the car's speed to only 10-11 km/h, the higher stress level that we detected in the autonomous scenario than the other two scenarios is most significant. Indeed, If we detected a higher stress level in the autonomous scenario at a speed of 10-11 km/h, this means that at a higher speed, the stress should be much higher.

Finally, comparing the correlation analysis results and the questionnaire's answers, we discovered that the questionnaire answers confirmed the results extracted from the physiological sensors. Indeed, questionnaire answers confirmed that the autonomous drive is the most stressful scenario in both the degree of agitation and stress. Furthermore, the questionnaires highlighted that the participants felt a lack of control in the autonomous scenario. As we expected, the participant's degree of control was almost maximum in manual driving.

C H A P T E R

6. Conclusion and future research direction

The thesis work presented here brings new possibilities of study for the emerging domain represented by the passenger's comfort as a physiological issue. There is little available knowledge in the on-road passenger's comfort study, but it will become crucial for the future autonomous cars' acceptance and trust.

To address this purpose, we developed a platform that allows studying, in a real environment, the interaction between driver and car, focusing on the driver stress and the driving factor that impacts driver stress. Its design, from the sensors to the acquisition and processing software, was discussed in detail, aiming to provide a view of challenges existing in the domain.

The acquisition and processing software has been developed keeping the entire system modular and allowing the system to be flexible. For example, in our system, it is easy to add a new sensor with the relative ROS acquisition driver or add new driving measures or physiological indexes to the processing software. Such modularity is essential since it allows us to evolve the whole system easily and test several configurations.

We developed a software able to detect, using the K-nn algorithm, the probability of correlation between a time-series extracted from Physiological stress indexes and an event sequence extracted from Vehicle or Environmental's data. A correlation between a time-series and an event-sequence indicates an increase in the driver's stress level.

The on-field experiments that we conducted using our platforms allowed us to achieve our thesis's primary goal, which was to validate the entire system. Indeed, the correlation analysis results demonstrated that our platform

is able to acquire and process physiological and driving features extracting information on the driver's stress that are consistent for different physiological stress indexes and are confirmed by the questionnaire answers.

From all Physiological indexes considered in our study, the ones that obtained higher correlation were the respiration rate and the skin conductance's phasic component. The information obtained from the two physiological indexes is similar. Both demonstrate that the longitudinal jerk, the linear acceleration on y and the angular velocity on x-axis are the driving factors that impact the driver comfort/stress. This answered our research questions on which physiological data provides more information on the driver's stress level and the driving factors that impact driver comfort/stress.

With the respiration rate and the skin conductance's phasic component we demonstrated that the stress level in an autonomous drive is higher than a manual drive or a manual drive as a passenger. Questionnaire answers confirmed the higher stress level in the autonomous drive. Furthermore, the questionnaires proved that the higher stress level in the autonomous drive is also due to the lack of control.

Our research has provided a hardware and software platform that is the base from which we can perform different studies on the interaction between driver/passenger and car, expanding the little available knowledge on this field.

6.1 Future work

There are a series of future developments for the work we have conducted.

Firstly, we will make new acquisitions to have more data to study and have participants with different backgrounds and experiences.

With the data acquired and future data, we will develop and test new algorithms to detect stress, comparing different methods' results. We will also add to the analysis more stress indexes like the temporal analysis of the heart rate variability (HRV) and new driving parameters as the acceleration and braking pedals movements and the steer encoder.

This research is part of the I.DRIVE project where it is available a fixed-base simulator. The next step of our research will be to validate our simulator with the vehicle data. We have already done a first validation study [33] where we presented the methodological and conceptual description of a validation protocol. For the previous validation study, we used only the galvanic

skin response and the speed measures since, in the previous version of the set-up, the vehicle had only a GPS and a skin conductance sensor. We will validate the simulator with the actual vehicle set-up by comparing the physiological responses and the driving features in the same scenario. The future validated link between the I.DRIVE simulator and the vehicle will allow the development of integrated research paradigms, optimizing both experimental settings' pros and cons.

Another opportunity for improvements is the use of less invasive physiological sensors inside the vehicle. Indeed, our physiological sensors need electrodes that are connected by wires to the Procomp. These wires could annoy the driver during some maneuvers and force him to perform unnatural movements increasing the stress. We already tested two methods to measure the Remote PhotoPlethysmoGraphy (rPPG) that allows extracting cardiac information just by analyzing a video stream of a person's face.

In the first test, we used a high-rate (120fps) RGB camera to acquire the driver's face and tested it in a real scenario. It was revealed to be inaccurate due to light condition changes during the drive and poor face tracking.

In the second test, we used a Single-Photon Avalanche Diode (SPAD) camera in order to perform rPPG with higher accuracy, especially in low illumination conditions, exploiting the higher sensitivity of the SPAD sensors [27]. We tested it inside the car in a static condition and the results showed high accuracy. In the feature, we will test it on real scenarios to see if the accuracy stays high.

Also for eye tracking, there are systems based on cameras mounted on the cockpit of the car. An eye tracker based on cameras would avoid using glasses that could be annoying, especially for those who already wear glasses. An interesting future development will be to apply the correlation information between stress and driving features to an autonomous car local planner. Indeed, our system can detect the driving features that impact the driver's stress, giving the information on which value (percentile) of each stressful driving feature has a high impact on the stress. With this information, we could modify an autonomous car planner's parameters to make the passenger feel less stressed. The development of an autonomous car planner able to considered the passenger's stress information would help a broad acceptance of the car automation technology.

We are developing a local planner that will control our vehicle using the Model Predictive Control (MPC) framework [4]. With this approach, we will add, remove, or modify the control parameters' constraints with the infor-

mation coming from our platform, making the driving style more relaxing for the passenger.

Bibliography

[1] The most popular database for modern apps, . URL https://www.mongodb.com.

[2] ROS.org | Powering the world's robots, . URL https://www.ros.org/.

[3] Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. 2018. doi: 10.4271/j3016_201609.

[4] L. Bascetta, M. Farina, A. Gabrielli, and M. Matteucci. A feedback linearisation algorithm for single-track models with structural stability properties. *arXiv:2003.13793 [cs, eess]*, Apr. 2020. URL http://arxiv.org/abs/2003.13793. arXiv: 2003.13793.

[5] C. Basu, Q. Yang, D. Hungerman, M. Sinahal, and A. D. Draqan. Do You Want Your Autonomous Car to Drive Like You? In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*, pages 417–425, Mar. 2017. ISSN: 2167-2148.

[6] M. Beggiato, M. Pereira, T. Petzoldt, and J. Krems. Learning and development of trust, acceptance and the mental model of ACC. A longitudinal on-road study. *Transportation Research Part F: Traffic Psychology and Behaviour*, 35:75–84, Nov. 2015. ISSN 1369-8478. doi: 10.1016/j.trf.2015.10.005. URL http://www.sciencedirect.com/science/article/pii/S1369847815001564.

[7] H. Bellem, T. Schönenberg, J. F. Krems, and M. Schrauf. Objective metrics of comfort: Developing a driving style for highly automated vehicles. *Transportation Research Part F: Traffic Psychology and Behaviour*, 41:45–54, Aug. 2016. ISSN 1369-8478. doi: 10.1016/j.trf.2016.05.005. URL http://www.sciencedirect.com/science/article/pii/S136984781630064X.

[8] H. Bellem, M. Klüver, M. Schrauf, H.-P. Schöner, H. Hecht, and J. F. Krems. Can We Study Autonomous Driving Comfort in Moving-Base Driving Simulators? A Validation Study. *Human Factors*, 59(3):442–456, May 2017. ISSN 0018-7208. doi: 10.1177/0018720816682647. URL https://doi.org/10.1177/0018720816682647. Publisher: SAGE Publications Inc.

[9] H. Bellem, B. Thiel, M. Schrauf, and J. F. Krems. Comfort in automated driving: An analysis of preferences for different automated driving styles and their dependence on personality traits. *Transportation Research Part F: Traffic Psychology and Behaviour*, 55:90–100, May 2018. ISSN 1369-8478. doi: 10.1016/j.trf.2018.02.036. URL http://www.sciencedirect.com/science/article/pii/S1369847817301535.

[10] M. Benedek and C. Kaernbach. Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology*, 47(4):647–658, July 2010. ISSN 0048-5772. doi: 10.1111/j.1469-8986.2009.00972.x. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2904901/.

[11] G. E. Billman. The LF/HF ratio does not accurately measure cardiac sympatho-vagal balance. *Frontiers in Physiology*, 4, Feb. 2013. ISSN 1664-042X. doi: 10.3389/fphys.2013.00026. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3576706/.

[12] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom. nuScenes: A multimodal dataset for autonomous driving. *arXiv:1903.11027 [cs, stat]*, May 2020. URL http://arxiv.org/abs/1903.11027. arXiv: 1903.11027.

[13] M. Elgendi. Fast QRS Detection with an Optimized Knowledge-Based Method: Evaluation on 11 Standard ECG Databases. *PLOS ONE*, 8(9):e73557, 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0073557. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0073557. Publisher: Public Library of Science.

[14] I. Galimberti. I.DRIVE: elaborazione e analisi di segnali fisiologici per la valutazione dello stress alla guida. 2016.

[15] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? The KITTI vision benchmark suite. In *2012 IEEE Conference*

on Computer Vision and Pattern Recognition, pages 3354–3361, June 2012. doi: 10.1109/CVPR.2012.6248074. ISSN: 1063-6919.

[16] I. A. GmbH. CarMaker, Aug. 2020. URL `https://ipg-automotive.com/products-services/simulation-software/carmaker/`. Publisher: IPG Automotive GmbH.

[17] F. Hartwich, M. Beggiato, A. Dettmann, and J. F. Krems. Drive me comfortable: customized automated driving styles for younger and older drivers. 8. *VDI-Tagung "Der Fahrer im*, 21:442–456, 2015.

[18] F. Hartwich, M. Beggiato, and J. F. Krems. Driving comfort, enjoyment and acceptance of automated driving – effects of drivers' age and driving style familiarity. *Ergonomics*, 61(8):1017–1032, Aug. 2018. ISSN 0014-0139. doi: 10.1080/00140139.2018.1441448. URL `https://doi.org/10.1080/00140139.2018.1441448`. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/00140139.2018.1441448.

[19] F. Hartwich, C. Witzlack, M. Beggiato, and J. F. Krems. The first impression counts – A combined driving simulator and test track study on the development of trust and acceptance of highly automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, 65:522–535, Aug. 2019. ISSN 1369-8478. doi: 10.1016/j.trf.2018.05.012. URL `http://www.sciencedirect.com/science/article/pii/S1369847816306489`.

[20] J. Healey and R. Picard. SmartCar: detecting driver stress. In *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, volume 4, pages 218–221 vol.4, Sept. 2000. doi: 10.1109/ICPR.2000.902898. ISSN: 1051-4651.

[21] J. Healey, J. Seger, and R. Picard. Quantifying driver stress: Developing a system for collecting and processing bio-metric signals in natural situations. *Biomedical sciences instrumentation*, 35:193–198, 1999.

[22] J. A. Healey and R. W. Picard. Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems*, 6(2):156–166, June 2005. doi: 10.1109/TITS.2005.848368.

[23] M. Helander. Applicability of drivers' electrodermal response to the design of the traffic environment. *Journal of Applied Psychology*, 63(4): 481–488, 1978. ISSN 1939-1854(Electronic),0021-9010(Print). doi: 10.1037/0021-9010.63.4.481. Place: US Publisher: American Psychological Association.

[24] M. Kassner, W. Patera, and A. Bulling. Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct Publication - UbiComp '14 Adjunct*, pages 1151–1160, Seattle, Washington, 2014. ACM Press. ISBN 978-1-4503-3047-3. doi: 10.1145/2638728.2641695. URL `http://dl.acm.org/citation.cfm?doid=2638728.2641695`.

[25] D. N. Lee. A Theory of Visual Control of Braking Based on Information about Time-to-Collision. *Perception*, 5(4):437–459, Dec. 1976. ISSN 0301-0066. doi: 10.1068/p050437. URL `https://doi.org/10.1068/p050437`. Publisher: SAGE Publications Ltd STM.

[26] C. Luo, J.-G. Lou, Q. Lin, Q. Fu, R. Ding, D. Zhang, and Z. Wang. Correlating events with time series for incident diagnosis. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '14, pages 1583–1592, New York, NY, USA, Aug. 2014. Association for Computing Machinery. ISBN 978-1-4503-2956-9. doi: 10.1145/2623330.2623374. URL `https://doi.org/10.1145/2623330.2623374`.

[27] M. Paracchini, L. Marchesi, K. Pasquinelli, M. Marcon, G. Fontana, A. Gabrielli, and F. Villa. Remote PhotoPlethysmoGraphy Using SPAD Camera for Automotive Health Monitoring Application. In *2019 AEIT International Conference of Electrical and Electronic Technologies for Automotive (AEIT AUTOMOTIVE)*, pages 1–6, July 2019. doi: 10.23919/EETA.2019.8804516.

[28] R. W. Picard, E. Vyzas, and J. Healey. Toward machine emotional intelligence: analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(10):1175–1191, Oct. 2001. ISSN 1939-3539. doi: 10.1109/34.954607.

[29] M. Quigley, B. Gerkey, K. Conley, J. Faust, T. Foote, J. Leibs, E. Berger, R. Wheeler, and A. Ng. ROS: an open-source Robot Operating System. In *ICRA workshop on open source software*, volume 3, page 5, 2009.

[30] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You Only Look Once: Unified, Real-Time Object Detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 779–788, June 2016. doi: 10.1109/CVPR.2016.91. ISSN: 1063-6919.

[31] G. Rigas, C. D. Katsis, P. Bougia, and D. I. Fotiadis. A reasoning-based framework for car driver's stress prediction. In *2008 16th Mediterranean Conference on Control and Automation*, pages 627–632, June 2008. doi: 10.1109/MED.2008.4602162.

[32] G. Rigas, Y. Goletsis, and D. I. Fotiadis. Real-Time Driver's Stress Event Detection. *IEEE Transactions on Intelligent Transportation Systems*, 13(1):221–234, Mar. 2012. ISSN 1558-0016. doi: 10.1109/TITS.2011.2168215. Conference Name: IEEE Transactions on Intelligent Transportation Systems.

[33] D. Ruscio, L. Bascetta, A. Gabrielli, M. Matteucci, D. Ariansyah, M. Bordegoni, G. Caruso, and L. Mussone. Collection and comparison of driver/passenger physiologic and behavioural data in simulation and on-road driving. In *2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, pages 403–408, June 2017. doi: 10.1109/MTITS.2017.8005705.

[34] S. Scherer, A. Dettmann, F. Hartwich, T. Pech, A. C. Bullinger, and G. Wanielik. How the driver wants to be driven - Modelling driving styles in highly automated driving. 2015.

[35] M. F. Schilling. Multivariate Two-Sample Tests Based on Nearest Neighbors. *Journal of the American Statistical Association*, 81(395): 799–806, 1986. ISSN 0162-1459. doi: 10.2307/2289012. URL `https://www.jstor.org/stable/2289012`. Publisher: [American Statistical Association, Taylor & Francis, Ltd.].

[36] F. Shaffer and J. P. Ginsberg. An Overview of Heart Rate Variability Metrics and Norms. *Frontiers in Public Health*, 5, Sept. 2017. ISSN

2296-2565. doi: 10.3389/fpubh.2017.00258. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5624990/.

[37] A. M. Shahsavarani. The effects of Stress on Visual Selective Attention: The Moderating Role of Personality Factors. *Journal of American Science (ISSN: 1545-1003).*, 2013. URL https://www.academia.edu/4422402/The_effects_of_Stress_on_Visual_Selective_Attention_The_Moderating_Role_of_Personality_Factors.

[38] A. M. Shahsavarani, E. Azad Marz Abadi, and M. Hakimi Kalkhoran. Stress: Facts and Theories through Literature Review. *International Journal of Medical Reviews*, 2(2):230–241, June 2015. ISSN 2345-525X. URL http://www.ijmedrev.com/article_68654.html. Publisher: Baqiyatallah University of Medical Sciences.

[39] F. W. Siebert, M. Oehl, F. Bersch, and H.-R. Pfister. The exact determination of subjective risk and comfort thresholds in car following. *Transportation Research Part F: Traffic Psychology and Behaviour*, 46:1–13, Apr. 2017. ISSN 1369-8478. doi: 10.1016/j.trf.2017.01.001. URL http://www.sciencedirect.com/science/article/pii/S1369847817300311.

[40] G. J. Székely and M. L. Rizzo. Energy statistics: A class of statistics based on distances. *Journal of Statistical Planning and Inference*, 143 (8):1249–1272, Aug. 2013. ISSN 0378-3758. doi: 10.1016/j.jspi.2013.03.018. URL http://www.sciencedirect.com/science/article/pii/S0378375813000633.

[41] J. Taelman, S. Vandeput, A. Spaepen, and S. Huffel. Influence of Mental Stress on Heart Rate and Heart Rate Variability. In *IFMBE Proceedings*, volume 22, pages 1366–1369. Jan. 2009. ISBN 978-3-540-89207-6. doi: 10.1007/978-3-540-89208-3_324. Journal Abbreviation: IFMBE Proceedings.

[42] U. Technologies. Unity Real-Time Development Platform | 3D, 2D VR & AR Engine. URL https://unity.com/.

[43] B. L. Welch. The significance of the difference between two means when the population variances are unequal. *Biometrika*, 29(3-4):350–362, Feb. 1938. ISSN 0006-3444. doi: 10.1093/biomet/29.3-4.350. URL https://

academic.oup.com/biomet/article/29/3-4/350/256577. Publisher: Oxford Academic.

[44] P. Welch. The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15(2): 70–73, June 1967. ISSN 1558-2582. doi: 10.1109/TAU.1967.1161901. Conference Name: IEEE Transactions on Audio and Electroacoustics.

[45] G. F. Wilson. An Analysis of Mental Workload in Pilots During Flight Using Multiple Psychophysiological Measures. *The International Journal of Aviation Psychology*, 12(1):3–18, Jan. 2002. ISSN 1050-8414. doi: 10.1207/S15327108IJAP1201_2. URL https://doi.org/10.1207/S15327108IJAP1201_2. Publisher: Taylor & Francis _eprint: https://doi.org/10.1207/S15327108IJAP1201_2.

[46] G. F. Wilson, J. D. Lambert, and C. A. Russell. Performance Enhancement with Real-Time Physiologically Controlled Adaptive Aiding. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 44(13):61–64, July 2000. ISSN 2169-5067. doi: 10.1177/154193120004401316. URL https://doi.org/10.1177/154193120004401316. Publisher: SAGE Publications Inc.

[47] Y. Yan. traveller59/second.pytorch, Sept. 2019. URL https://github.com/traveller59/second.pytorch. original-date: 2018-10-03T23:05:52Z.

[48] Y. Yan, Y. Mao, and B. Li. SECOND: Sparsely Embedded Convolutional Detection. *Sensors*, 18(10):3337, Oct. 2018. doi: 10.3390/s18103337. URL https://www.mdpi.com/1424-8220/18/10/3337.

A P P E N D I X

A. Questionnaires

These are the pre and post driving questionnaires proposed to the experiment participant as defined in the protocol at subsection 5.2.2 to understand the participant backgrounds and experiences and if there has been any change after a driving scenario in driver's stress and mood.

A.1 Pre-Driving Questionnaire

1. What is your gender?
    - Male
    - Female
    - Other
    - Prefer not to say
2. How old are you? .....
3. Do you own a driving license?
    - Yes
    - No
4. How long have you had your driving license?
    - Less than 5 months
    - From 5 months to 1 year

- From 1 year to 2 years
- From 2 years to 5 years
- For more than 5 years

5. Do you or your family own a car?

   - Yes
   - No

6. How often do you drive a car?

   - Never in last 3 months
   - Twice a month
   - Once a week
   - Twice or three times a week
   - Everyday

7. The following questions relate to your experience as a **driver** in an everyday driving scenario. Please circle the appropriate number - there are no right or wrong answers.

   - I feel unsafe at high speeds:
     Not at all / Somewhat / Moderately / Very much
   - I get uncomfortable with heavier congestion:
     Not at all / Somewhat / Moderately / Very much
   - Getting close to other vehicles make me feel uncomfortable:
     Not at all / Somewhat / Moderately / Very much

8. The following questions relate to your experience as a **passenger** in an everyday driving scenario. Please circle the appropriate number - there are no right or wrong answers.

   - I feel unsafe at high speeds:
     Not at all / Somewhat / Moderately / Very much
   - I get uncomfortable with heavier congestion:
     Not at all / Somewhat / Moderately / Very much

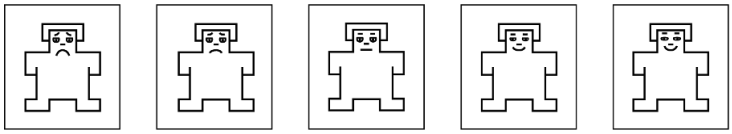
- Getting close to other vehicles make me feel uncomfortable:
  Not at all / Somewhat / Moderately / Very much

9. How do you consider your driving style?
   - Very Defensive
   - Defensive
   - Aggressive
   - Very Aggressive
10. Which is your favourite driving style as passenger?
    - Very Defensive
    - Defensive
    - Aggressive
    - Very Aggressive
11. Have you ever ridden an autonomous vehicle?
    - Yes
    - No
12. On a scale from 1 to 10 (being 1 **Not Trusting** and 10 being **Fully Trusting**) please rate your overall trust in autonomous vehicles and their safety:
    1 / 2 / 3 / 4 / 5 / 6 / 7 / 8 / 9 / 10

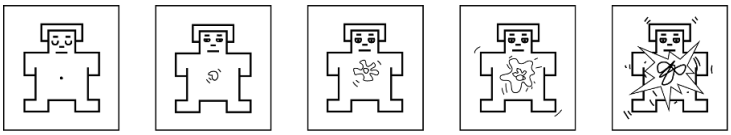
Emotional State Analysis
SAM (Self-Assessment Manikin) Scale

Choose the figure that better describes your emotional state

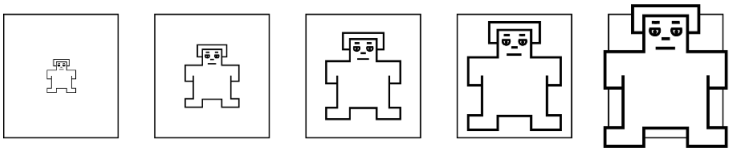
1. Emotive state
   From very distressed to very euphoric



2. Degree of agitation
   From very calm to very agitated



3. Degree of self-control
   From no control to full control of situation



A.2 Post-Driving Questionnaire

1. How do you rate the experience of autonomous driving?
   From 0 very negative to 5 very positive
   0 / 1 / 2 / 3 / 4 / 5

2. Did you feel comfortable wearing physiological sensors?
   From 0 very uncomfortable to 5 very comfortable
   0 / 1 / 2 / 3 / 4 / 5

3. Would you like to repeat the autonomous driving?
   From 0 absolutely no to 5 absolutely yes
   0 / 1 / 2 / 3 / 4 / 5

Please read each statement and circle the appropriate number to indicate how you felt during the drive about various situations and driving characteristics. There are no right or wrong answers. Do not spend too much time on any one statement.

1. Vehicle speed?
   Very Uncomfortable / Uncomfortable / Comfortable / Very Comfortable

2. Vehicle acceleration?
   Very Uncomfortable / Uncomfortable / Comfortable / Very Comfortable

3. Vehicle jerk?
   Very Uncomfortable / Uncomfortable / Comfortable / Very Comfortable

4. Vehicle speed on turns?
   Very Uncomfortable / Uncomfortable / Comfortable / Very Comfortable

5. Vehicle braking?
   Very Uncomfortable / Uncomfortable / Comfortable / Very Comfortable

6. Distances from other vehicles?
   Very Uncomfortable / Uncomfortable / Comfortable / Very Comfortable

7. Distances from pedestrians?
   Very Uncomfortable / Uncomfortable / Comfortable / Very Comfortable

Please read each statement and circle the appropriate number to indicate how you felt during the drive. There are no right or wrong answers. Do not spend too much time on any one statement.

1. I felt nervous
   Not at all / Somewhat / Moderately / Very Much

2. I felt jittery
   Not at all / Somewhat / Moderately / Very Much

3. My body felt tense
   Not at all / Somewhat / Moderately / Very Much

4. I felt tense in my stomach
   Not at all / Somewhat / Moderately / Very Much

5. My body felt relaxed
   Not at all / Somewhat / Moderately / Very Much

6. My heart rate was racing
   Not at all / Somewhat / Moderately / Very Much

7. I felt my stomach sinking
   Not at all / Somewhat / Moderately / Very Much

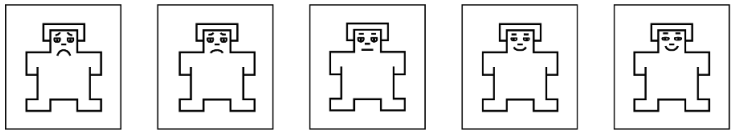
8. My hands were clammy
   Not at all / Somewhat / Moderately / Very Much

9. My body felt tight
   Not at all / Somewhat / Moderately / Very Much

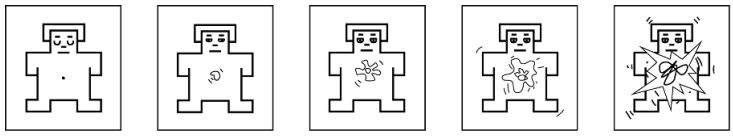
Emotional State Analysis
SAM (Self-Assessment Manikin) Scale

Choose the figure that better describes your emotional state

1. Emotive state
   From very distressed to very euphoric



2. Degree of agitation
   From very calm to very agitated



3. Degree of self-control
   From no control to full control of situation