Anna Stróż
Olga Zdzienicka

R O B O T E C . A I

HANDBOOK ON DRIVER STRESS DETECTION

Warsaw, 2023

ROBO

Anna Stróż
Olga Zdzienicka

ROBOTEC AI
Warsaw, 2023

THE ROBOTEC.AI

HANDBOOK ON DRIVER STRESS DETECTION

Published by:
Robotec.ai Sp. z o.o.
Koszykowa 61B St.
00-677 Warsaw
www.robotec.ai

ISBN 978-83-969303-1-6

Warsaw, 2023.
To cite the book, please use the following APA citation:
Stróż, A. & Zdzienicka, O. (2023). *The Robotec.ai Handbook on Driver Stress Detection*. Robotec.ai Sp. z o.o.

Date	Version	Authors	Task
29.08.2023	1.0	Anna Stróż, Olga Zdzienicka	Document ready for review
20.10.2023	1.1	Sylwia Mac, Anna Olejniczak-Serowiec, Mikołaj Pacek, Mikołaj Sokołowski Proofreading: Michał Torzecki, Justyna Toton	Review & proofreading
24.10.2023	1.2	Anna Stróż, Olga Zdzienicka	Final version

DTP:
STUDIO
grafpa
www.grafpa.pl

Table of Contents

List of abbreviations

ADAS	Advanced Driver Assistance System	**HPA**	Hypothalamic-Pituitary-Adrenal
ADB	Anxious Driving Behavior	**HRV**	Heart Rate Variability
AI	Artificial Intelligence	**IMU**	Inertial Measurement Unit
CFQ	Cognitive Failures Questionnaire	**k-NN**	k-Nearest Neighbors
CNN	Convolutional Neural Network	**MEG**	Magnetoencephalography
DALI	Driving Activity Load Index	**ML**	Machine Learning
DBI	Driving Behaviour Inventory	**MRQ**	Multiple Resource Questionnaire
DBI-Gen	Driving Behaviour Inventory – General	**NASA – TLX**	NASA Task Load Index
DBS	Driving Behavior Survey	**NDS**	Naturalistic Driving Study
DBQ	Driver Behavior Questionnaire	**PAT**	Pulse Arrival Time
DCQ	Driver Coping Questionnaire	**PD**	Pupil Diameter
DMS	Driver Monitoring System	**PPG**	Photoplethysmography
DSI	Driving Stress Inventory	**PSS**	Perceived Stress Scale
DSSQ	Dundee Stress State Questionnaire	**PTT**	Pulse Transit Time
DT	Decision Tree	**RBF**	Radial Basis Function
DVQ	Driving Vengeance Questionnaire	**RF**	Random Forest
ECG	Electrocardiography	**ROI**	Region of Interest
EDA	Electrodermal Activity	**SACL**	Stress Arousal Checklist
EEG	Electroencephalography	**SAM**	Sympathetic Adrenomedullary
EMG	Electromyography	**SDSQ**	State Driver Stress Questionnaire
EOG	Electrooculography	**SFS-SVM**	Sequential Forward Selection Support Vector Machine
F-DBQ	Freight Driver Behavior Questionnaire	**SPR**	Skin Potential Response
FFT	Fast Fourier Transform	**STAXI**	State-Trait Anger Expression Inventory
fMRI	Functional Magnetic Resonance Imaging	**SVM**	Support Vector Machine
fNIRS	Functional Near-Infrared Spectroscopy	**UMACL**	UWIST Mood Adjective Checklist
GSR	Galvanic Skin Response	**VAS**	Visual Analog Scale

Over

OVERVIEW OF THE HANDBOOK STRUCTURE

THE MAIN MOTIVATION BEHIND THE PREPARATION of this handbook was to search for answers not only to the question *"how to measure stress?"*, but to go even further and ask *"how to measure stress in a driver?"*. Such a challenge, which goes beyond the doors of a well-controlled laboratory and requires expert knowledge of measurement tools in the field of psychophysiology, is what our Human Factors team at Robotec.ai is passionate about. Based on our experience in studying drivers in variable experimental conditions and environments, we would like to introduce another interesting challenge that will hopefully grasp the attention of developers working on driver assistance and monitoring systems – the phenomenon of stress among drivers.

However, before getting to the point, we will start with a brief conceptualization of the stress phenomenon. What does it mean to be under stress? How do our bodies work then? Are there any cognitive or behavioral correlates to this state? Can we easily detect the period when we are stressed and quantify it? As a starting point, we would like to provide an overview of the stress-related matter, based on available research works and our expertise in scientific methodology.

Right after the "Stress 101" intensive course from the previous paragraph, we start our dive, targeting a specific use case for stress detection or prediction—driving-related context. In order to do so, we introduce you to two major ways to estimate stress—subjective and physiological assessment, and further discuss them, with examples derived from selected research works in the domain of driving under stress. After running through some fascinating research examples, you will probably notice that things are getting really complex—we are here to help, and this is why we also wrote a brief section about multimodal approaches, and discussed several applications of Artificial Intelligence algorithms in driving studies. Finally, we close the handbook with several conclusions, take-home messages, and an overview of the current market.

The document is divided into four pieces, each devoted to different aspects of the phenomenon of stress:

The introductory section S*tress – easy to feel, difficult to describe?* serves as a brief initiation to the concept of stress, with references to supplementary readings in this domain. We further narrow our scope of interest to the phenomenon of acute stress, which is the subject of subsequent investigations.

Chapter *Our bodies and stress – an overview* covers the more sophisticated image of stress from psychological and physiological perspectives. Further, we summarize selected stress assessment strategies both in subjective and physiological contexts.

Chapter *Diving into driving: on driver stress recognition* is the core part of the document, with a careful selection and discussion of research works in the domain of stress and driving studies. Please note that the handbook is not a systematic review nor a meta-analysis of existing research work. Its scope is limited to numerous examples of research work in that domain.

Chapter *Market overview and concluding remarks* serves as a space for comments on potential applications of stress assessment strategies in the driving context, and on eventual limitations inherent in assessing—sometimes very complex—cognitive states. This section also provides an overview of the commercial solutions tailored to stress detection or mitigation in the automotive context. Finally, we summarize the core points and concepts for the assessment of stress. If you reach this point, starting from the first chapter, congratulations!

Intro

INTRODUCTION

STRESS – EASY TO FEEL, DIFFICULT TO DESCRIBE?

STRESS IS A PHENOMENON PRESENT IN EVERYONE'S lives, occurring in different conditions and daily life situations, with varying intensity and period. We may face stress during different times of the day or because of certain events. Our experience of stress may vary, just like our ways of coping with it. Workplace, meetings, or human interaction contexts may influence stress as social or cognitive factors. Yet, there are also physiological, bodily stressors such as hunger, pain (e.g., Kogler et al., 2015), or uncomfortable temperature conditions (e.g., Baek et al., 2009). Although short-term stress may be crucial for our survival through metabolic induction of rapid reactions to stressors, longitudinal exposure to stress may negatively influence our well-being and general health status (e.g., Dhabhar, 2014; McEwen, 2008). Some diseases may be linked to chronic stress exposure, but the area of such clinical consequences, and the chronic stress itself, is outside the scope of the handbook. For general reviews on stress, please refer to e.g., Dhabhar (2014), McEwen (2008), Lupien et al. (2009).

Recent developments in consumer technology and personal health tracking (e.g., through wearable technology, mobile applications) are rapidly broadening the scope of possibilities to include sensors in digital monitoring of our health statuses. One of the fields of interest in digital health assessment is tracking the occurrence of stress. Currently, there is a great interest in the development of technologies that could help us mitigate the effects of exposure to stress or detect it and react onward. In some contexts, such as car driving, stress may seriously contribute to safety-critical events and crash risks. Early detection of stress-related behavioral consequences could help counteract its effects on road safety, or, as predicted in the case of semi-autonomous driving solutions, signal to the system to take control of the car. It is, therefore, interesting to what extent observable correlates of stress may help predict or detect its presence and further mitigate the influence of negative stress outcomes. In the present review, we would like to narrow its scope to the recent ideas and advancements in acute stress detection and prediction, with further potential application in the domain of automotive assistance and monitoring systems.

In the following chapter of the handbook, we lean into the phenomenon of stress from both psychological and physiological perspectives in order to grasp the basic knowledge necessary for further driver-oriented considerations. Then, we discuss existing ways to assess the occurrence of stress, among which, tentatively, we can distinguish subjective measures and techniques based on physiological data. The main part of this handbook is an elaborate overview of research works aimed at stress evaluation in driving conditions, in order to finally get to the place where we will discuss the potential possibilities of stress detection in drivers operating vehicles. In addition to summing up some of the existing subjective assessment techniques, we discuss the most popular physiological measurement techniques in the field. Further, we take a quick look at the stress classification problem itself, which, thanks to the recent boost in Artificial Intelligence (AI), especially Machine Learning (ML), seems to be much more “doable” than ever before. Some examples, arguments in favor, and limitations of AI in the setting we focus on are summarized.

In the last part of the handbook, we look at some of the current commercial attempts to apply stress estimation approaches proposed in the automotive industry. This chapter also serves as a conclusion, with a summary of the potential and limitations of technology in the estimation of driver stress.

Chap

CHAPTER 1.

OUR BODIES AND STRESS – AN OVERVIEW

BASED ON OUR OWN EXPERIENCES, WE can usually pretty accurately estimate when we feel stressed. However, in addition to our subjective assessment of the feeling of stress and its accompanying emotions, there are a number of processes in our bodies that prepare us to confront a stressor, and that will bring the body back to homeostasis. Thus, we will see that the overall stress state can be considered in several subcategories, with explanations of its mechanisms spanning from the cellular levels through higher levels of organization (e.g., physiological at the level of body systems) up to psychological or psychosocial contexts. In this chapter, we briefly highlight the key physiological mechanisms, as well as the effects observable in the operation of particular human body systems. We also turn our attention to some psychological correlates, considering the effects of stress on human cognitive functioning. We consider short-term exposure to stress, i.e., what happens when we are confronted with a stressor and immediately afterward. However, we do not address the correlates and consequences of long-term exposure to stress (for review see e.g., Marin et al., 2011; Romero-Martínez et al., 2020; McEwen, 2017).

It is also worth noting that difficulties exist in reaching a precise, operational *explanans* of "stress". Depending on the boundary conditions set for the definition, it may seem too broad or too narrow for different levels of explanation[1].

1 If you are interested in a thorough discussion of the concept of stress, and a critique of its current scope of usage, please refer to e.g., Koolhaas et al. (2011).

Due to lack of specificity, the term “stress” is frequently criticized (e.g., Karatsoreos, 2018; Koolhaas et al., 2011), and differences in conceptualization of stress may influence the discussion and reproducibility of experimental results. However, in this section we will not dive into the discussion of terminology and appropriateness of the term “stress”, as the primary focus is to describe some basic, well agreed upon aspects of stress mechanisms.

Another key piece of information that should constantly be kept in mind is that there can be significant inter- and intraindividual variability in the causes and effects that accompany the onset of stress, whether at the physiological, psychological, or behavioral level. It is also important to recognize the associations and differences in stress reactivity due to, among other things, personal characteristics (e.g., gender, age, genetic factors), and variables affecting the condition, such as sleep quality or medications intake (Moses et al., 2023). All of these factors, as well as differences in conditions and the experimental environment, should be remembered when making comparisons of results from different sources.

At the level of systemic organization, there exist two processing streams—the hypothalamic-pituitary-adrenal (HPA) axis and the sympathetic adrenomedullary (SAM) system (e.g., Koolhaas et al., 2011), which activation is crucial for the initiation of physiological processes that enable the management of the body’s resources in order to respond appropriately to a stressor (e.g., Tsigos & Chrousos, 2002; Turner et al., 2020). Hypothalamic and pituitary structures influence an increase in cortisol levels produced in the adrenal cortex (Chrousos, 2009 as cited in Turner et al., 2020). This hormone, in turn, influences glucose secretion and the activity of the sympathetic part of the autonomic nervous system (e.g., Dunn, 2008). The other key hormones in the process are adrenaline and noradrenaline (Chrousos, 2009 as cited in Turner et al., 2020). Consequently, results of the mechanisms enable the body to prepare for a response to a stressor by mobilizing the action of specific organ systems, including the cardiovascular system—with, for example, increasing blood flow in skeletal muscles or increasing heart rate (e.g., Herd, 1991), the respiratory system—e.g., increasing respiratory rate or occurrence of breathing irregularity (e.g., Grossman, 1983), the urinary functions—such as a need to urinate (e.g., Shimizu et al., 2021), or the muscular system, with increasing tension in some of the skeletal muscles (e.g., in trapezius muscle, see Nilsen et al., 2007; Wijsman et al., 2013, and frontalis muscle, see Nilsen et al., 2007).

The mobilization of organ systems in response to stress is known popularly as the “fight or flight” (plus “or freeze” among some of the later works) response, which signifies the body’s readiness to respond in relation to the stressor (Bracha et al., 2004; Germer & Neff, 2015), with “fight” response to face the stressor, “flight” response—to distance oneself from the stressor, or the “freeze” response—which is a response through which no defensive or escape action is undertaken. In popular scientific terminology, the “fight or flight (or freeze)”[2] response is associated with the activity of the sympathetic part of the autonomic nervous system, while the relaxed state, also known as “rest and digest”, is associated, in turn, with increased activity of the parasympathetic part of the autonomic nervous system, which promotes the occurrence of digestion and resting processes of the body (e.g., Wehrwein et al., 2016)[3]. This division serves only as a pictorial representation of the actions performed by the autonomic nervous system components, as many researchers point to more complex interactions in the maintenance of the body’s homeostasis (see e.g., Shaffer & Ginsberg, 2017; Wehrwein et al., 2016).

---

2 For a thorough discussion of the *fight or flight or freeze* conceptualization, see Bracha et al., 2004. For discussion of the concept of freezing, you can refer to Noordewier et al., 2020.

3 The popular-scientific terms “fight or flight” and “rest and digest” are not precise scientific descriptions, and the relationship between the sympathetic and parasympathetic parts of the autonomic nervous systems is more complex. For a discussion on that matter, please refer to e.g., Wehrwein et al., 2016. Furthermore, in the behavioral context, next to the “fight or flight” strategy, there is proposed the “tend and befriend” strategy, and there may occur gender-related differences in strategy adoption (Taylor et al., 2000 as cited in Mayo & Heilig, 2019).

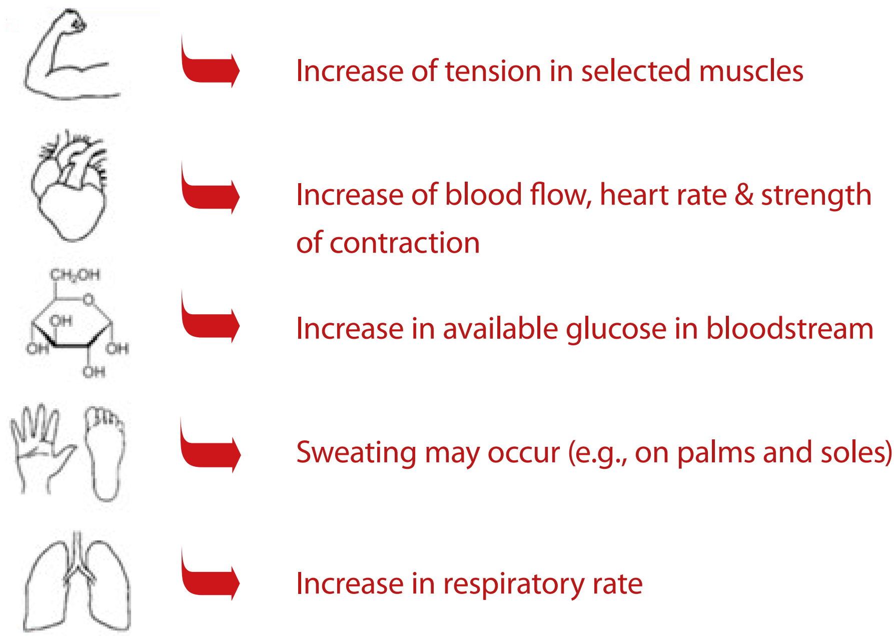


Figure 1.

Selected physiological correlates of stress responses. Please take into consideration existing inter- and intraindividual variability.

With that knowledge, we can move on to higher-level effects, that is, physiological correlates of stress in human organ systems. From the perspective of this handbook, the key point is that changes in physiological parameters may exist during the stress response, some of which can be measured under experimental conditions. Because of that, we aim to summarize a group of well-established observations before introducing you to selected stress evaluation techniques. Basically, first, we need to learn what may happen in the body, and then we can focus on selecting the proper testing environment and measurement techniques.

The abovementioned SAM system and the HPA axis affect a cascade of processes in the body, preparing it to respond to a stressor, yet please note that the intensity and direction of changes may be subject to inter- and intraindividual variability. In the case of the cardiovascular system, there are usually changes associated with an increase in the frequency (e.g., Lundberg et al., 2002; Nilsen et al., 2007) and strength of heart contractions, as well as changes in vascular throughput, resulting in the delivery of more blood to skeletal muscles (e.g., Herd, 1991; Carter et al., 2008; Nilsen et al., 2007)[4]. In a study by Ogorevc et al. (2011), participants were exposed to both mental and physical stressors, and authors found that the stress exposure influenced not only the increase of heart rate, but also of systolic (during heart muscle contraction) and diastolic (during heart muscle relaxation) blood pressure levels.

Interesting processes also happen in the muscular system during the reaction to a stressor. Selected groups of skeletal muscles may be characterized by increased activity—for the sake of trying to undertake some kind of bodily response to a stressor, this seems to be a reasonable consequence. Among the muscle groups identified by researchers is the trapezius muscle—it has been noted to have increased tension among those subjected to a stressful stimulus (Lundberg et al., 2002; Nilsen et al., 2007; Wijsman et al., 2013). A meta-analysis by Eijckelhof et al. (2013) reviewed a group of studies on the activity of selected muscle groups during work-related stress and noted increased activity of forearm muscles and neck-shoulder muscles. It is also worth noting the responses of mimic muscles (e.g., frontalis muscle, Nilsen et al., 2007). However, it should be emphasized that, as Mayo and Heilig (2019) pointed out, there may be large interindividual variability in how stress may influence facial expressions. Fatima et al. (2020) measured facial muscle and trapezius muscle activity during exposure to mental stress. While both muscle groups yielded significant results, trapezius muscle activity proved to be a better indicator than facial muscles. Thus, we can see that among selected

4 For a thorough review on cardiovascular activity, especially on heart rate variability, you can refer to Kim et al., 2018.

skeletal muscle groups, we can find potential candidates for measuring activities that may be related to the occurrence of stress.

Important processes happen in the largest—in terms of total area—system of our body, the integumentary system. We ourselves have probably experienced sweating palms or temples more than once during the experience of stress, for example, before a public appearance. Indeed, studies indicate increased sweating during the experience of stress, with particular emphasis on the palms and feet (e.g., Ohmi et al., 2009), but also the face and armpits (e.g., Harker, 2013)[5]. Jumping from skin to lungs, what happens in the respiratory system as we are reaching the stress phase? You may expect, among other things, an increased respiratory rate, as shown, for example, in studies by Hernando et al. (2015), Nilsen et al. (2007), Suess et al. (1980), decreased end-tidal CO_2 (Suess et al., 1980), or an increase in minute ventilation (Masaoka & Homma, 1997).

Aspects of body temperature changes that can accompany increasing levels of stress are warming up researchers (pun intended) interested in stress detection. Using thermal imaging of faces in the visible and infrared light spectra, N. Sharma et al. (2013) proposed a method of stress detection based on such an approach, with 72% accuracy in stress recognition rate. In a study by Levine et al. (2009), which was also oriented toward thermal data analysis, authors noted an association between the onset of stress and an increase in temperature at the site of the corrugator muscle, which is responsible for frowning. Another attempt to build a classification model with thermal imaging data was proposed in a study by Gioia et al. (2022). The authors observed, among other things, a decreased temperature at the nose tip and increased mean forehead temperature, which, like in the Levine et al. (2009) study, may be related to increased corrugator muscle activity.

5 See the review by Harker (2013) for an overview of the neural and biochemical aspects of psychological sweating.

The last of the physiological phenomena we describe here is what happens to pupils during stress. From the perspective of non-intrusive driver monitoring systems development, this may be a particularly important variable since these systems typically track features related to gaze. A study by Ren et al. (2014) involved the Pupil Diameter (PD) signal with the aim of assessing its usefulness in distinguishing between relaxation and stress states. Based on the classification results obtained, the researchers came to the observation that the PD signal performed better than the Galvanic Skin Response signal, which describes conductivity on the skin and can change under the influence of increased sweating, among others. In the paper by Sakamoto et al. (2009), authors examined the relationship between cognitive load, accompanying stress, and the properties of the PD signal. They noted that the PD signal analyzed in the frequency domain correlated with the state induced by cognitively stressful tasks. A similar observation, favoring the use of PD as a predictor in the assessment of mental stress, was proposed by Torres-Salomao et al. (2015), as well as by Zhai et al. (2005), with the latter work also allowing us to observe the value of the so-called multimodal approach, which will be discussed in one of the following chapters in the handbook. The added value of eye-tracking can also be noticed in blink frequency estimation, as exemplified by the work of Nagasawa et al. (2020), in which this variable was used in addition to PD to estimate stress levels.

The physiological mechanisms discussed above are among the well-described stress-related processes. If you are interested in the relationship between stress and specific body systems, we encourage you to see e.g., Kim et al. (2018), Wehrwein et al. (2016). In the next paragraph, however, we will focus on the cognitive-behavioral aspects that can accompany the onset of stress or occur afterwards.

As cognition is not a single entity, it would be better to describe the effects of stress or its correlates with respect to specific processing streams and functions. It is a relevant starting point for this summary, as the stress state may impact specific cognitive functions in a different manner and with varying intensity, which was mentioned e.g., in Sandi (2013). There is no single causal chain for exposition to stress, as stressors and the environment around us, internal, bodily states, and past experiences create a unique set of intertwined properties. Sandi (2013) discussed that notion and pointed out that the type of the stressor (e.g., mental, physical), its intensity, and co-occurring conditions may impact processing of stimuli differently. Some of the differences may correlate with other factors, such as age or gender (e.g., Moses et al., 2023; Sandi, 2013).

It is also worth emphasizing that when reading and discussing the results, attention should be paid to the health status of the study groups, as depending on whether it is an experimental study among healthy individuals or a study oriented to a specific population of patients, such as those with neurological damage, the results for cognitive functioning may differ, as mentioned e.g., by Moses et al. (2023)[6].

When it comes to the aspects related to memory, its subdivisions such as working memory, memory consolidation and memory retrieval may be differently affected by the occurrence of acute stress (LeBlanc, 2009). In the meta-analysis by Shields et al. (2016), authors concluded that stress may negatively influence working memory, although there may occur mediating effects of memory workload or gender. Proposed models of the ways in which working memory could be affected and therefore impaired, were described, e.g., in Moses et al. (2023), with a suggestion that the deactivation of the dorsolateral prefrontal cortex and Central Executive Network in general

6 The review paper by Moses et al. (2023) is a very thoroughly prepared resource on cognitive and emotional aspects in the light of research on the impact of stress. It contains a summary of commonly used psychosocial stimuli, descriptions of networks that may be involved in processing when one is exposed to a stressor, as well as a summary of reports with studies using neuromodulation in the preset context of psychosocial stressors.

may take part in this process. Several examples in favor of the hypothesis on the disruptive effect of acute stress on working memory were also described in the review article by LeBlanc (2009). However, in the work of Porcelli et al. (2008), the authors indicate that the results for the relationship between working memory and stress seem to be mixed, with works suggesting either no effect or a negative one. The eventual reasons for inconsistent results may be due to different methodological approaches and utilized techniques, stressor manipulations, or measures of participants' performance (see Porcelli et al., 2008). In their work, Porcelli and colleagues noticed that the worsened performance was associated with the workload put on working memory but not with the occurrence of stressors themselves. Contrary to the abovementioned suggestion by Moses et al. (2023), they observed increased activation of prefrontal cortex in stress-related conditions, but authors suggested that it may serve as a mechanism necessary to sustain some performance of working memory under such difficult circumstances as stress. For memory consolidation and retrieval interesting commentary was provided in LeBlanc (2009). In the case of moderate stress, or if there is a relationship between the content to be consolidated and the stressful experience, there may occur an improvement in terms of consolidation. When it comes to the recall of information, stress may negatively impact the process. However, it can depend on the task and the content themselves. Two other executive functions that were also the subject of the meta-analysis by Shields et al. (2016) are cognitive inhibition, which can be seen as a mental ability of restraining from running a process[7], and cognitive flexibility, which can be defined as "the capacity for objective appraisal and appropriately flexible action" (American Psychological Association, n.d.). Through the analysis of experimental works, authors concluded that both functions can be negatively affected by

7 A detailed explanation of the concept of cognitive inhibition can be found in the work of MacLeod, 2007. The author proposed a following definition of the term: "Cognitive inhibition is the stopping or overriding of a mental process, in whole or in part, with or without intention." (MacLeod, 2007, p. 5).

stress, with an exception for the phenomenon of response inhibition, which can be, on the other hand, improved in a stressful condition. Attention is another aspect of cognition that would be interesting to discuss. Based upon the work of LeBlanc (2009), it seems that the outcomes for attention during stress exposure can be mixed. In the work, the author attempted to explain the confusion caused by the observation that sometimes stress can improve attentional processes by narrowing the scope of processed information to the most relevant aspects, yet in different contexts it may negatively influence our abilities to distinguish relevant information. LeBlanc (2009) also points at another—very important from the perspective of driving studies—aspect of the so-called divided attention tasks, in which performance during exposure to stress may be indeed worsened. As in driving studies on stress and distraction, an approach known as dual task paradigm is commonly used, such knowledge about the influence of stress may be relevant to the study protocols.

Decision-making can also be negatively affected by the presence of acute stressors. An example can be found in the study conducted by Wemm and Wulfert (2017), in which 56 participants were first exposed to a stressful task (Trier Social Stress Test), and, subsequently, to a task involving decision-making (Iowa Gambling Task). Among the participants exposed to the stressful situation, authors observed physiological changes consistent with previously described results, such as increased heart rate. The stressful situation was discussed as negatively influencing the decision-making process, as the participants had a narrower perspective of decision consequences. The authors observed also some differences correlating with gender for the relation between the influence of stress and decision-making.

Aside from higher cognitive operations, stress may also affect processing of information from different sensory modalities. In the review article by Jafari et al. (2017), authors summarized that the existing results support the notion that acute stress may negatively affect auditory processing. Yet, not only hearing can be somehow affected. In the work of Shackman et al. (2011), authors tested hypotheses about the effects of task-irrelevant stress on visual processing. In an EEG-based study they found significant differences in time-locked EEG components between the threat and safe conditions, both for early and late processing, thus observing how our vigilance and visual attention may possibly be influenced by the threat of a stressor.

Having an overview of the very complex relationships between the state of acute stress and some bodily systems and cognitive functions we described, we can finally proceed to the next step—to investigate if, how, and to what extent we are able to track the occurrence of acute stress. We try to target the following questions: *What could be the most popular methods for that purpose? How can we categorize our "hammers", "nails" and "screwdrivers" in the existing "toolbox"? What are the arguments in favor and against the tools, i.e., subjective and physiological techniques?*

1.1. Exploring methods for assessment of acute stress

As we narrowed the scope of the review to some preselected levels of description for stress, we can right now try to target them by introducing the means to assess stress, which have been in use up-to-date, and could be helpful in description and tracking of that phenomenon. The "tracking" can be tentatively split into subjective and objective measures, and out of them, we narrow the scope to standard subjective assessment techniques, such as questionnaires or surveys, and for, let's say, more objectivity-oriented measures, we will follow the path investigating physiological data collected with non-intrusive techniques.

1.1.1. Subjective assessment: questionnaires and surveys

With a brief yet careful introspective look, we can usually recognize whether we are stressed at a given time. This is why the subjective assessment of stress is of great importance. Questionnaires and surveys are standard tools for collecting introspective reports in psychology. Over the course of several decades, various methods for reporting have been developed in the field of stress research. We can collect subjective answers on different aspects, such as perceived stress levels, susceptibility to stress, or risks related to stress. There exist both non-specific stress questionnaires and questionnaires tailored to the needs of specific groups.

In Section 2.1. *Driving studies: subjective assessment,* we discuss several examples of questionnaires in the driving context that target the stress phenomenon.

The development of tools for subjective assessment of stress has been driven by the impact of stress on various areas of life and its implications on

behavior in the most diverse types of situations. The development of stress research has certainly been important in clinical contexts, as well as in general behavioral research, or specifically narrowed to various fields of psychology—e.g., transport psychology (more details in Chapter 2) or industrial, work, and organizational psychology (e.g., Salas et al., 2017; Sharit & Salvendy, 1982).

Speaking of self-report studies, it is important to remember that they can take really diverse forms—they can be written or oral; collected on paper, online, or even via phone; retrospective or real-time-based, qualitative and quantitative, to name a few. For this reason, such type of research can include closed questionnaires, open-ended interviews, various types of scales, and even one-off written reports or protocols (Pekruna, 2020).

When, on the other hand, we talk about self-report surveys specifically on the topic of stress, it is important to pay attention to what exactly respondents can be asked about—here, we can talk either about questions on stressful responses themselves (so-called *stress responses*) or oriented toward exposure to events and situations that cause stress, recognized as *stressor exposure* (Crosswell & Lockwood, 2020). The choice of appropriate questions and tools depends on the premises of the study.

One type of tools for subjective measurement of stress are questionnaires of different kinds, which can be found in various studies. Those studies might involve specific age groups, and questionnaires themselves are often tailored for them—for example, the Adolescent Stress Questionnaire for the certain age group (Byrne et al., 1995, as cited in Byrne et al., 2007), or Undergraduate Stress Questionnaire for people with certain educational backgrounds (Crandall et al., 1992) or professions (Psychology Student Stress Questionnaire, Cahir & Morris, 1991). In addition, because work is often a source of daily both short- and long-term stress, a variety of questionnaires related to self-reported stress at work have been developed, e.g., Caplan's Job Stress Questionnaire (Caplan et al., 1975, as cited in Harris et al., 1999) or Work Stress Questionnaire (Holmgren et al., 2009).

As was already mentioned, stressor exposure is one of the variables considered in stress research. This is a variable that can be examined with the use of questionnaires such as the Life Event Checklist (Gray et al., 2004), or the Stress and Adversity Inventory (Slavich & Shields, 2018). For each stressor to which the participant was exposed, additional questions are asked, such as the age of contact with the event in question, how long it lasted, or how stressful it was. However, due to differences in the way people may perceive certain situations, it is not easy to design a list of questions that would cover all potentially stressful events.

Often, responses and reactions to stressful events can be a more interesting object of study than the events themselves. Responses to stressful events or, in general, perceptions of specific situations and events as stressful or not can be highly individual, and thus the reactions themselves can be extremely varied. Such reactions can be both physical—for some, the reaction to stress is headache, migraine, abdominal pain, or dizziness (Healey & Picard, 2000), and psychological—here mood swings, general fatigue, irritability, or nervousness can be given as examples (Natvig et al., 1999). Therefore, sometimes more critical than the stressor exposure itself may be the individual response to stressful situations—this refers to the second variable mentioned earlier, namely the response to the stressor. In this context, in addition to the other more specific questionnaires mentioned earlier, there are also more general ones, such as the well-known global measure of stress, used precisely to measure response to the stressor—the Perceived Stress Scale (PSS, Cohen et al., 1983). The PSS is a tool that departs from the idea of testing exposure to a stressor and is used to assess the degree to which certain situations or events are stressful for the respondent. It captures respondents' perception of how overwhelmed they are by their current life situation and events. It originally used events taken from a certain period back in the respondent's life, and based on 14 questions with a 5-point Likert-type scale, it assessed how stressful the events given by the respondent were for them. In this way, the researcher moves away from general questions and statements and relies entirely on the respondent's evaluation of both the events themselves and their impact on them.

The number of questionnaires designed and developed specifically for stress research is so large that it would be impossible to describe, or at least list all of them, and this is not the primary goal of the handbook. Yet, they are not the only subjectivity-oriented tools used in stress research, as in such studies, researchers also look for various correlates of stress, for example, specific personality traits or coping strategies. Such questionnaires are usually not tailored explicitly for stress investigation, and yet may be helpful in the investigation of personality or behavioral covariates—usually in combination with other tools, and therefore can also be included in stress research. An example of a tool used in this way is Spielberger's State-Trait Personality Inventory (Spielberger, 1979), which helps assess several emotional conditions, e.g., anxiety or anger (e.g., Van Wijk, 2014).

To conclude, the primary benefit stemming from the application of subjective measurement techniques is that we, researchers, have the ability to reach some insight into participants' subjective experiences, such as their thoughts, feelings, and beliefs, which would not be currently possible with other methods of data collection. Meaning that subjective methods, by definition, provide insights into the subjective feelings of the participants that cannot be obtained by any objective measures in any way. This is an undeniable and huge advantage of this type of research, as well as an advantage over objective measures.

Another undoubted advantage, especially for researchers, is the economic aspect of self-report studies. Conducting an interview or doing a questionnaire, whether on paper or online, is inexpensive and can be done in different, also out-of-laboratory, environments. Opting for self-report studies is a fairly simple way to collect data from many people quickly and at a low cost with the use of fewer resources than for other research techniques.

Additionally, this type of data can be collected in a variety of ways—both online or over the phone, as well as in person, by interview or questionnaire, orally or in writing. It gives researchers tremendous opportunities to tailor the tool and form of data collection to their needs, optimize the entire process, and create a truly naturalistic environment for the study if needed. In this way, they can gather data from people spanning both large geographic areas and various social, age, or ethnic groups to which they may not have direct access. Additionally, self-report studies are easy to administer and can be completed quickly.

Self-report research methods, like any other research tools, have advantages, but there also exist limitations that we need to keep in mind. Opposed to the possibly wide geographic and ethnic range of self-report studies, in such a paradigm there may arise a problem of cultural or language barriers. Such studies might not be useful for participants who struggle to comprehend the language used in a questionnaire or survey. Aside from language comprehension, another important aspect lies in perception of different stressors, which possibly can be influenced by culture or the way the questionnaire is constructed. All these factors can impact the accuracy of responses, particularly in studies with culturally diverse populations. However, this is not an insurmountable problem—it just means that this type of research requires a decent cultural and linguistic adaptation.

Also, the validity of questionnaires as a research tool can and does come into question. Research based on self-reporting is inherently biased due to the feelings and emotions of the participant at the very moment of completing the questionnaire. Participants may always be prone to bias and, therefore, answer questions, e.g., in a socially desirable way. They may forget important details or knowingly lie—thus, in such studies, the honesty of the respondents is the key, and careful preparation of questions is needed to minimize eventual biases. It is a bargain that the researcher must consciously decide on. The possibility of receiving not entirely truthful information from the participant versus insight into the participant's mind—most researchers, however, decide one way or another that the pros outweigh the cons, or combine the tools to cross-check for the eventual biases.

Another shortcoming may be the respondents' potential lack of ability to give a reasonably accurate and relevant answer (e.g., answering complex questions in a perfunctory manner). To address this issue, researchers usually take care of a proper explanation of the study's relevance so that participants feel that they should be compliant. If such an approach fails, inspection of outlier responses can be performed. Yet, such a task requires great care—cross-check questions, for example, may help investigate the participant's non-compliance (e.g., through monitoring the presence of contradictory answers).

In addition, such studies may fail to accurately capture what they are intended to measure. Participants may not always be aware of their biases or be unable to accurately report on their behavior or attitudes.

Of course, tools and methods are being developed to compensate for the natural shortcomings of self-report-based research, but doubts remain. In summary, self-report surveys can provide valuable insights into individuals' subjective experiences, but researchers must be aware of their limitations and potential biases. Self-report surveys are an incredibly valuable tool in any type of social or psychological research. Additionally, they can be combined, both with each other and with objective methods, so they can be used to create extremely comprehensive and complete studies and produce valuable data and results.

1.1.2. Physiological *echoes* of stress

Frequently, physiological monitoring is considered as a means of receiving "more objective" indices, in contrast to the subjective one, which we can target with reporting (see Section 1.1.1. *Subjective assessment: questionnaires and surveys*). As we observed in the previous sections, different physiological correlates occur during exposure to stress. Some of them could be monitored through nonintrusive technology and their overview would be the leading

topic of the section. With technological advancements observed in the past several decades and the rapid development of nonintrusive data collection solutions, we reached the point in which physiological data collection does not necessary entail a complicated protocol and very expensive equipment. With that being said, we can proceed to an overview of several physiological measures that are frequently recognized in studies within the domain of stress research.

Keeping to the chronological order of our considerations, we will start with the means for estimation of cardiovascular system activity. There are several techniques being recognized in this field, with the major interest in application of

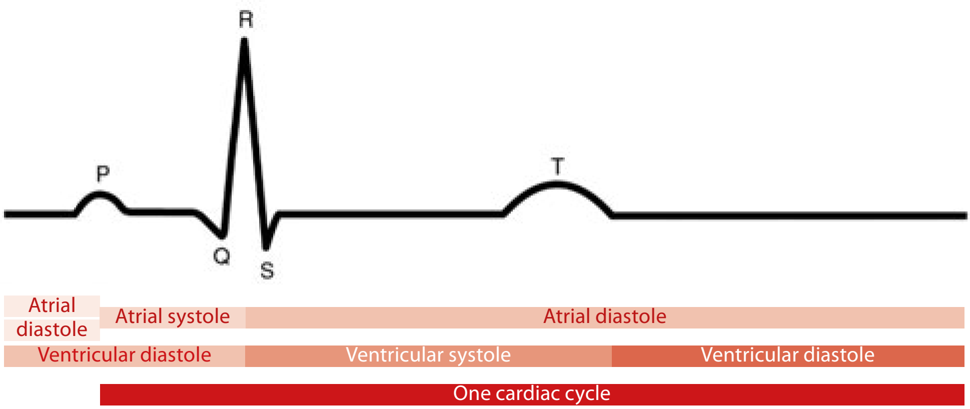


Figure 2.

A pictorial description of what actions occur during a single cardiac cycle and how it is reflected in ECG signal. Source: Betts et al. (2022). Licensed under CC-BY 4.0.

electrocardiography (ECG) and photoplethysmography (PPG). Electrocardiography is a technique common in clinical settings, helping to diagnose various pathological conditions in the cardiovascular system. In research settings, ECG is usually being applied in a simplified manner, with a lower number of leads—for example, three leads in the case of Einthoven montage instead of conventional, clinical twelve leads.

ECG involves the application of surface electrodes on the participant's body, through which values of electrical potential between pairs of electrodes are being collected, to be further amplified and digitized. Changes in electrical potential are the observable correlates of a cardiac cycle, during which heart muscle contracts and relaxes, pumping blood over the arteries in a quasi-periodic manner (see Figure 2). Afterward, recorded ECG time series can be analyzed in multiple ways, and there exist tens of standard parameters calculated upon the signal (for review see e.g., Shaffer & Ginsberg, 2017; see Table 5 in Section 2.2 for a description). Next to the popular measure of heart rate, heart rate variability being analyzed in time-, frequency-, and nonlinear domains may result in a relevant set of information derived from the signal[8]. Descriptions of the selected parameters can be found in Table 5 in Section 2.2.

Photoplethysmography is a different technique from ECG, as functionally, it helps to track changes in blood volume and flow in the vessels, which to some extent follow the quasi-periodic character of the cardiac cycle but are also sensitive to the influence of body thermoregulation and respiration (Allen, 2007). The physical-physiological basis for that measurement results from the optical properties of blood (such as e.g., light scattering or absorption), and the recording setup has two key elements—light-emitting diodes and photodetectors, with the latter collecting "what was left" from the emitted light source (for a detailed description, please refer to Allen, 2007)[9]. Optical aspects are thus posing some limitations for the selection of a proper light source for PPG, which is usually in the near-infrared spectrum. The resultant time series collected via PPG presents changes in the absorption of light emitted from the optical device, which is usually mounted on the participant's fingertips or ear. The obtained signal is further analyzed in terms of slowly and rapidly occurring changes, as the sources of various components may be associated with different physiological phenomena (Allen, 2007). Several popular PPG parameters are described in more detail in Table 6 in Section 2.2.

We can now move to the muscular system and ways to assess the changes that may occur during the stress responses. As we saw in the previous

8 If you are interested in diving into more details about ECG, you can read one of Robotec.ai's company blogposts, serving as a great starting point for further readings on ECG. For that matter, please see Kotynia & Stróż (2023).

9 The topical review by Allen (2007) is an excellent source of information regarding the physical, physiological and applicational contexts of PPG. We highly recommend the source work for a thorough understanding of the PPG technique.

chapter, several groups of muscles were found to be changing their activity level during the presence of stressors, both mental and physical. How could such activity be measured? For that reason, we can refer to surface electromyography (EMG), which is a non-intrusive technique, contrary to intramuscular EMG (iEMG), for which electrode needles have to be inserted through the skin surface. The latter gives better signal quality and precision of measurement, but due to its invasiveness, iEMG is applied only in clinical contexts. Surface EMG, similarly to ECG, uses one or more pairs of electrodes, and these are placed on the skin areas above the investigated muscle. With electrodes of a smaller diameter, one can also successfully record EMG signal from facial muscles, such as corrugator muscle, mentioned earlier. During contraction and relaxation of muscles, changes in electrical potential can be observed, and such time series can be analyzed in terms of temporal or spectral properties, or both, resulting in time-frequency representations.

Skin, or more generally, integumentary system, is another candidate for assessing and tracking stress-related changes. As we already know, sweating may occur during stress. The activity of sweating glands is a contributing factor for the signal derived with a measurement technique known as Electrodermal Activity (EDA)[10]. Over the years, since the first measurements of skin electrical properties, different terms were developed to describe these phenomena (e.g., Galvanic Skin Response) leading to a lack of standardized nomenclature, but efforts were made to propose a single (Johnson & Lubin, 1966, as cited in Tronstad et al., 2022; Critchley, 2002), umbrella term for "changes in electrical conductance of the skin […], that result from sympathetic neuronal activity" (Critchley, 2002, p. 132). During increased sympathetic activity (e.g., because of a stressor), there may occur significant changes in

10 A common view that sweat on the skin surface is the only causal factor for EDA is being confronted by evidence in favor of more complex interactions occurring in the skin. For a comprehensive review on that matter, please refer to Tronstad et al. (2022). It is also important to emphasize the possible existence of complex interactions between the brain and EDA, as described e.g., in Picard et al. (2016).

skin conductivity (and therefore, resistivity), changing the local environment for the flow of electric current. There are both active and passive means to measure EDA, with active measuring conductivity/resistivity using a very low current passing between the electrodes, and passive measuring skin electrical properties as they are (for review see M. Sharma et al., 2016).

In the EDA signal, we can recognize two major contributions—rapid changes due to external or internal causes (e.g., stimulation), known as phasic changes or responses, and slower fluctuations, known as tonic changes or levels. A set of various parameters can be derived from the signal, with a major focus on its amplitude, event durations, but also on properties of the response curve, such as slopes or total areas (for review of EDA data processing methods see Posada-Quintero & Chon, 2020).

Activity of the respiratory system can also provide us with valuable information during or after a stress response. In order to measure respiratory properties, various techniques were developed, differing in precision or in parameters under investigation. In research settings, devices based on periodic chest expansion are used frequently due to unobtrusiveness in their application. A carefully prepared review article by Liu et al. (2019) covers the broad area of tools used for respiratory system performance assessment and algorithms, which can aid with the derivation of respiratory information from some of the other biosignals. The standard parameters obtained from such signal are frequency of breaths or their intensity during stress exposure (e.g., Masaoka & Homma, 1997; Suess et al., 1980).

Body temperature can also serve as a peripheral signal assisting in stress estimation, which was exemplified in one of the previous sections. For that purpose, thermal cameras or wearable body thermometers can be used. What is usually being investigated is the change of local temperature in comparison with local baseline, as different body parts tend to have a variable temperature,

and thus a focus on particular Regions of Interests (ROIs) is needed. Analysis of thermal data requires an understanding of various image processing techniques. However, the recent development of AI-based tools may facilitate, for example, the integration of thermal and RGB images, if needed (such an approach was described in, e.g., Cardone et al., 2020).

In order to measure pupil-related changes during stress, the most popular technique is eye-tracking. Eye-tracking relies on image-based processing, facilitating the identification of such areas as eyelids, eye corners, eye and its components or features—iris, pupil, or Purkinje images. For example, with eye-tracking one can receive data about pupil diameter (PD) and its changes, and therefore can investigate correlations between PD and stress. However, as the main purpose of an eye-tracking application is to estimate a participant's gaze vectors, such a tool can also be utilized to assess behavioral correlates during stress. It is possible that our visual attention and the way we focus or distract may change during stress exposure, yet to our best knowledge, no particular indices were established so far. Some research cues can be found in the works targeting cognitive workload—see, e.g., the influence of mental stress on the gaze-click patterns in M. Huang et al. (2016), or driver gaze behavior during increased cognitive load, resulting in the reduction of peripheral vision (Reimer, 2009), but research at the intersection of mental workload, stress, and gaze patterns is still needed.

Last, but not least—brain activity. For its observation, researchers may attempt the challenge through various means, such as functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG), functional near-infrared spectroscopy (fNIRS), or electroencephalography (EEG). However, the first two methods are rather limited to strict laboratory settings due to immobility of the measuring devices and required experimental conditions. Yet, still, it should not be a problem for study protocols with a static/passive setup, in which the participant is asked to watch a recording or solve a simple task using buttons. On the other hand, the last two methods can be used out of experimental room, but not without limitations. Still, as the muscular activity would be the major source of signal artifacts, it is usually recommended to perform non-dynamic studies, with participants being seated. fMRI and fNIRS rely on the same physiological phenomenon, which is a flow of oxygenated and deoxygenated blood in the brain vessels, but the physical background of measured signals is different. It goes analogously for MEG and EEG, as both methods are oriented toward the measurement of activity from the cortex and yet focused on two sides of the phenomenon—measurement of magnetic fields and electrical fields over time, respectively.

As an example of stress research incorporating neuroscientific methods, in work by Wang et al. (2019), the authors studied the influence of acute stress on creative processes in a sample of young male participants. Using EEG, they found significant differences in the power of the upper alpha band (one of the frequency bands usually recognized in EEG data processing, defined in the study as 10–12 Hz) in pre- and post-stress conditions, which they associated with other measures, including salivary cortisol and alpha-amylase, to infer about the impairing influence of stress on a creative process. Other than EEG, also fNIRS and fMRI have already found their applications in study settings created for stress-inducing paradigms, with examples from Kalia et al. (2018) for fNIRS and cognitive flexibility testing, or Zhang et al. (2019) for fMRI in a large sample study of connectivity patterns before and after stress exposure.

As all the abovementioned methods look perfectly pretty on paper, and we can imagine that ideal stress recognition is at our fingertips (pun intended), in reality things are getting way more complicated. There are several key points that should be taken into consideration as barriers in physiological data collection: artifacts, inter- and intraindividual differences, and ecological validity.

In psychophysiology or, more generally, in the area of biological data acquisition, artifacts are recognized as disruptions occurring in a given dataset whose source is different than the one we would like to measure. Let's put it into a specific use case. While we record ECG or EEG data, muscular activity from limbs or extraocular muscles activity and eyeball movements—respectively—may be co-recorded in the signal without our intention, and therefore, we are accounTable for overseeing such possibilities. Motion seems to be the primary source of artifacts in different study settings. Consequently, it is highly recommended to limit the participant's movement to the necessary minimum to perform study tasks and take care of the proper study setup so that the leads and the participant's movement will not additionally interfere with each other.

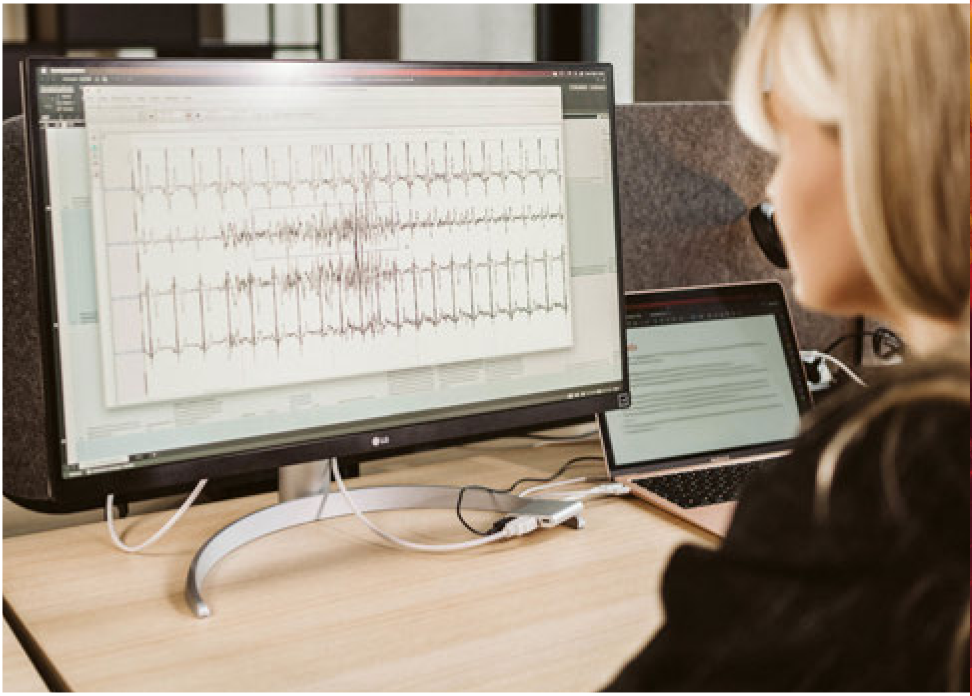
A proper setup design may be of great help on many levels, as we can PREVENT, PREDICT, and PREPROCESS—let's call it a 3P RULE for now. Through PREVENTION, we can think about any aspects of the procedure environment that can be prepared in a manner limiting the contribution of artifacts, such as proper seating of a participant, adequately scheduled breaks in the study, and an accurate montage of recording tools, being non-intrusive for the participant, but also minimizing the chance of leads' movement, for example. Through PREDICTION, we can consider carefully

what could possibly happen in the study in terms of internal artifacts—those that are not associated with the environment and cannot be solved with prevention, and what we can do to control such influence—for example, if we record EEG and expect that participants will blink frequently for some reason, it would be great then to co-record electrooculographic (EOG) signal for the further step—PREPROCESSING. Prediction can help us to solve some of the technical issues that will arise in the preprocessing step, as having co-registered signals can help to disentangle signal sources in the data, which are affected by the internal artifacts. The preprocessing phase is very challenging because, in this step, we need to clean the data and prepare it for subsequent analysis. Expertise in the interpretation of biosignal data is of great importance, as we need to understand the physiological basis, transfer it to clarification of what we observe in the signal, and eliminate the influences coming from different sources than the one of our interest. For each of the discussed biosignals, there are usually sets of standards or rules of thumb in terms of primary signal filtering in order to get a readable output of the recorded physiological phenomenon. Co-recorded signals from the prediction step can be not only analyzed on their own, but also help to clean other potentially affected signals. In the case of EEG, having EOG co-recorded, we can manage to statistically separate the influence of ocular movements activity on EEG through the means of such techniques as, for example, Independent Component Analysis (ICA). This is just an illustration of how important it is to have a global-local perspective on study protocols. However, we need to remember that not the methods, but the hypotheses drive the study, and the methods are simply the tools that we need to integrate during the operationalization phase. On the other hand, if the experimental protocol is under preparation and we would like to apply a certain method in the study, we need to make sure that the requirements for the method used are also met. For example, suppose we would like to record ECG and compare between conditions. In that case, the conditions should be prepared in a manner enabling us to interpret the data further correctly, e.g., not being too short, as the resultant parameters will not correctly "grasp" the physiological phenomenon (e.g., a recommendation for HRV analysis to have at least 5 minutes of recording, see Heathers, 2014). It is also important to emphasize the role of baseline recording for physiological data. There are at least two reasons for that—if we want to see whether *A* influences or co-exist with *B*, the baseline or control condition is needed; also, inter- and intraindividual variability can be an important factor, so if we want to observe any group effects, the baseline is necessary to see from what "place" we start.

Inter- and intraindividual variability are the two important limitations also for the physiological data interpretation, even if those are frequently seen as "objective" measures. Responses to stressors vary to some extent among participants, but may also vary in a single person, because of different factors such as sleep quality or medication intake over time (see Chapter 1. *Our bodies and stress – an overview*). With longitudinal or repeated-measures study, it is necessary to consider such factors as possible moderators of the effects. Conducting a baseline session may at least serve as a referential measure in each study separately.

Another limitation, which could be observed in stress research as well, is that when using physiological data acquisition, we need to follow some rules that, as a result, may impact the ecological validity and interpretability of the results. Subjective assessment may also face previously mentioned issues, e.g., with biases or answering in a socially desirable manner. Physiological recording, on the other hand, may limit the participants' freedom of movement due to the recording setup, influence their personal ways of responding to stressors, or even induce additional stress because of the laboratory setting and strict protocol. These can be seen as serious limitations on the path to a naturalistic approach in data collection. However, there are possible ways to try at least to minimize such contributions. One of them is a promising development of wearable technologies for physiological data acquisition, which could be beneficial in terms of participant's overall comfort throughout the study. The second suggestion is related to the previous statements about prevention—having a clear and well-prepared protocol for the study, we can actually help participants immerse themselves in the environment and minimize the "laboratory experience" through careful cooperation and explanation of the study steps.

Now, we will *immerse* ourselves in Chapter 2, which will be specifically devoted to stress recognition in the context of driving. As our main interest is in



searching for possible correlates, which could be helpful in the development of automated systems for stress recognition, we will go through some of the measures introduced above, then dive into them and describe in detail some of the parameters, which could be potentially used in future testing of stress detection solutions.

Chap

CHAPTER 2.

DIVING INTO DRIVING: ON DRIVER STRESS RECOGNITION

AS WE LEARNED ABOUT THE FOUNDATIONS of stress and some basic tools for its assessment, let's now narrow down the context to the specific use case—identification of stress in driving context.

In recent years, recognition of driver stress has attracted attention in the automotive industry, in particular due to the development of Advanced Driver Assistance Systems (ADAS), and specifically Driver Monitoring Systems (DMS). DMS are recognized as a promising tool to address human-related risk factors in road safety, such as drowsiness, distraction, or stress. Through estimation of the driver's visual attention targets and the overall level of arousal and alertness, DMS can classify in real-time whether the driver retains the ability to drive safely. Different strategies can be developed upon recognition of the driver's state of arousal—these could be alerting messages, delivered through one or more of several modalities (e.g., tactile, auditory, or visual), or—in the case of semi-autonomous vehicles—these could be safety-oriented maneuvers, aiming at finding a safe spot in which the vehicle can be stopped, thus limiting the risk of safety-critical events.

A challenging question arises: how would the systems be able to recognize that the driver is under stress? Most of the DMS work on the driver's image data. Can stress be easily assessed in such a setup? This is not an easy question, as the recorded purely behavioral data can lead to ambiguous conclusions on someone's state. Therefore, there is a need for referential measures, which could be good candidates for stress recognition on the basis of earlier research efforts, in which they were co-recorded in stress-inducing paradigms. Having such referential measures (also recognized as “ground truth”) can be valuable for further investigation of the relationships between the state of stress and the observable behaviors, which could be recorded with DMS cameras. Referential measures could then serve as some sort of “labels” in the subsequent steps of DMS development and training of the algorithms. Still, it is not an easy task, as there are absolutely important factors of inter- and intraindividual variability that should be taken into consideration. The phenomenon of stress is susceptible to these aspects, and even if there are group tendencies in the recorded information, there still could be difficulties in its interpretation and implementation on the level of a single driver.

In this section, we describe some of the exemplary findings from existing research works aiming at the recognition of stress in a driver-related context. Please remember that this work is an overview, not a meta-analysis or a systematic review. As you will see, referential measures will aim to cover both the subjective and objective aspects of stress responses. Using these two perspectives may be crucial for a successful research project in the domain of stress recognition.

Another trend, which is worth noting, is that most of the reviewed studies attempted to target the phenomenon of stress by the use of more than a single referential measure—such approaches are recognized as multimodal, and aim at grasping various levels of information, to further make stress recognition more accurate, based on multiple information sources. Multimodal approaches and some existing examples, including the usage of Machine Learning techniques, are covered in more detail in Section 2.4.

Before we delve into examples of driving stress studies for which subjective and psychophysiological measurements found their applications, let’s turn our attention to the issue that will repeatedly be mentioned in discussed studies—the stressors. As we already know, to study stress under controlled conditions, it is also essential to understand how stress can be induced. So far, we have seen that the main dividing axis separates stressors into physical and mentally taxing ones. Based on observations and the publications we refer to in subsequent paragraphs, we note that mental stressors are much more frequently used in studies of stress in drivers. From the area of physical stressors, we can find studies in which, for example, the temperature inside the vehicle is manipulated (e.g., Baek et al., 2009). The spectrum of mentally taxing stressors is very broad and could merit a separate publication. In research related to stress and driving, stressors in the form of tasks that tax working memory (e.g., mental arithmetic tasks, *Trier Social Stress Test,* or *N-back*) are relatively common (e.g., Balters et al., 2021; Huang et al., 2020; Lanatà et al., 2014), and next to them stimuli directed at attentional processes (e.g., visual search in Anzengruber & Riener, 2012), or cognitive-manual resources (e.g., Ashton et al., 1972) are another methods for stress elicitation. Elements of simulation or, in the case of naturalistic studies, a changing environment, such as the complexity of the road infrastructure or the volume of traffic at a given time, are also sometimes used (e.g., Affanni et al., 2022; Bitkina et al., 2019; Cardone et al., 2020; Dobbins & Fairclough, 2018; Healey & Picard, 2000; Kerautret et al., 2022), thus resulting in a more ecological way of providing stimulation. Therefore, in the field of driver stress research, there is a significant diversity of scientific studies due to variability in the choice of stressors, but also differences in methodological categories, such as exposure time, the complexity of the experimental scenario, or the measurement techniques used. Consequently, it is not easy, if not impossible, to propose a single coherent meta-analysis of stress studies in drivers, and a more reasonable approach seems to be covered by a thorough review of the results for a group of studies with a methodological approach as similar as possible. In such subgroups, it could be achievable to draw high-level conclusions, nevertheless, with the caveat that the studies still may differ in some methodological details.

In addition to the characteristics of the stressor as such, it is also important to verify whether the stressor had affected the driver and, if so, how it was perceived. Accordingly, a fairly popular approach is to try to ascertain this by conducting a subjective evaluation of the level of stress, for which topic-specific scales or questionnaires can be used (e.g., in Dobbins & Fairclough, 2018; Lanatà et al., 2014), although in some studies we encounter a very simple evaluation in the form of a discrete scale, basically asking to rate how stressed one felt or feels at a given moment (e.g., Baek et al., 2009; Balters et al., 2021; Healey & Picard, 2000; Kerautret et al., 2022)—for example on a 1–10 scale. Unfortunately, there are also studies in which the stressor was not subjected to additional evaluation, but only assumed *a priori* that certain traffic situations would be more stressful than others, which, of course, is not necessarily true for every participant (e.g., Cardone et al., 2020; Gruden et al., 2019). We will dwell on the issue of stress level evaluation and specialized questionnaires aimed at vehicle drivers in the next section.

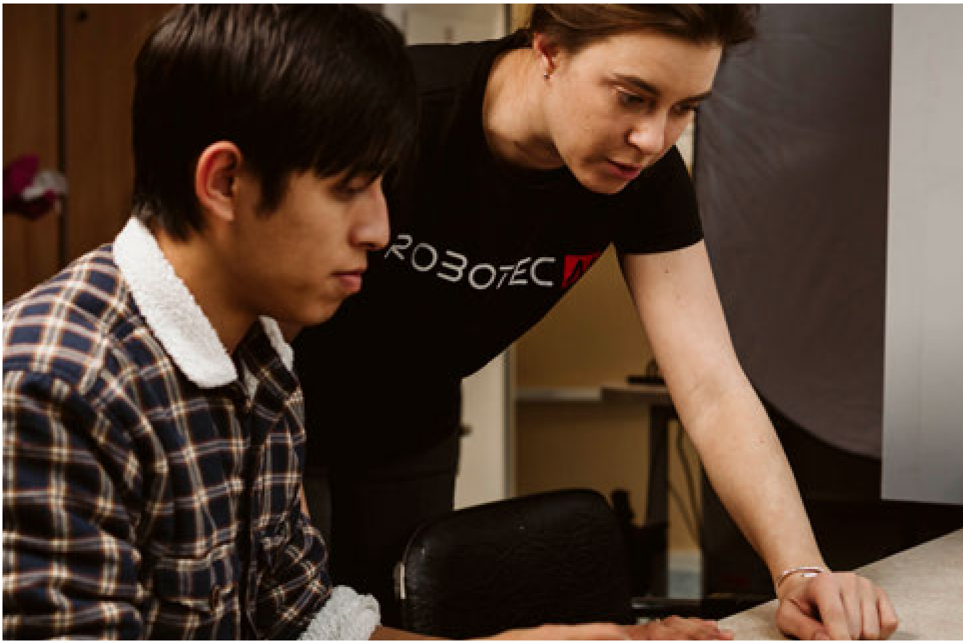
2.1. Driving studies: subjective assessment

Questionnaires and surveys of different kinds are used to measure both the emotional and perceived physical states of drivers, as well as their opinions—e.g., about a newly introduced system or solution in vehicles, or its impact on the quality of driving and the feelings associated with it (Reimer et al., 2016). Since driver stress may seriously impact road safety, research on this particular condition is very common in the automotive field (see e.g., Bustos et al., 2021; Reimer et al., 2016).

Similarly, as for stress research in general, self-report studies are used when examining both stress itself and its impact—in this case, on driving behavior. Here, too, researchers often turn to various sorts of questionnaires, scales, and interviews, which they implement on drivers both pre- and post-task (Chung et al., 2019). With advancements in digital assessment, answering throughout the driving study is also possible if safety conditions are addressed first.

Some standard subjective measures of stress in drivers include rating scales that ask drivers to indicate the level of stress they experience on a numerical, Likert-type scale or a visual analog scale (VAS) that ranges, e.g., from "not at all stressed" to "extremely stressed" (similar examples can be found in e.g., Baek et al., 2009; Balters et al., 2021; Healey & Picard, 2000; Kerautret et al., 2022). Other subjective measures of stress may include open-ended questions or checklists that allow drivers to describe specific stressors or factors that contribute to their stress.

In the area of research on drivers and driving stress, among the commonly used tools, there are both questionnaires tailored specifically to study stress in this very specific group and those dedicated to dealing with stress related to exact traffic events. However, as in non-domain-specific stress research,



some of personality questionnaires can be used here, and they can carry valuable information in large sample studies.

Sometimes, in order to seek answers to specific hypotheses, more than one questionnaire is used—they are either combined into one, or specific questions corresponding to the study's objectives are derived from them. It is worth noting that in studies of stress in drivers, researchers often try to find out what traffic situations increase the stress level, under what conditions it is the highest, what impact it has on drivers, what its consequences are, and how it affects the quality of driving and coping in traffic. In this type of research, sometimes researchers may need more than one tool. For example, Hennessy and Wiesenthal (1999),

in their research on traffic congestion and driver stress, included a set of questionnaires: *Driving Behavior Inventory—General*, *State Driver Stress Inventory*, and *State Driving Behavior Checklist* to measure driver stress and driving behaviors in actual low – and high-congestion conditions. A list of chosen stress questionnaires, the majority of which are tailored specifically for the driver stress use-case, is presented in Table 1 (see Table 1A for the non-specific examples, which may also occur in driver stress research).

Table 1. Selected stress questionnaires frequently used in driving stress-related studies.

Driving Behavior Survey (DBS) Authors: Clapp et al. 2011	
Purpose	Measurement and evaluation of the occurrence of potentially problematic anxiety-related driving behavior.
Details	The objective of the survey is to establish the frequency with which specific actions take place when driving in stressful situations. It is a questionnaire based on the 7-point Likert scale. It has 21 questions divided into 3 subscales.
Examples of use	Clapp et al., 2014
Driving Stress Inventory (DSI) Authors: Gulian et al., 1989	
Purpose	Assessment of drivers' vulnerability to stress reactions while driving.
Details	The DSI consists of 25 items that measure four dimensions of driving stress: emotional tension, cognitive anxiety, physiological arousal, and behavioral avoidance. Participants rate the frequency with which they experience various driving-related stressors on a 5-point Likert scale.
Examples of use	Funke et al., 2007; Mohamad, 2022
Stress Arousal Checklist (SACL) Authors: Mackay et al., 1978	
Purpose	Measurement of self-reported stress and arousal.
Details	The SACL is a short checklist consisting of 30 items that assess three dimensions of stress arousal: somatic arousal, cognitive arousal, and emotional arousal. Participants rate the extent to which they experience various physiological and emotional symptoms when faced with stressors, such as heart palpitations, sweating, racing thoughts, and feeling overwhelmed.
Examples of use	Raggatt & Morrissey, 1997
Dundee Stress State Questionnaire (DSSQ) Authors: Matthews et al., 1999	
Purpose	Assessment of transient states associated with performance-related stress.
Details	The original questionnaire contains 90 items. One version of the DSSQ is administered before a task and another version is administered after a task. Items are designed to sample three main domains: mood, motivation, and cognition. Some of the items used in this questionnaire were taken and repurposed from already existing scales, such as UMACL (Table 2).
Examples of use	Funke et al., 2007; Saxby et al., 2008

Driving Behaviour Inventory (DBI) Authors: Gulian et al., 1989	
Purpose	Driver stress instrument designed to examine the dimensions of stress in drivers and individual driving performances related to stress. The tool recognizes the impact of stressors related to activities other than driving, such as family life or problems at work, on driving quality.
Details	The survey consists of 97 questions that cover various topics such as demographics, details about car usage, frequency and amount of driving, preferred road types, accident, and conviction history, attitudes towards such incidents, information about health and its impact on driving, work, and personal relationships, as well as questions about personal, domestic, and occupational issues. Additionally, the survey explores respondents' moods, emotions, and attitudes towards driving and various traffic situations, including their interactions with other road users. Responses to questions on the frequency of a behavior or categorizing a behavior/opinion were presented on a 4-point scale.
Examples of use	Gulian et al., 1990; Hennessy & Wiesenthal, 1999; Kontogiannis, 2006; Lajunen & Summala, 1995; Matthews & Desmond, 1998

Driving Behaviour Inventory–General (DBI–Gen) Authors: Gulian et al., 1989	
Purpose	Evaluation of a range of driving behaviors that may affect safety on the road.
Details	A variation of the DBI that contains some of the items drawn from the DBI. It is a subset of DBI, therefore, the original work of the creation of both questionnaires is considered the same paper and both questionnaires have the same authors. The DBI-Gen includes 16 items that assess a general inclination, or trait susceptibility, towards experiencing stress while driving.
Examples of use	Hennessy & Wiesenthal, 2001; Li et al., 2004; Wickens & Wiesenthal, 2005

State Driver Stress Questionnaire (SDSQ) Authors: Hennessy & Wiesenthal, 1997	
Purpose	Assessment of the state experience of driver stress. This questionnaire was designed to be administered verbally in actual driving situations.
Details	It is comprised of 11 items similar to those from the DBI-Gen, as well as 10 items from the Stress Arousal Checklist. Note that the abbreviated name of this tool (SDSQ) is also used as an abbreviation for names of different tools e.g., Stress and Depression Screening Questionnaire.
Examples of use	Hennessy & Wiesenthal, 1999; Wickens & Wiesenthal, 2005

Driver Coping Questionnaire (DCQ) Authors: Matthews et al., 1997	
Purpose	Assessment of habitual style of coping with demanding events encountered on the road.
Details	The questionnaire consists of 5 coping scales, each having 7 items. The five coping dimensions include: Confrontive coping, Task-focused, Emotion-focused, Reappraisal, and Avoidance. It uses 5-point Likert scale.
Examples of use	Skippon et al., 2008

Perceived Stress Scale Authors: Cohen et al., 1983	
Purpose	Assessment of how individuals perceive situations in their lives as stressful. It does so by asking questions about their experiences during the previous month and how much they feel their life has been characterized by unpre – dictability, lack of control, and feeling overwhelmed.
Details	There are three versions of the PSS. The original instrument is a 14-item scale (PSS-14) that was developed in English (Cohen et al., 1983), with 7 positive items and 7 negative items rated on a 5-point Likert scale. Five years after the introduction of the PSS-14, it was shortened to 10 items (PSS-10) using factor analysis based on data from 2,387 U.S. residents. A four-item PSS (PSS-4) was also introduced as a brief version for situations requiring a very short scale or telephone interviews (Cohen & Williamson, 1988).
Examples of use	Dobbins & Fairclough, 2018; Ge et al., 2014

Table 1A. Selected questionnaires that are used in driving stress-related studies, however, their main purpose and usefulness lies elsewhere.

Anxious Driving Behavior (ADB) Authors: Clapp et al., 2011	
Purpose	Assessment of the degree to which an individual experiences anxiety and fear while driving.
Details	The scale consists of 19 items that measure anxiety in different driving situations, such as driving on highways, driving in heavy traffic, and driving in bad weather. Each item on the ADB scale is rated on a 5-point scale.
Examples of use	Mohamad, 2022

Driving Vengeance Questionnaire (DVQ) Authors: Wiesenthal et al., 2000	
Purpose	Assessment of driver's reactions to perceived threats and probe which situations would elicit the strongest reactions.
Details	The DVQ is comprised of 15 different scenarios that are presented to drivers. They are then asked to indicate how they would respond if they found themselves in each of the given situations.
Examples of use	Hennessy & Wiesenthal, 2002

Driving Cognitions Questionnaire Authors: Ehlers et al., 1994	
Purpose	Measurement of the severity of driving phobia.
Details	The initial item pool of 49 items in 4 categories (panic-related concerns, accident-related concerns, concerns about other adverse events when driving, and social concerns). The questionnaire uses a 5-point scale.
Examples of use	Barnard & Chapman, 2018

Freight Driving Behavior Questionnaire (F-DBQ) Authors: Useche et al., 2021	
Purpose	Assessment of the driving behaviors and risk factors specific to long-haul freight drivers.
Details	It is a concise version of the DBQ that is modified to suit the driving conditions and common road risk behaviors observed among long-distance freight drivers.
Examples of use	Djordjević, 2022

Driver Skill Inventory Authors: Lajunen & Summala, 1995	
Purpose	Used to measure perceptual-motor and safety skills.
Details	A 29-item self-reported measure of perceptual-motor and safety skills. In the Driver Skill Inventory, drivers are asked to rate how weak or strong they feel they were in each given skill, using a 5-point Likert scale.
Examples of use	Kontogiannis, 2006; Martinussen et al., 2014

Driver Behavior Questionnaire (DBQ) Authors: Reason et al., 1990	
Purpose	Assessment of the self-reported driving behaviors of individuals.
Details	The scale contains 50 items assessing three distinct classes of behaviors: violations, errors, and lapses.
Examples of use	Kontogiannis, 2006; Smorti & Guarnieri, 2016

In addition, there can be questionnaires on specific personality traits or a variety of driving situations. A short list of such emotion- and personality-oriented questionnaires used for driver research is presented in Table 2. Combining such questionnaires with other tools may help reach a broad perspective of the driver's cognitive and emotional states. Of course, it does not mean it is impossible to conduct good research on stress in drivers without using the aforementioned questionnaires.

There is also another branch of interest in stress research—these are examples of cognitive load questionnaires. Researchers use them to evaluate the impact of events or additional tasks and their potential correlation with stress. Additionally, they enable comparison between driving quality and the decision-making process in risky situations. These questionnaires typically ask drivers to rate their mental effort, or the amount of attention required to perform various driving tasks, such as navigating, maintaining speed, and avoiding hazards.

Table 2. Selected examples of emotion-, cognition-, and personality-related tools, which can be found in driving research.

Cognitive Failures Questionnaire Authors: Broadbent et al., 1982	
Purpose	The evaluation of an individual's probability of making a mistake while carrying out a routine task.
Details	The CFQ consists of 25 items; each item is drawn from one of the three categories: perception, memory, or motor function.
Examples of use	Allahyari et al., 2008; Larson & Merritt, 1991
State-Trait Anger Expression Inventory (STAXI) Authors: Spielberger, 1988	
Purpose	Assessment of three components of anger, which are state anger, trait anger, and anger expression, and measurement of the influence of those components on general health.
Details	The scale consists of 44 items comprised of 5 scales and two subscales. With the use of STAXI, participants can report the intensity and the source of experienced anger. In general, STAXI examines how individuals experience and express anger in different situations.
Examples of use	Brandenburg et al., 2017; Dobbins & Fairclough, 2018

UWIST Mood Adjective Checklist (UMACL) Authors: Matthews et al., 1990	
Purpose	Measurement of an individual's mood state by assessing their self-reported feelings and emotions.
Details	The scale consists of 24 adjectives describing mood. The mood factors are categorized as *Hedonic Tone*, *Tense Arousal*, and *Energetic Arousal*, each comprising eight adjectives. *Hedonic Tone* encompasses adjectives that describe pleasantness or unpleasantness, *Tense Arousal* ranges from anxiety to calmness, while *Energetic Arousal* encompasses adjectives describing vigor to fatigue.
Examples of use	Dobbins & Fairclough, 2018; Dorn & Matthews, 1995

Minnesota Multiphasic Personality Inventory (MMPI-2) Authors: Dahlstrom et al., 1989	
Purpose	First major revision of MMPI (Hathaway & McKinley, 1943) with some new subscales introduced. MMPI-2 replaced original MMPI in use as an upgraded version. Both MMPI and MMPI-2 were developed to help professionals with evaluating psychiatric disorders.
Details	The current form of the MMPI-2 consists of 567 items, but there is a less commonly used abbreviated form of the test that consists of the first 370 items of the MMPI-2. This form of the questionnaire is used when time constraints on the test do not allow the use of the full version. Items in the questionnaire are in the form of statements such as "I feel uneasy indoors", "I feel weak all over much of the time", or "I am afraid of losing my mind" which are supposed to help professionals in the diagnosis of certain mental conditions.
Examples of use	Tinella et al., 2021

Impulsiveness-Venturesomeness-Empathy questionnaire (IVE) Authors: Eysenck et al., 1984; Eysenck et al., 1985	
Purpose	IVE was designed to assess three personality traits which are: impulsiveness, venturesomeness, and empathy. The IVE has been used to examine the relationship between impulsivity and illicit drug use.
Details	There are two original papers for this scale, as one is dedicated toward children and the other one toward adults. In the assumptions of the scale, impulsivity is related to psychoticism and is associated with taking risky actions without considering the consequences. Venturesomeness, on the other hand, is associated with extraversion and risk-taking behavior, in which the individual is aware of the risks but engages in the behavior for the thrill.
Examples of use	Owsley et al., 2003; Ismail et al., 2016

NEO Five Factor Inventory (NEO-FFI) Authors: Costa & McCrae, 1989	
Purpose	The NEO-FFI scale was designed to characterize people based on five personality traits: Neuroticism, Extraversion, Agreeableness, Openness to Experience, and Conscientiousness.
Details	The NEO-FFI consists of 60 items (5 scales with 12 items each), which are rated on 5-point scales. The scale contains states ranging from "strongly agree" to "strongly disagree". Might be used in studies on drivers to explore the impact of those personality traits on driving safety.
Examples of use	Guo et al., 2013; Guo et al., 2016

Raven's Standard Progressive Matrices (SPM) Authors: Raven & Court, 1998	
Purpose	The goal behind the creation was the need to be able to test abstract reasoning independent of language. The tool is considered a non-verbal estimate of cognitive skills and fluid intelligence.
Details	The test consists of 60 "questions" of increasingly difficult pattern-matching tasks and has little dependence on language skills. In addition, a set of 36 questions of greater difficulty has been designed. The test containing all 96 questions is called the Raven's Advanced Progressive Matrices Test.
Examples of use	Biernacki & Tarnowski, 2011

Cognitive load questionnaires typically involve a series of questions or items that ask drivers to rate their mental effort or workload on a numerical scale, such as the *NASA-TLX* (Hart & Staveland, 1988) or the *Multiple Resource Questionnaire* (Boles & Adair, 2001). Through these questionnaires, we may also ask drivers to rate the difficulty or complexity of specific driving tasks or scenarios, such as merging onto a busy highway or driving in inclement weather. By assessing cognitive load or mental workload, such tools can provide insights into how different driving conditions and stressors affect drivers' cognitive processing, decision-making, performance, and how drivers allocate their attention and cognitive resources while driving.

Cognitive load questionnaires can also be used to evaluate the effectiveness of interventions aimed at reducing the mental workload and stress in drivers, such as driver training programs or in-vehicle technologies. A list of chosen cognitive load questionnaires used in the studies on drivers is presented in Table 3. In summary, subjective measures of stress in drivers can be useful for understanding how drivers perceive and cope with stress while driving. Some of the measures presented here may to some extent intersect between the stress response *per se,* the assessment of personality or coping strategies, and the assessment of effort or load related to the driving task associated with a stressor. Although subjective measures of stress can be prone to biases and inaccuracies, as drivers may not always be fully aware of their own stress levels or may report their stress levels differently depending on their mood or context, they can provide valuable insights into the specific stressors that drivers experience, such as traffic congestion, road rage, or inclement weather, and help identify effective strategies for managing stress on the road. Hence, they can be extremely valuable tools in both stress and other research areas. However, due to their limitations, it is important to use a combination of subjective and objective measures of stress in drivers to obtain a comprehensive understanding of the impact of stress on driving behavior and safety.

Table 3. Selected examples of questionnaires targeting cognitive load, which can be found in driving research.

NASA-TLX Authors: Hart & Staveland, 1988	
Purpose	Obtaining subjective workload ratings in different experiments.
Details	NASA-TLX dimensions are mental demand, physical demand, temporal demand, effort, and frustration. This questionnaire assesses workload on five 7-point scales. Increments of high, medium, and low estimates for each point result in 21 gradations on the scales.
Examples of use	Anzengruber & Riener, 2012
Driving activity load index (DALI) Authors: Pauzié, 2008	
Purpose	DALI is a revised version of NASA-TLX adapted to driving tasks and studies.
Details	DALI uses the same factors as NASA-TLX: temporal demand, mental demand, effort, performance, physical demand, frustration level. The main difference between DALI and NASA-TLX is the fact that DALI is tailored specifically for assessing the workload of driving tasks while NASA-TLX can be used in various domains. DALI considers factors relevant to driving, including traffic conditions, road complexity, and the need for quick decisions.
Examples of use	Gabaude et al., 2012; Piranveyseh et al., 2022
Multiple Resource Questionnaire (MRQ) Authors: Boles & Adair, 2001	
Purpose	Prediction of interference during multitasking, done by determining the extent to which tasks use the same resources. This is achieved by assessing the overlap in resources used by each task.
Details	The questionnaire consists of a 17-item measure for subjective workload assessment. Based on a 5-point Likert scale.
Examples of use	Boles et al., 2007

2.2. Driving studies: physiological measures

Up to this point we successfully learned that stress is a phenomenon accompanied by *some* physiological changes. Accordingly, one of the most important and widely used approaches in the literature within the field of stress research is the estimation of physiological parameters. In this chapter, we look at the physiological data acquisition techniques that are used in measuring the stress response, and then identify the parameters that have been highlighted so far in stress research. All the recording techniques discussed in this handbook are non-invasive, but they can vary in the complexity of the measurement system, as well as in their susceptibility to artifacts in signal recording phase.

A key aspect to keep in mind to avoid false positives is that these responses may not be specific to the stress phenomenon alone, and therefore may also accompany other physical or cognitive states (e.g., changes in Pupil Diameter due to varying light conditions; increase of arterial blood pressure during mirthful laughter, see Fry & Savin, 1988), and the level and detectability of the response itself will be susceptible to intra- and interindividual factors.

Table 4. A summary of selected non-invasive techniques utilized in stress detection research in driving context.

Electrocardiography (ECG)	
Description	A non-invasive measurement of the heart's electrical activity, which is causally related to the cardiac cycle properties of heart muscles contractions and relaxations. The technique is based upon application of electrodes on participant's chest or limbs, and continuous recording. ECG data is usually analyzed in terms of time-based, frequency-based, or nonlinear characteristics.
Examples of use	Studies: Ashton et al., 1972; Baek et al., 2009; Balters et al., 2021; Cardone et al., 2020; Chui et al., 2020; Dobbins & Fairclough, 2018; Eilebrecht et al., 2012; Gruden et al., 2019; Healey et al., 1999; Healey & Picard, 2000; Huang et al., 2020; Kerautret et al., 2022; Lanatà et al., 2014. Discussed in reviews: Burlacu et al., 2021; Giannakakis et al., 2019.

Electroencephalography (EEG)	
Description	A non-invasive technique for measurement of brain electrical activity. Electrical potential of neural tissue in cortex changes in time both during rest and during different cognitive processes, and such changes can be monitored via EEG. During EEG recording, participant wears a cap with electrodes, which collect the signal from the head surface. There are various techniques for EEG analysis, with frequency bands-based measures, time-locked analysis, or time-frequency representations as the most common.
Examples of use	Studies: Affanni et al., 2022. Discussed in reviews: Giannakakis et al., 2019.
Electromyography (EMG)	
Description	In non-clinical settings, EMG is typically recorded from the skin surface, right above the muscles of which activity we are interested to investigate. During muscle contraction, electrical properties of muscle cells change for a brief period, and the changes can be monitored with a nonintrusive technique, such as surface EMG. With EMG, one can record and track the periods of muscle contractions and relaxations. EMG is usually being analyzed in terms of time- and frequency-based frameworks.
Examples of use	Studies: Healey et al., 1999; Healey & Picard, 2000. Discussed in reviews: Giannakakis et al., 2019.
Photoplethysmography (PPG)	
Description	PPG is a non-invasive technique enabling real-time monitoring of changes in blood flow and volume in vessels. Usually PPG is recorded from participant's fingertips or ears, with a setup consisting of LEDs and photodetectors. The time series can be analyzed in terms of slow and rapid changes.
Examples of use	Studies: Baek et al., 2009; Dobbins & Fairclough, 2018; Gruden et al., 2019. Discussed in reviews: Giannakakis et al., 2019.
Eye-tracking	
Description	Eye-tracking is based on estimation of gaze vectors on the basis of video images of participant's face, with prior identification of eyes' features, such as pupils, irises, or Purkinje reflections. Eye-tracking is measured nonintrusively, and for modern setups there is no need to mount the system on participant's head or to restrict head movements. Typically, gaze vectors are estimated and analysis of gaze in certain regions of interest (ROI) is performed. Usually, information about changes of Pupil Diameter can also be derived from the output data.
Examples of use	Studies: Mou et al., 2021; Pedrotti et al., 2014. Discussed in reviews: Giannakakis et al., 2019; Kerautret et al., 2021.

Respiration	
Description	In research settings, respiration is based on nonintrusive measurement of chest expansion during breathing. Usually it is being measured with conductive sensors reacting to expansion of the band around the chest. Respiration can be analyzed in terms of e.g., frequency or intensity of breaths.
Examples of use	Studies: Ashton et al., 1972; Baek et al., 2009; Healey et al., 1999; Healey & Picard, 2000; Lanatà et al., 2014.
Electrodermal Activity (EDA)	
Description	A non-invasive technique to estimate skin conductivity changes, which may occur e.g., during stress. Conductivity is measured with electrodes, which can be placed on fingertips or foot, where many sweat glands are located. EDA can be analyzed in terms of tonic (long-term) and phasic (short-term) changes occurring in the signal.
Examples of use	Studies: Baek et al., 2009; Bitkina et al., 2019; Dogan et al., 2019; El Haouij et al., 2015; Healey et al., 1999; Healey & Picard, 2000; Lanatà et al., 2014. Discussed in reviews: Giannakakis et al., 2019.
Thermal imaging	
Description	Thermal imaging is a technique involving measurement of emitted infrared light. Thermal imaging camera can record changes in skin temperature from a predefined distance, and with specified resolution. Thermal imaging data can be investigated in terms of ROI analysis.
Examples of use	Studies: Anzengruber & Riener, 2012; Cardone et al., 2020. Discussed in reviews: Giannakakis et al., 2019.

In the field of stress research, two prominent recording techniques are electrodermal activity (EDA) and electrocardiography (ECG). Many research articles (see Table 4) have evaluated EDA and ECG data, looking for sources of so-called "ground truth" or more robust referential measures, and we will start the overview with these signals. As keeping track of all the abbreviations and parameters discussed in this Chapter may be to some extent challenging, Tables 4–7 are here to help you better understand the most relevant techniques and associated variables.

Because of the convenience and non-invasiveness, ECG is widely used in research with vehicle drivers, both in simulated conditions and naturalistic studies. In one of the oldest studies from the field of stress physiology in drivers (Ashton et al., 1972), authors set out to evaluate the effect of the stressor occurrence on physiological parameters while operating a control panel, thus eliciting cognitive-manual workload, and simultaneously participating in a simulation. Compared to today's high-fidelity driving simulators, this is a simplified solution, yet it provided very interesting results, opening the horizon of new research possibilities in that period. The researchers measured heart activity and blood pressure in parallel and observed a positive relationship between heart rate and an increase in task complexity. However, for changes in blood flow and blood pressure, they observed no significant differences between levels of difficulty set in the task.

In a very different setting, a study by Baek et al. (2009) verified the effects of physical and mental stressors on physiological data during naturalistic driving study. The researchers were interested in HRV parameters, Pulse Arrival Time (PAT) analysis and RR intervals, which they compared to baseline values, but without further investigation of group statistics. A relevant take-home message from the study is that significant variability in physiological responses was observed both inter- and intraindividually.

Balters et al. (2021) conducted a study using wearable technology—the Zephyr BioHarness 3.0™—to examine cardiac function while driving. The participants were asked to perform the *Trier Social Stress Test* as a stress-eliciting stimulus or to relax before the start of the drive. Mean heart rate (HR) and mean RMSSD were selected from HRV measures to be analyzed further. They noted a higher HR in the stress stimulus condition than in the relaxation condition, as well as a statistically insignificant trend of reduced RMSSD in the stress condition compared to the relaxation condition. In the study by Cardone et al. (2020), the stressors during simulated driving were emergency situations and traffic. RR intervals were analyzed, as well as the so-called *Baevsky's Stress Index* (SI, see Cardone et al., 2020), which was used as a referential measure. The participants were not asked about their subjective feelings with respect to perceived stress, so it is difficult to demonstrate that the SI was the correct ground truth measure.

Table 5. A summary of the most popular parameters derived from ECG in the stress-related driving studies.

Inter-beat interval (IBI) / RR interval	
Description	A period between two subsequent R peaks in ECG [ms]. Typically, normal RR (known as NN) intervals are taken into consideration, after exclusion of abnormal peaks (Shaffer & Ginsberg, 2017).
Overview results	Dobbins & Fairclough, 2018: mean IBI among top features in stress discrimination. Huang et al., 2020: mean IBI significantly different among different stress levels.
Examples of use	Baek et al., 2009; Dobbins & Fairclough, 2018; Eilebrecht et al., 2012; Huang et al., 2020; Lanatà et al., 2014.
Heart rate (HR)	
Description	Number of heart beats per minute [bpm].
Overview results	Ashton et al., 1972: HR increased with stress/task difficulty. Balters et al., 2021: HR increased in task after stressful stimulation in contrast to task after relaxation. Dobbins & Fairclough, 2018: HR among top features in stress discrimination. Eilebrecht et al., 2012: positive correlation of HR and TLX-based subjective workload. Healey & Picard, 2000: HR feature with 52.6% performance in 4-class stress classification.
Examples of use	Baek et al., 2009; Ashton et al., 1972; Balters et al., 2021; Dobbins & Fairclough, 2018; Eilebrecht et al., 2012; Healey & Picard, 2000.

SDNN	
Description	Standard deviation of the NN intervals[11] [ms].
Overview results	Huang et al., 2020: SDNN significantly different among different stress levels.
Examples of use	Baek et al., 2009; Dobbins & Fairclough, 2018; Eilebrecht et al., 2012; Gruden et al., 2019; Huang et al., 2020; Lanatà et al., 2014.
RMSSD	
Description	A parameter calculated as a root mean square of differences between subsequent NN intervals (Kleiger et al., 2005; Shaffer & Ginsberg, 2017) [ms].
Overview results	Balters et al., 2021: RMSSD decreased in task after stressful stimulation in contrast to task after relaxation (not statistically significant). Huang et al., 2020: RMSSD significantly different among different stress levels.
Examples of use	Baek et al., 2009; Balters et al., 2021; Dobbins & Fairclough, 2018; Gruden et al., 2019; Huang et al., 2020; Lanatà et al., 2019.
pNN50	
Description	Percentage of differences between the adjacent NN intervals greater than 50 ms [%].
Examples of use	Baek et al., 2009; Gruden et al., 2019; Lanatà et al., 2014.
LF power	
Description	Power of the HRV low-frequency band, usually defined as 0.04–0.15 Hz (see Kleiger et al., 2005; Shaffer & Ginsberg, 2017) [ms^2].
Overview results	Eilebrecht et al., 2012: negative correlation of percentage of LF power in a parameterized FFT spectrum and TLX-based subjective workload.
Examples of use	Baek et al., 2009; Dobbins & Fairclough, 2018; Eilebrecht et al., 2012; Lanatà et al., 2014.
HF power	
Description	Power of the HRV high-frequency band, usually defined as 0.15–0.4 Hz (see Kleiger et al., 2005; Shaffer & Ginsberg, 2017) [ms^2].
Examples of use	Baek et al., 2009; Dobbins & Fairclough, 2018; Eilebrecht et al., 2012; Lanatà et al., 2014.

11 The NN intervals are the RR intervals calculated after the exclusion of abnormal peaks in the ECG signal (Shaffer & Ginsberg, 2017).

LF/HF ratio	
Description	A ratio of low-frequency and high-frequency bands power in HRV. After calculating the LF and HF power, the ratio can be established.
Overview results	Healey and Picard, 2000: Autonomic Balance feature (similar to LF/HF) with 52.5% performance in 4-class stress classification.
Examples of use	Baek et al., 2009; Dobbins & Fairclough, 2018; Healey & Picard, 2000.
VLF power	
Description	Power of the HRV very low-frequency band, usually defined as 0.0033–0.04 Hz (see Kleiger et al., 2005; Shaffer & Ginsberg, 2017) [ms^2].
Examples of use	Dobbins & Fairclough, 2018; Lanatà et al., 2014.
Total power	
Description	Total power estimated for the signal (see Kleiger et al., 2005; Shaffer & Ginsberg, 2017) [ms^2].
Examples of use	Dobbins & Fairclough, 2018.

Dobbins & Fairclough (2018) recorded ECG and PPG activity in a group of participants during daily drives to and from work. The study had no additional stressors other than situations naturally encountered on the road. Multiple parameters of ECG and PPG signals were analyzed, and the mean and standard deviation of Pulse Transit Time, heart rate and mean IBI were identified as the best features in stress discrimination. However, a limitation of this study is that the researchers' subjective inference about the possible occurrence of a stressful situation was based on contextual photographs, supported only with pre/post mood assessment questionnaires.

Eilebrecht et al. (2012) tested the hypothesis that driving in the city would lead to greater workload than driving on the highway. Participants also answered the *NASA-TLX* scale during the study. The researchers analyzed multiple ECG indices and identified mean HR as correlating with subjective workload ratings and percentage of LF power from parameterized FFT spectrum as negatively correlating with subjective workload ratings. However, an important limitation of the present study is the small sample size, and the selection of HRV parameters was left unjustified and analyzed without explicitly stated hypotheses first.

In a study by Healey & Picard (2000), the ECG signal collected during naturalistic driving study (NDS) conducted in areas of varying complexity (e.g., highways, toll booths) was analyzed, and the participants' subjective ratings of perceived stress levels were collected. Two variables were analyzed—heart rate and *autonomic balance*, which is a similar parameter to the LF/HF ratio. The study design covering four stress level categories evaluated the indices in stress classification, indicating 52.6% for heart rate and 52.5% for *autonomic balance* performance in stress classification. In a study by Huang et al. (2020), ECG was measured under simulated conditions, with mental arithmetic being the stress-inducing task. It was noted that all three analyzed indices (mean IBI, SDNN, RMSSD) significantly differed between separated stress levels. However, in the study description, there is a lack of information about the sample size. Therefore, discussing these results' potential relevance and generalizability is challenging.

In a study by Kerautret et al. (2022), participants drove in a simulator in an experimental condition with three levels of threat predictability: unfamiliar and unpredictable, unfamiliar and predictable, familiar and predictable. They observed an increase in HR after a traffic event in both predicTable conditions for all participants and unpredicTable conditions for some of them. In the study by Lanatà et al. (2014), participants were enrolled to drive in a simulated environment. In addition to the control condition, they were presented with two levels of stress driving conditions (skids and skids + mental arithmetic). The ECG signal was analyzed with multiple indices (including mean NN, SDNN, RMSSD, pNN50) and indicated that statistically significant differences were found for all characteristics for at least one condition.

Only a few studies discuss the use of EMG for stress testing in the context of driving. One study by Healey & Picard (2000) measured EMG from the

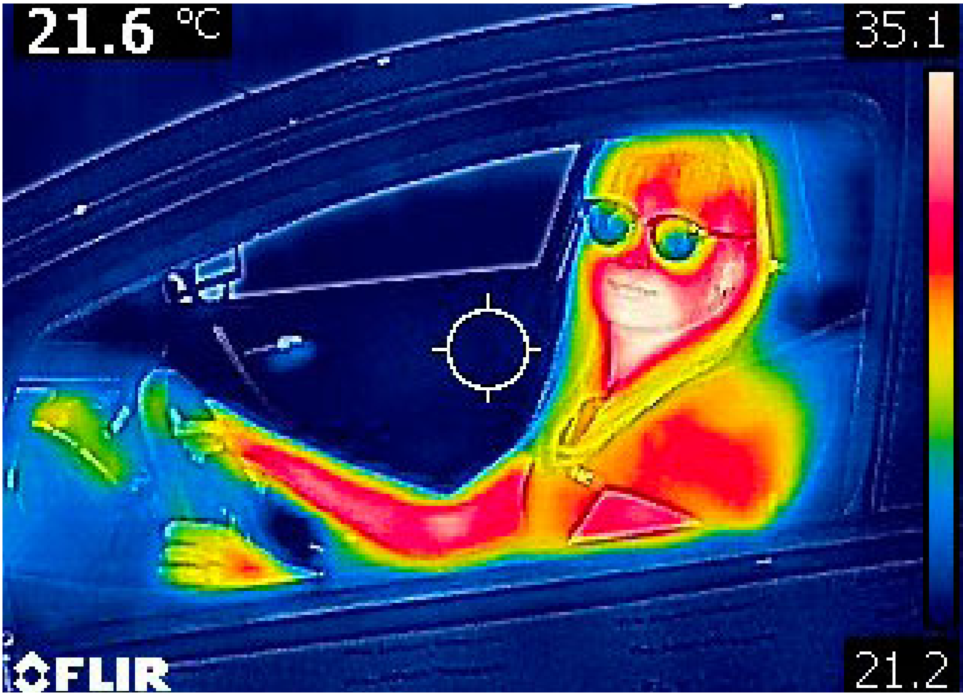
participant's back (the precise location was not specified in the report), and drivers participated in an NDS procedure with naturally occurring stressors along the way. The mean and variance of the EMG signal was determined for each one-minute segment, and the performance of each feature was then determined for 4-class classification, with mean EMG resulting in 58.3% and variance of EMG with 53.5% of performance.

Photoplethysmography is another technique besides ECG that allows us to observe specific changes occurring in the cardiovascular system. Because of its ease of use, it is popular in psychophysiological research, but in the context of stress-related driver studies, it appears to be used less frequently than ECG. Baek et al. (2009) study used a PPG sensor on the steering wheel of a vehicle, and the study itself looked at the effects of mental and physical stressors on the driver. The variable of interest to the researchers was PAT. At the same time, no group results were presented, making it difficult to systematically assess the usefulness of this feature in monitoring stress. In contrast, Dobbins and Fairclough (2018) study used a PPG sensor attached to the driver's ear, and the participants drove in an NDS procedure in which only stressors naturally occurring in traffic appeared. The researchers were interested in the Pulse Transit Time (PTT) variable, and then identified mean PTT and standard deviation of PTT among the top features in stress discrimination.

Table 6. A summary of the most popular parameters derived from PPG in the stress-related driving studies.

Pulse Arrival Time (PAT)	
Description	Time interval between ECG R peak and the characteristic point of the pulse wave of the peripheral artery (Baek et al., 2009).
Overview results	Baek et al., 2009: they investigated mean PAT, but the data were presented individually, with no group analysis.
Examples of use	Baek et al., 2009
Pulse Transit Time (PTT)	
Description	"Time interval between an R peak in the ECG and the subsequent S peak of the PPG [33]." (Dobbins & Fairclough, 2018, p. 6).
Overview results	Dobbins & Fairclough, 2018: mean PTT and SD PTT among the top features in stress discrimination.
Examples of use	Dobbins & Fairclough, 2018

There are also studies in which the temperature measurement of drivers under stress exposure is of interest. In a dual-task-paradigm simulator study by Anzengruber and Riener (2012), the authors used thermal imaging to assess the relationship between temperature changes and cognitive load increases. Among the areas of interest studied were the temple and inner corners of the



eyes. Some associations were noted between temperature changes in these areas and workload, but without clear conclusions. Furthermore, we need to keep in mind that cognitive workload and stress are not the same phenomena, and their co-occurrence would need to be further investigated before transferring or interpreting such outcomes from cognitive load studies as potential stress indicators. The study by Cardone et al. (2020) used thermal imaging in a simulator study considering stressors, and the areas of interest were the nose tip, nostrils, and glabella. A group of parameters from the thermal signal were determined for the indicated areas. Then, the relationships between their values and the *Stress Index*, which was also obtained in the study, were analyzed. The authors

then indicated which calculated parameters were most significant in predicting the *Stress Index*. For the ROI nose tip, it was kurtosis and skewness. For nostrils, the standard deviation, and for glabella, it was the 90^{th} percentile. However, a limitation of the study is the adoption of the *Stress Index* as a reference measure without the subjective assessment of stress levels from the participants' perspective.

In a study by Yamakoshi et al. (2008), the authors also investigated skin temperature measurement to search for driver stress correlates. The study was conducted with 25 participants driving a monotonous simulated scenario. Next to the temperature measurements, blood pressure measures were taken into consideration, and subjective stress was reported for each 10-minute period, with a 1–9 Likert-type scale. The prior assumption in the study was that monotonous driving would presumably increase stress levels, as drivers need to keep focused throughout the drive. The authors collected the data from preselected ROIs, such as finger, nose, cheek, jaw, or forehead, divided to represent peripheral (finger, nose) and truncal (cheek, jaw, forehead) parts. They observed a slight but significant decrease in nose and fingertip temperatures. However, temperatures of the truncal parts did not rise, and they claim this is a good sign, as these may be used to represent general disturbances of temperature because their outcome was not different than the one for baseline. However, stress caused by monotony covers one from a plethora of different factors, and, for example, forehead temperature measurement in a different setting and stressor exposition could potentially be a good candidate for stress assessment (e.g., in a setup inducing cognitive load forehead temperature may rise, see e.g., Abdelrahman et al., 2017).

The next physiological source of information we discuss is the use of respiration-related techniques. It is a measurement of some interest in studies of stress among drivers, usually associated with the application of sensors placed on the chest to monitor the degree of chest expansion.

In the Ashton et al. (1972) study, which was mentioned earlier with respect to ECG, respiratory measures were investigated in addition to other physiological measurements. The parameter further analyzed was mean respiration rate, yet no statistically significant differences were found for this feature between the baseline task and tasks likely to cause higher levels of stress in the participants. In a study by Baek et al. (2009), also discussed earlier, the sensor for measuring respiration was attached to the seat belt. They reported the outcomes for mean respiratory rate by comparing mean values from baseline and experimental conditions for each participant separately. There was a large variability both in terms of inter- and intraindividual aspects, yet some of the individual results were significant. In the Healey and Picard (2000) study, a respiratory signal was collected from the chest. The researchers determined two parameters: mean and variance of respiratory signals. Then, they determined performance in 4-category stress classification for these features with a score of 62.2% for mean and 50.2% for variance of the signal. In the simulator study by Lanatà et al. (2014), the authors determined the following parameters: respiration rate and the mean and standard deviation of the first and second derivatives of the signal. Among the investigated features, they found interesting observations in respiration rate, signal power in 0.1–0.2 Hz band, and skewness-related features.

Table 7. Summary of overview results for Respiratory Rate.

Respiratory Rate	
Description	Number of breaths per minute [bpm].
Overview results	Ashton et al., 1972: no significant differences between the levels inducing more stress and the basic levels. Baek et al., 2009: for mean respiratory rate comparison of baseline vs experiment levels was conducted for each participant separately. Some results were significant. Large inter- and intraindividual variability. Healey and Picard, 2000: mean respiration with 62.2% performance in 4-category stress classification; variance of respiration with 50.2% performance. Lanatà et al., 2014: they investigated respiratory rate, mean and SD of the 1st and 2nd derivatives. They found some significance for respiratory rate, 0.1–0.2 Hz band power, and the skewness values.
Examples of use	Ashton et al., 1972; Baek et al., 2009; Healey & Picard, 2000; Lanatà et al., 2014.

One of the most popular psychophysiological measurement techniques in driver stress research is undoubtedly Electrodermal Activity/Galvanic Skin Response. It appears in a variety of configurations, and due to the relative convenience of the data collection procedure, as well as the clearer psychophysiological interpretation, it is of great interest to the research community, as well as for possible industrial applications.

For EDA/GSR sensors, there are various approaches in their application, such as on the steering wheel (Baek et al., 2009), but usually the test is conducted with electrodes placed on the fingers (e.g., Dogan et al., 2019; Lanatà et al., 2014). A study by Bitkina et al. (2019) looked at a number of parameters calculated from the GSR signal, including features of amplitude and duration of activity. From a subset of the parameters analyzed, some were identified as significant predictors in stress assessment, including, among others, the number of occurrences, duration, and amplitude of responses. However, by virtue of the fact that this was a longitudinal, but single case study, and the assumptions regarding stress induction were significantly simplified, the identified parameters would require confirmation in subsequent experimental studies on stress in a larger group of drivers. The Healey and Picard (2000) study used foot and hand GSR recording. The GSR signal's mean and variance were calculated for one-minute signal segments, and also the values associated with the so-called *GSR startle response* were determined. The authors indicated how the individual traits fared in the 4-class classification of stress levels, and of the GSR signal traits analyzed, it was mean GSR that showed the most promise, with 62.0% of performance.

In the study by Lanatà et al. (2014), which we also described earlier, the GSR signal was analyzed while considering the tonic and phasic components of the signal. The detailed contribution of each calculated feature was not described. Still, the authors reported that the parameters determined for both components allowed them to distinguish between three levels of the procedure complexity, i.e., baseline condition, condition with physical stimuli (lateral skids), condition with physical and cognitive stimuli (lateral skids + mental arithmetic).

As this is just an overview and not a systematic review or meta-analysis, we cannot provide you with a summary of the section, having clear-cut recommendations or *ground truths* with regard to physiological data acquisition

and analysis in driver stress research. However, we hope that we successfully guided you through some of the complexities associated with including psychophysiological measures, with possibly clear explanations of the specific signals. It is still an open field for further research explorations. Yet, after reading many experimental works in this domain, we see that emphasis should be put on larger study samples and a more thorough description of hypotheses standing behind the choice of certain tools and parameters. However, in the next section, we will briefly target an aspect that was not specifically mentioned before, namely the application of physical and, to some extent, behavioral measurements in studying stress among drivers.

2.3. Selected examples of physical and behavioral measures in driving context

Next to the measurement of internal bodily signals, and those externally "visible"—such as eye-tracking and thermal monitoring, which may indicate changes in the activity of specific organs or systems—there exist measures derived from biomechanical data, and also those which can be obtained from the vehicle information. In the work of J. Lee et al. (2021), authors proposed the existence of three types of data, which can be useful for stress detection in drivers: vehicle data, facial behavior data, and physiological data. Based on our work and experience with a variety of data collected in driving studies, we decided to tentatively disentangle data sources into subjective assessment, physiological input, and a "mechano-behavioral" category to cover all the aspects in a satisfying manner. The "mechano-behavioral" category could be further split into bodily movements and forces, facial behavior, and driver performance data/vehicle feedback.

Among ways that fall into the first from the listed categories, there exist sensors measuring the grip forces or posture, and also actigraphs recording acceleration of different body parts, e.g., limbs. These are not very common in studying stress specifically, yet we found two examples of studies using such tools. In one of them (Dogan et al., 2019), force-sensing resistors (FSR) were used to measure grip forces put on the steering wheel, next to the physiological measurement of the EDA signal. Authors discussed the FSR-derived results in light of gender differences during stress, yet with no general conclusions on usability of such metrics in a general application for stress level determination. The second research work by D. S. Lee et al. (2016) included an Inertial Measurement Unit (IMU) sensor on top of the participants' hands, mounted in a wearable glove, to estimate the steering wheel motion and further derive some features for stress estimation. As force sensors may still be viable, posture and acceleration measuring devices, which were mentioned above, seem to be of limited usability in terms of driver stress assessment because both could be affected due to driver's specific mobility in a vehicle cabin. No clear associations were found regarding postural or acceleration-based stress correlates, which would be valid specifically for driver's posture and mobility.

Another category of data can be associated with the derivation of features from the facial behavior of a driver. One example of such a study can be found in Gao et al. (2014), in which authors assumed that two of the basic emotions, anger and disgust, will be stress indicators. A stress state will be detected if those emotions are to be recognized from a driver's face in a specified time interval. The authors trained a Machine Learning model on the basis of existing databases of emotional facial expressions and evaluated its performance with their own data, including office and vehicle scenarios. Unfortunately, the authors did not provide detailed information regarding detected landmarks or features, which heavily contributed to successful detection. One limiting factor of this study is that the assumption that those two emotions convey stress may not be certainly correct, and the authors did not create a setup enabling them to naturalistically challenge the assumption, as participants were only asked to mimic certain emotions. Yet, the research work itself may be a useful resource from the perspective of camera-based system development.

When it comes to mechanical properties, not only the data collected directly from the driver may be valuable. For example, in driving simulator studies, we can collect information about the properties of steering wheel behavior. Alternatively, having the sensors mounted on steering wheel, we can collect the data about the grip force directly from it. Aside from that, we can collect behavioral measures in terms of reaction times, speed, or distance to the lane center, which may be valuable in terms of estimation of the influence of stress on driver performance. Such an example can be found in the work of Balters et al. (2021), in which authors propose utilization of steering wheel angle data for driver stress detection. Yet, one limitation that should be highlighted is that the proposed model's performance was limited to curvy roads. Another example can be found in one of the works cited earlier, Lanatà et al. (2014), in which changes in selected mechanical parameters (steering wheel angle correction, vehicle velocity difference) obtained from some predefined periods of the simulated drive were found to be significantly correlated with stress increase.

The abovementioned examples of physical and behavioral measurements were provided here not as a result of comprehensive, systematic review on that matter but rather to present a complete scope of tools we already have in our toolbox. However, as it goes for all the methods we discuss in the handbook, hypotheses, not the methods, drive the study. Therefore, a careful selection of our "screwdrivers" and "hammers" is still in our hands (pun intented, again).

2.4. Driving research meets multimodal assessment and Machine Learning-based approaches

A primary observation from the research discussed up to this point is that stress studies in drivers typically involve many variables, which could be of potentially vital interest when searching for answers to the questions with which we started the handbook. *What is stress? How can we measure the phenomenon? How can we tackle the issue of stress among drivers?* Well, right now, we already know that plenty of research tools are available, yet also that the analysis and interpretation of the outcomes may be seriously complex. As there is no single "ground truth" measure to which we can refer as an excellent, always-reliable marker of stress, we still need to keep to the path of continuous searching whether—for example—any combinations of methods will suffice in terms of stress detection. It is the primary reason behind a multimodal approach, which allows us to consider in parallel PSYCHOLOGICAL VARIABLES, which can be derived from questionnaires and subjective assessment, PHYSIOLOGICAL VARIABLES recorded in the study course, and sometimes BIOMECHANICAL VARIABLES or BEHAVIORAL and DRIVING PERFORMANCE features. The essence of this approach lies in an attempt to capture the multidimensionality of the correlates and influences accompanying the occurrence of stress. So far, we have discussed the aspects of how stress impinges on the occurrence of a physiological response, how the occurrence of stress may affect a person's cognitive functioning, as well as their subjective assessment, or their behavior. In the case of driver studies, such opportunity arises, as in a properly designed experimental procedure, observations can be made in parallel for each of these areas. Thus, a multimodal approach entails that the data accompanying the different levels of correlates of the stress phenomenon can be analyzed and interpreted in light of the hypotheses and experimental conditions created. However,

multimodality means "increased complexity" as well. A multitude of data streams originating from different functional levels—behavioral, physiological, or introspective/subjective—with no clear-cut nor obvious causal relationships, seem to be a challenging task to solve.

Our Team efforts are dedicated to much more than the concept of driver stress itself, and as we had been researching the topic of yet another widely discussed phenomenon—driver sleepiness—we coined the term "Driver State Bermuda

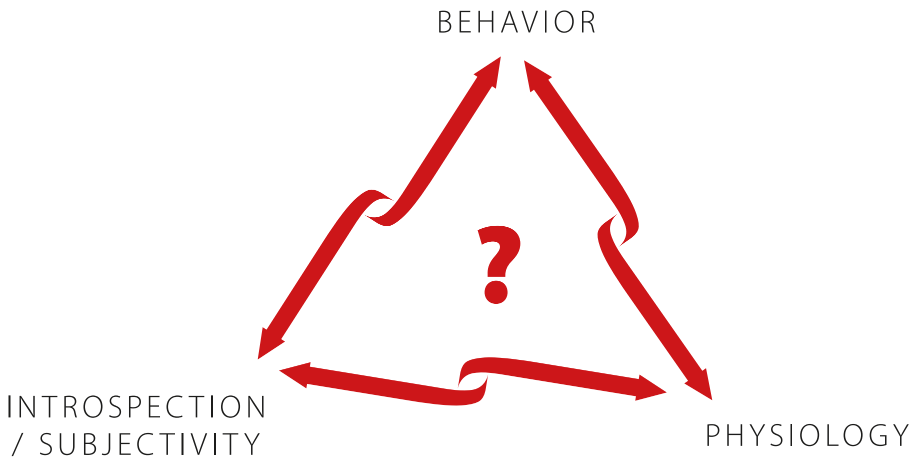


Figure 3.

A pictorial representation of the term "Driver State Bermuda Triangle".

Triangle"[12], which was supposed to picture the complexity of driver drowsiness assessment with vertices representing the three major levels of scientific investigation, and the center of the triangle, with a large question mark, representing "a (still) big unknown" of causal interrelations between those levels (see Figure 3). After deep diving into different aspects of driver-oriented human factors research, we started to observe that it may follow a similar picture for other cognitive states investigated.

12 With subsequent investigations, we observed that the term "Bermuda triangle" in the area of cognitive science, neuroscience or psychology seems to be quite popular in different contexts (e.g., Friedman, 2021; Pettersson et al., 2018), yet to our best knowledge we used it for the first time in the given context within the field of driver state research.

The Triangle is comprised of three vertices—*behavior, introspection* and *physiology,* each representing a different explanatory level of a certain phenomenon—it could be drowsiness or a stress state, being the core topic of the book. The state of causal relationships between these factors seems to be currently unknown, as no strict or perfectly reliable chains of causality seem to be established. For example, the occurrence of physiological stress correlates does not need to be necessarily followed by any facial expression nor by the participant's conscious perception of stress declared through subjective assessment. Will the "Driver State Bermuda Triangle" be solved over time? It may be so, but as for the current state of knowledge, causality between each of the vertices is not established. Therefore, incongruences between the outcomes from different levels grasping the phenomenon may occur. Let's imagine us as study experimenters observing what follows: a physiological variable, e.g., low-frequency HRV; subjective assessment—e.g., an introspective report from the driver that they feel stressed "a little" at some time point, and a behavioral metrics, e.g., more frequent gaze vector shifts during a given period. Is it enough to claim that the driver is under stress? Are there any better or worse metrics out of those? No such answers will be given here, but this is where we can start some really deep diving investigations. Things can get seriously complex if we aim at monitoring the time courses of physiological, behavioral, and introspective responses, as without prior assumptions on causal factors intermediating in expressions on each level, we may be stuck in a "which was first—chicken or egg?" position. Does it mean that our efforts at reaching a proper assessment of different cognitive states are pointless? Certainly not, as only through the means of continued research we can push the boundaries further, understand the phenomenon better, and proceed with technological solutions, step-by-step covering more complex stress assessment—starting with obvious, clear-cut cases in which the three vertices equally voice for the occurence of stress events, and then, with increasing knowledge, aiming at grasping more nuanced cases from the Triangle. Therefore, no worries, roll up your sleeves. We still have a lot of fascinating research work to be done in this domain. Now, let's take a look at some of the concepts and tools that might be helpful in this research journey.

There are a variety of approaches used to analyze multimodal data. The most basic and widespread is the one in which the data from each modality is analyzed in a standard way for it—e.g., for the ECG signal, it is the appropriate filtering, followed by the detection of R peaks and the calculation of characteristic HRV measures; for thermal data it could be preprocessing, ROI selection and subsequent analysis of any changes between conditions. In the next steps, the obtained data features can be subjected to correlation analysis to trace possible co-variation in the data, as the search for similarities or trends may facilitate the analysis through data dimensionality reduction or simplify the study setup in the following iterations if there is unnecessary redundancy present. Having a referential measure (e.g., the subjective assessment of stress levels by the participant or experimental conditions that evoke different stress levels—e.g., low, medium, high), observations can be made as to which variable (physiological, mechanical, driving parameters, etc.) possibly reproducibly "follows" the referential measure and thus could be a measure supporting detection or prediction of the occurrence of stress response under driving conditions.

As some interactions can be more complex or nonlinear, Artificial Intelligence (AI) algorithms are of particular interest to researchers, enabling the construction of classifiers and predictive models that are based on the identification of very complex relationships between variables. Right now, we will take a look at some of the exemplary research works aimed at driver stress estimation, which utilized AI tools in their multimodal approach. We will not dwell on the aspects of implementation, explanation, or selection of certain Machine Learning models in the reviewed articles, as this is not the book's core. Still, we will provide some additional comments where needed.

The section is finished with a brief summary of arguments in favor, but also risks of inconsiderate application of ML tools in the domain.

In the research work of Zontone et al. (2021), the authors investigated the application of ML techniques in binary stress classification upon the Skin Potential Response (SPR) signal. For that purpose, they conducted a driving simulator study with two separate conditions: the one presumed to be more stressful, with other road users presenting unpredicTable behavior, and the second, with no such potential stressors present. Authors presented four classifiers: SVM, RF, DT, k-NN[13], which were trained in their earlier study, and having several features calculated from the SPR signal in the current study, they approached binary classification of states as "stress" vs. "no stress". However, only the percentage of performance for single participants and the mean were presented for the percentage of intervals labeled as "stress". A limiting factor of the study lies in a lack of additional referential measures, assuring that the participants were indeed facing stress. Another example, the paper mentioned in the previous section, from J. Lee et al. (2021), explores the possibility of applying Convolutional Neural Networks (CNNs) in the domain of stress detection. To do so, the authors used short-time data from hand and foot GSRs and heart rate measurements. From the time domain, the authors shifted the signal data to a phase space, creating nonlinear data representations, which were further provided as the input data to the CNNs, which aimed to produce a binary output on stress state detection, distinguishing between high and low stress levels. The models were evaluated on a publicly available stress dataset, with assumptions that *rest = low stress*, *highway = medium stress*, and *city = high stress*. For brief, 30-second signals (from a preselected subset of driving data, not from the whole drive), authors reported *95.67%* accuracy; for 10-second signals, it was *92.33%*. Their work also highlighted the superiority of multimodal over unimodal approach, as the

13 See *List of abbreviations* for the abbreviations describing Machine Learning models exemplified in this chapter.

integration of signals resulted in better accuracies of their models. However, assumptions of induced driver stress level, which were not backed by drivers' feedback, and predefined chunks of signals that were under consideration may presumably limit the model's generalizability. Yet, the approach with the application of signal phasic representations seems to be interesting. An example of using another biosignal data as the state evaluation basis can be found in the research work by D. S. Lee et al. (2016). The authors proposed an SVM classifier relying on features obtained from a wearable glove worn by drivers in the study. They used an IMU for movement estimation, such as steering wheel motion and PPG, to obtain features related to drivers' pulse waves. The study was conducted in a driving simulator, with three scenarios covering city, highway, and urban driving and simulated random weather conditions. A good remark is that the authors took care of subjective assessment of driver state, with Driver Behavior Survey pre – and post-session and a survey on physical statuses of drivers. The authors also investigated drivers' facial behaviors to search for features for further stress classification. The study seems to be an excellent example of multimodality, as next to physiological, biomechanical (steering wheel movement), and subjective indices, a physician's assessments were also considered. After feature selection, SVM classifiers were trained, and one of them (SFS-SVM with RBF kernel) reached approximately *95%* accuracy in a binary stress classification. A study by Cardone et al. (2020), which was already mentioned as an example of using ECG and thermal imaging, was also oriented toward developing an ML model based on the recorded input data. "Stress Index" calculated from the ECG recordings served as a referential measure to label the data, while features from thermal monitoring were utilized in the SVM-based model development. The results obtained were a sensitivity of *77%*, a specificity of *78%*, and the correlation between the "Stress Index" and stress predicted from thermal images was $r = 0.61$ (p reported as ~ 0). From the perspective of human factors research, the assumptions for "stress" and "no stress" categories could be better supported, with, e.g., subjective assessment and the usage of the ECG-derived "Stress

Index" as a gold standard could be justified more thoroughly. One of the earlier works on multimodal approaches in driver stress recognition was published by Healey and Picard (2000). Having ECG, respiratory, EMG, and GSR measurements, but also self-reported questionnaires in a naturalistic driving study setting, the authors calculated variables and further subjected them to a feature selection procedure to create a ranking of individual predictors. Within selected features, authors investigated performance for correct stress state classification in a four-class setting (low, neutral, high, very high stress). The last study we exemplify here was reported by Lanatà et al. (2014), in which the authors aimed at recognizing three categories—normal driving, first – and second-level stress. On the basis of prior feature calculation and selection upon collected time series (ECG, EDA, respiration, simulator-derived data), the authors proposed three input datasets and three types of classifiers for further exploration. In their study, the model based on the Nearest Mean Classifier resulted in an accuracy greater than *90%*.

These were some illustratory examples of how multimodal approaches may serve us in practice and for what purposes AI can be sufficient. However, utilization of ML tools should be performed with proper care. Next, we will look over the primary benefits and some existing threats and risks that could occur while incorporating ML approaches in development. Surely, as we aim to somehow discretely or non-discretely detect and classify the state of stress, methods of Artificial Intelligence seem to be superior, in general, over heuristics – or rule-based algorithms, as the rules themselves do not seem to be clearly identifiable nor universal for stress state detection. Machine Learning tools can be particularly helpful in grasping nonlinearities and large dataset complexity, and data dimensionality can be reduced through ML-based feature selection approaches. Artificial Intelligence can greatly help integrate data or features from different modalities, as various approaches are being tailored for image data analysis, voice/natural language processing, or combinations of features from different domains. However, many open questions arise that need to be addressed when using ML tools in a project within the area of human factors research. One of them is the lack of transparency or explainability of the system, which can lead to difficulties in understanding the system's decisions. From a practical standpoint, it may lead to a lowered acceptance rate of a system if too many false positive detections induce the alerts, notifications, or means to reduce stress. There are also experimental risks in the development of AI-based systems, such as ethnic, gender, or cultural biases, due to the lack of heterogeneity of the dataset used for training algorithms. Yet another bias, which tends to be rarely discussed in such studies, is related to health statuses, as various diseases or disorders—usually due to study exclusion criteria—will probably be omitted in such ML tools and classifiers. Consequently, there will be a lack of representation of participants affected by certain medical problems in future datasets with which the systems are being trained, and this may lead to incorrect recognitions or a narrowed scope of algorithm usage. When it comes to datasets, two other shortcomings are worth discussing. One is that samples can be imbalanced in terms of "stress" and "no stress" events, which could lead to, e.g., algorithm overfitting, as there will be only a very limited representation of data associated with a given label. The classical problem of "trash in, trash out" cases also exists. Suppose the data was collected improperly (with no caretaking for proper referential measures, environment, or study sample) or improperly processed (e.g., poor selection of techniques, lack of proper filtering, or simply enough know-how about certain data processing). In that case, it will lead to incorrectly collected features, and thus, the model and its responses will be flawed.

However, most of the counterarguments listed here can be addressed with a proper study methodology and a clear pipeline for data processing. Gender, age or ethnic biases reduction can be attempted with recruitment of diverse participants. There exist methods for the class imbalances problem, e.g., data augmentation techniques, which could help reduce imbalances' effects.

However, care should and can be taken even at the level of a study protocol, with properly scheduled experimental and control conditions. When it comes to "trash in, trash out" situations, a careful selection of methods, clearly defined hypotheses, and an analysis pipeline will help to lower the possibility of such scenarios.

Artificial Intelligence techniques are currently revolutionizing different fields of human activity. Mobility and smart in-vehicle technologies are part of these activities, which are and will be largely influenced by ongoing developments in AI. ML-based algorithms may be very useful in the development of in-cabin sensing, yet care should be taken to understand and attempt to solve the challenges that are faced by ML implementations in different domains, such as explainability, liability, and biases.

Chap

CHAPTER 3.

MARKET OVERVIEW AND CONCLUDING REMARKS

AS WE LEARNED ABOUT A PLETHORA OF techniques that could be helpful in driver stress state determination, now seems to be a great time to take a look at what's going on in the industry in this matter. In the subsequent section, we will briefly describe some recent public announcements or patent applications, which are somehow related to the concept of stress detection among drivers. We will finish this brief chapter with some concluding remarks and acknowledgements, and this is the last of our journey in this handbook edition.

3.1. Market overview

Although stress detection is an ever-evolving concept in the context of driver monitoring systems, there are already several manufactured solutions and systems on the market, that take a wide variety of forms. Browsing online sources, one can find both articles and patent applications oriented at systems designed to detect, monitor, and sometimes even react to driver stress. These solutions come in different forms, including biometric systems, specialized equipment, wearable devices, and mobile applications. In this section, we will briefly present some examples of such systems, yet please note that this is not an advertisement of any described solutions, and the overview should serve only as a small set of market examples.

Among the systems available on the market, there are solutions such as "Ready Care"—a system developed by Harman that not only detects real-time stress factors but also aims to reduce stress in drivers by, for instance, suggesting the route with fewer stress factors (like traffic or bad weather) instead of the one recommended by the navigation (Soni, 2022; Harman Automotive, n.d.). Hyundai Mobis offers a solution called the "Smart Cabin Controller". Next to driver stress detection, this multi-purpose system also takes active measures to reduce its level. It incorporates features like an aroma diffuser, a 3D pattern mood lamp, and the ability to prevent motion sickness or monitor carbon dioxide levels in the car cabin and adjust it by opening the windows (Ang, 2022; Ryu, 2022). In addition to systems of this type, there are also a variety of equipment solutions and car accessories on the market to help combat stress. One of these is "The Active Wellness" seat developed by Faurecia. It is a seat that, through a biometric sensing system, senses stress or a drop in the driver's energy and reacts accordingly—a message appears on a screen in the cabin with a recommended course of action if the seat detects stress or fatigue. If the driver accepts the system's recommendations the cabin is cooled or warmed depending on the detected state of the driver (Forvia Faurecia, 2017; Stock, 2015).

It is also not unheard of for such solutions to work in combination with other tools—in some systems, for instance, "The Active Wellness", the seat is designed to work in conjunction with any wearable device, and synchronize the data with it (Stock, 2015). Such fusion of wearable devices and the vehicle's response to the data thus acquired can also be found in patent descriptions of other existing solutions (Siddiqui et al., 2017). Solutions similar to those used in Hyundai Mobis "Smart Cabin Controller", such as aroma diffusers, can also be found in patent descriptions—for example, a system that spreads cedrol in the vehicle cabin depending on the driver's stress and fatigue levels (Shinji & Teruhisa, 2018). Among the patented solutions there can also be found systems that attempt to reduce stress in drivers by making sounds or controlling the environment inside the vehicle cabin (Biss et al., 2020).

As we can conclude, the topic of driver detection and stress reduction system development is alive, and the interest of OEMs and Tier 1&2 companies is growing. We can presume that in the following years, more solutions will be market-ready, and prototypes reaching more nuanced aspects of stress detection will also be announced.

3.2. Concluding remarks

In the previous chapters, we dissected many of the key aspects accompanying the study of the complex phenomenon of driver stress. We did an accelerated "101" course about the effects of stress on the body, including statistical trends in terms of changes occurring in its various systems. We learned what tools can be useful in studying the phenomenon of stress in drivers, taking into account subjective, psychophysiological testing techniques and behavioral plus mechanical measurements, such as data from a vehicle or simulator. Then, we also learned that data from the aforementioned multiple sources can be traced integratively, using multimodal approaches that are currently more available thanks to the accelerated development of AI, among other things. We have achieved the fundamental premise for this handbook, which is the integration of information to serve as a toolbox. Now, the tools will remain in your hands and enable you to develop successful projects in driver stress research. From here, we wish you good luck and encourage you to share the results of your own research in this area. In time, perhaps we will collect enough of them to create the version 2.0 of the handbook, from which we will all learn even more about the phenomenon of stress in drivers, as well as ways to alleviate it and thus have a real impact on improving road safety.

Acknowledgements

We would like to deeply express our gratitude to our supportive Human Factors team members at Robotec.ai, who were always there for open discussions, with advices and iterative reviews of the document. Thank you Marta Kasprzak, Sylwia Mac, and Anna Olejniczak-Serowiec. We would also like to thank our colleagues from Robotec.ai – Dominika Radzimińska and Olga Reunamo, for their help with the production and dissemination of the book.

Last but not least, we would like to thank our CEO, Michał Niezgoda, who simply "trusted the process" and gave us support, space, and time to explore new research possibilities in the domain of driving studies.

List of figures and tables

References

Abdelrahman, Y., Velloso, E., Dingler, T., Schmidt, A., & Vetere, F. (2017). Cognitive heat: exploring the usage of thermal imaging to unobtrusively estimate cognitive load. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(3), 1–20.

Affanni, A., Aminosharieh Najafi, T., & Guerci, S. (2022). Development of an EEG headband for stress measurement on driving simulators. *Sensors*, 22(5), 1785.

Allahyari, T., Saraji, G. N., Adi, J., Hosseini, M., Iravani, M., Younesian, M., & Kass, S. J. (2008). Cognitive failures, driving errors and driving accidents. *International journal of occupational safety and ergonomics*, 14(2), 149–158.

Allen, J. (2007). Photoplethysmography and its application in clinical physiological measurement. *Physiological measurement*, 28(3), R1.

Ang, A., (2022) Hyundai Mobis launches health monitoring controller. *MobiHealthNews*. Retrieved May 30, 2023, from https://www.mobihealthnews.com/news/asia/hyundai-mobis-launches-health-monitoring-controller.

Anzengruber, B., & Riener, A. (2012, October). "FaceLight" potentials and drawbacks of thermal imaging to infer driver stress. In *Proceedings of the 4th international conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 209–216).

American Psychological Association. (n.d.). Cognitive flexibility. *APA Dictionary of Psychology*. Retrieved September 4, 2023, from https://dictionary.apa.org/cognitive-flexibility.

Ashton, H., Savage, R. D., Thompson, J. W., & Watson, D. W. (1972). A method for measuring human behavioural and physiological responses at different stress levels in a driving simulator. *British journal of pharmacology*, 45(3), 532–545.

Baek, H. J., Lee, H. B., Kim, J. S., Choi, J. M., Kim, K. K., & Park, K. S. (2009). Nonintrusive biological signal monitoring in a car to evaluate a driver's stress and health state. *Telemedicine and e-Health*, 15(2), 182–189.

Balters, S., Gowda, N., Ordonez, F., & Paredes, P. E. (2021). Individualized stress detection using an unmodified car steering wheel. *Scientific reports*, 11(1), 20646.

Barnard, M. P., & Chapman, P. (2018). The moderating effect of trait anxiety on anxiety-related thoughts and actions whilst driving. *Personality and individual differences*, 135, 207–211.

Betts, G., Young, K. A., Wise, J. A., Johnson, E., Poe, B., Kruse, D., Korol, O., Johnson, J., Womble, M., & DeSaix, P. (2022). Cardiac Cycle. In Anatomy and Physiology 2e. *OpenStax*. https://openstax.org/books/anatomy-and-physiology-2e/pages/19–3-cardiac-cycle. License: CC-BY 4.0.

Biernacki, M., & Tarnowski, A. (2011). The effect of age and personality on the main cognitive processes in drivers. *International Journal of Occupational Medicine and Environmental Health*, 24, 367–379.

Biss, R., Hall, S., Melvin, S. (2020). Driver stress reduction and mitigation (Application number WO2020153932 (A1)). Espacenet. Retrieved May 30, 2023, from https://worldwide.espacenet.com/publicationDetails/biblio?II=21&ND=3&adjacent=true&locale=en_EP&FT=D&date=20200730&CC=WO&NR=2020153932A1&KC=A1#.

Bitkina, O. V., Kim, J., Park, J., Park, J., & Kim, H. K. (2019). Identifying traffic context using driving stress: A longitudinal preliminary case study. *Sensors*, 19(9), 2152.

Boles, D. B., & Adair, L. P. (2001, October). The multiple resources questionnaire (MRQ). *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 45, No. 25, pp. 1790–1794). Sage CA: Los Angeles, CA: SAGE Publications.

Boles, D. B., Bursk, J. H., Phillips, J. B., & Perdelwitz, J. R. (2007). Predicting dual-task performance with the Multiple Resources Questionnaire (MRQ). *Human factors*, 49(1), 32–45.

Bracha, H. S., Ralston, T. C., Matsukawa, J. M., Williams, A. E., & Bracha, A. S. (2004). Does "fight or flight" need updating?. *Psychosomatics*, 45(5), 448–449.

Brandenburg, S., Oehl, M., & Seigies, K. (2017). German taxi drivers' experiences and expressions of driving anger: Are the driving anger scale and the driving anger expression inventory valid measures?. *Traffic injury prevention*, 18(8), 807–812.

Broadbent, D. E., Cooper, P. F., FitzGerald, P., & Parkes, K. R. (1982). The cognitive failures questionnaire (CFQ) and its correlates. *British journal of clinical psychology*, 21(1), 1–16.

Brown, G. W., & Harris, T. O. (1978). Social origins of depression: A study of depressive disorder in women.

Burlacu, A., Brinza, C., Brezulianu, A., & Covic, A. (2021). Accurate and early detection of sleepiness, fatigue and stress levels in drivers through Heart Rate Variability parameters: a systematic review. *Reviews in cardiovascular medicine*, 22(3), 845–852.

Bustos, C., Elhaouij, N., Sole-Ribalta, A., Borge-Holthoefer, J., Lapedriza, A., & Picard, R. (2021, September). Predicting driver self-reported stress by analyzing the road scene. 2021 *9th International Conference on Affective Computing and Intelligent Interaction (ACII)* (pp. 1–8). IEEE.

Byrne, D. G., Davenport, S. C., & Mazanov, J. (2007). Profiles of adolescent stress: The development of the adolescent stress questionnaire (ASQ). *Journal of adolescence*, 30(3), 393–416.

Byrne, D. G., Byrne, A. E., & Reinhart, M. I. (1995). Personality, stress and the decision to commence cigarette smoking in adolescence. *Journal of psychosomatic research*, 39(1), 53–62.

Cahir, N., & Morris, R. D. (1991). The psychology student stress questionnaire. *Journal of clinical psychology*, 47(3), 414–417.

Cardone, D., Perpetuini, D., Filippini, C., Spadolini, E., Mancini, L., Chiarelli, A. M., & Merla, A. (2020). Driver stress state evaluation by means of thermal imaging: A supervised machine learning approach based on ECG signal. *Applied Sciences*, 10(16), 5673.

Carter, J. R., Durocher, J. J., & Kern, R. P. (2008). Neural and cardiovascular responses to emotional stress in humans. American Journal of Physiology-Regulatory, *Integrative and Comparative Physiology,* 295(6), R1898-R1903.

Chui, K. T., Lytras, M. D., & Liu, R. W. (2020). A generic design of driver drowsiness and stress recognition using MOGA optimized deep MKL-SVM. *Sensors*, 20(5), 1474.

Chung, W. Y., Chong, T. W., & Lee, B. G. (2019). Methods to detect and reduce driver stress: a review. *International journal of automotive technology*, 20, 1051–1063.

Clapp, J. D., Olsen, S. A., Beck, J. G., Palyo, S. A., Grant, D. M., Gudmundsdottir, B., & Marques, L. (2011). The driving behavior survey: Scale construction and validation. *Journal of anxiety disorders*, 25(1), 96–105.

Clapp, J. D., Baker, A. S., Litwack, S. D., Sloan, D. M., & Beck, J. G. (2014). Properties of the Driving Behavior Survey among individuals with motor vehicle accident-related posttraumatic stress disorder. *Journal of anxiety disorders*, 28(1), 1–7.

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of health and social behavior*, 385–396.

Costa, P. T., & McCrae, R. R. (1989). *NEO PI/FFI manual supplement for use with the NEO Personality Inventory and the NEO Five-Factor Inventory.* Psychological Assessment Resources.

Crandall, C. S., Preisler, J. J., & Aussprung, J. (1992). Measuring life event stress in the lives of college students: The Undergraduate Stress Questionnaire (USQ). *Journal of behavioral medicine*, 15, 627–662.

Critchley, H. D. (2002). Electrodermal responses: what happens in the brain. *The Neuroscientist*, 8(2), 132–142.

Crosswell, A. D., & Lockwood, K. G. (2020). Best practices for stress measurement: How to measure psychological stress in health research. *Health psychology open,* 7(2), 2055102920933072.

Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). The Minnesota Multiphasic Personality Inventory-2 (MMPI-2): Manual for Administration and Scoring.

Dhabhar, F. S. (2014). Effects of stress on immune function: the good, the bad, and the beautiful. *Immunologic research*, 58, 193–210.

Djordjević, J. (2022, June). Psychologically Profiling Drivers: A Questionnaire. *Proceedings of the MEi: CogSci Conference* (Vol. 16, No. 1).

Dobbins, C., & Fairclough, S. (2018). Signal processing of multimodal mobile lifelogging data towards detecting stress in real-world driving. *IEEE Transactions on Mobile Computing*, 18(3), 632–644.

Dogan, D., Bogosyan, S., & Acarman, T. (2019). Evaluation of driver stress level with survey, galvanic skin response sensor data, and force-sensing resistor data. *Advances in Mechanical Engineering*, 11(12), 1687814019891555.

Dorn, L., & Matthews, G. (1995). Prediction of mood and risk appraisals from trait measures: Two studies of simulated driving. *European Journal of Personality*, 9(1), 25–42.

Dunn, A. J. (2008). The HPA axis and the immune system. In A. Del Rey, G. Chrousos & H. Besedovsky (Ed.). The hypothalamus-pituitary-adrenal axis. *Elsevier.*

Ehlers, A., Hofmann, S. G., Herda, C. A., & Roth, W. T. (1994). Clinical characteristics of driving phobia. *Journal of Anxiety Disorders,* 8(4), 323–339.

Ehlers, A., Taylor, J. E., Ehring, T., Hofmann, S. G., Deane, F. P., Roth, W. T., & Podd, J. V. (2007). The driving cognitions questionnaire: Development and preliminary psychometric properties. *Journal of anxiety disorders,* 21(4), 493–509.

Eijckelhof, B. H. W., Huysmans, M. A., Bruno Garza, J. L., Blatter, B. M., Van Dieën, J. H., Dennerlein, J. T., & Van Der Beek, A. J. (2013). The effects of workplace stressors on muscle activity in the neck-shoulder and forearm muscles during computer work: a systematic review and meta-analysis. *European Journal of Applied Physiology*, 113, 2897–2912.

Eilebrecht, B., Wolter, S., Lem, J., Lindner, H. J., Vogt, R., Walter, M., & Leonhardt, S. (2012, September). The relevance of HRV parameters for driver workload detection in real world driving. In *2012 Computing in Cardiology* (pp. 409–412). IEEE.

El Haouij, N., Ghozi, R., Poggi, J. M., Ghalila, S., & Jaidane, M. (2015). Feature extraction and selection of electrodermal reaction towards stress level recognition: Two real-world driving experiences. 47emes *Journees de Statistique de la SFdS (JdS)*, 1–5.

Eysenck, S. B., Easting, G., & Pearson, P. R. (1984). Age norms for impulsiveness, venturesomeness and empathy in children. *Personality and Individual Differences*, 5(3), 315–321.

Eysenck, S. B., Pearson, P. R., Easting, G., & Allsopp, J. F. (1985). Age norms for impulsiveness, venturesomeness and empathy in adults. *Personality and Individual Differences*, 6(5), 613–619.

Fatima, M., Gulzar, K., Khan, K. R., Amjad, F., & Shafique, M. (2020, July). Trapezius or facial muscles: Which one is more suiTable for the measurement of stress using sEMG signals?. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (pp. 670–673). IEEE.

Forvia Faurecia. (2017). The "Active Wellness" car seat wins at iF DESIGN AWARD | Faurecia. *Faurecia*. Retrieved May 30, 2023, from https://www.faurecia.com/en/newsroom/active-wellness-car-seat-wins-if-design-award.

Friedman, J. H. (2021). Cognitive Sinks & the Bermuda Triangle of fMRI Cognitive Studies. *Rhode Island Medical Journal*, 104(3), 12–13.

Fry, W. F., & Savin, W. M. (1988). Mirthful laughter and blood pressure.

Funke, G., Matthews, G., Warm, J. S., & Emo, A. K. (2007). Vehicle automation: A remedy for driver stress?. *Ergonomics*, 50(8), 1302–1323.

Gabaude, C., Baracat, B., Jallais, C., Bonniaud, M., & Fort, A. (2012). Cognitive load measurement while driving. *Human Factors: a view from an integrative perspective*, 67–80.

Gao, H., Yüce, A., & Thiran, J. P. (2014, October). Detecting emotional stress from facial expressions for driving safety. In *2014 IEEE International Conference on Image Processing (ICIP)* (pp. 5961–5965). IEEE.

Ge, Y., Qu, W., Jiang, C., Du, F., Sun, X., & Zhang, K. (2014). The effect of stress and personality on dangerous driving behavior among Chinese drivers. *Accident Analysis & Prevention*, 73, 34–40.

Germer, C. K., & Neff, K. D. (2015). Cultivating self-compassion in trauma survivors. Mindfulness-oriented interventions for trauma: Integrating contemplative practices, 43, 58.

Giannakakis, G., Grigoriadis, D., Giannakaki, K., Simantiraki, O., Roniotis, A., & Tsiknakis, M. (2019). Review on psychological stress detection using biosignals. *IEEE Transactions on Affective Computing*, 13(1), 440–460.

Gioia, F., Greco, A., Callara, A. L., & Scilingo, E. P. (2022). Towards a contactless stress classification using thermal imaging. *Sensors*, 22(3), 976.

Gray, M. J., Litz, B. T., Hsu, J. L., & Lombardo, T. W. (2004). Psychometric properties of the life events checklist. *Assessment*, 11(4), 330–341.

Grossman, P. (1983). Respiration, stress, and cardiovascular function. *Psychophysiology*, 20(3), 284–300.

Gruden, T., Stojmenova, K., Sodnik, J., & Jakus, G. (2019). Assessing drivers' physiological responses using consumer grade devices. *Applied Sciences*, 9(24), 5353.

Gulian, E., Glendon, A. I., Matthews, G., Davies, D. R., & Debney, L. M. (1990). The stress of driving: A diary study. *Work & Stress*, 4(1), 7–16.

Gulian, E., Matthews, G., Glendon, A. I., Davies, D. R., & Debney, L. M. (1989). Dimensions of driver stress. *Ergonomics*, 32(6), 585–602.

Guo, F., Simons-Morton, B. G., Klauer, S. E., Ouimet, M. C., Dingus, T. A., & Lee, S. E. (2013). Variability in crash and near-crash risk among novice teenage drivers: a naturalistic study. *The Journal of pediatrics*, 163(6), 1670–1676.

Guo, M., Wei, W., Liao, G., & Chu, F. (2016). The impact of personality on driving safety among Chinese high-speed railway drivers. *Accident Analysis & Prevention*, 92, 9–14.

Harker, M. (2013). Psychological sweating: A systematic review focused on aetiology and cutaneous response. *Skin pharmacology and physiology*, 26(2), 92–100.

Harman Automotive. (n.d.). HARMAN Ready Care – In-vehicle Safety and Wellbeing. Connected by HARMAN. *Harman Automotive*. Retrieved May 25, 2023, from https://car.harman.com/experiences/ready-care.

Harris, J. A., Saltstone, R., & Fraboni, M. (1999). An evaluation of the job stress questionnaire with a sample of entrepreneurs. *Journal of Business and Psychology*, 13, 447–455.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology* (Vol. 52, pp. 139–183). North-Holland.

Hathaway, S. R., & McKinley, J. C. (1943). The Minnesota multiphasic personality inventory, Rev. ed., 2nd printing.

Healey, J., & Picard, R. (2000, September). SmartCar: detecting driver stress. In *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000* (Vol. 4, pp. 218–221). IEEE.

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

Heathers, J. A. (2014). Everything Hertz: methodological issues in short-term frequency-domain HRV. *Frontiers in physiology*, 5, 177.

Hennessy, D. A., & Wiesenthal, D. L. (1997). The relationship between traffic congestion, driver stress and direct versus indirect coping behaviours. *Ergonomics*, 40(3), 348–361.

Hennessy, D. A., & Wiesenthal, D. L. (1999). Traffic congestion, driver stress, and driver aggression. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression*, 25(6), 409–423.

Hennessy, D. A., Wiesenthal, D. L., & Kohn, P. M. (2000). The influence of traffic congestion, daily hassles, and trait stress susceptibility on state driver stress: an interactive perspective 1. *Journal of Applied Biobehavioral Research*, 5(2), 162–179.

Hennessy, D. A., & Wiesenthal, D. L. (2001). Gender, driver aggression, and driver violence: An applied evaluation. *Sex Roles*, 44(11–12), 661–676.

Hennessy, D. A., & Wiesenthal, D. L. (2002). The relationship between driver aggression, violence, and vengeance. *Violence and Victims*, 17(6), 707–718.

Herd, J. A. (1991). Cardiovascular response to stress. *Physiological reviews*, 71(1), 305–330.

Hernando, A., Lázaro, J., Arza, A., Garzón, J. M., Gil, E., Laguna, P., ... & Bailón, R. (2015, September). Changes in respiration during emotional stress. In *2015 Computing in Cardiology Conference (CinC)* (pp. 1005–1008). IEEE.

Holmgren, K., Dahlin-Ivanoff, S., Björkelund, C., & Hensing, G. (2009). The prevalence of work-related stress, and its association with self-perceived health and sick-leave, in a population of employed Swedish women. *BMC public health*, 9(1), 1–10.

Huang, J., Luo, X., & Peng, X. (2020). A novel classification method for a driver's cognitive stress level by transferring interbeat intervals of the ECG signal to pictures. *Sensors*, 20(5), 1340.

Huang, M. X., Li, J., Ngai, G., & Leong, H. V. (2016, October). Stressclick: Sensing stress from gaze-click patterns. In *Proceedings of the 24th ACM international conference on Multimedia* (pp. 1395–1404).

Ismail, R., Voon, N. L., Saad, M. H. M., Saleem, M., & Ibrahim, N. (2016). Aberrant driving among young Malaysian drivers. *Jurnal Teknologi*, 78(6–10), 55–62.

Jafari, Z., Kolb, B. E., & Mohajerani, M. H. (2017). Effect of acute stress on auditory processing: a systematic review of human studies. *Reviews in the Neurosciences*, 28(1), 1–13.

Kalia, V., Vishwanath, K., Knauft, K., Vellen, B. V. D., Luebbe, A., & Williams, A. (2018). Acute stress attenuates cognitive flexibility in males only: an fNIRS examination. *Frontiers in psychology*, 9, 2084.

Karatsoreos, I. N. (2018). Stress: Common themes toward the next frontier. *Frontiers in neuroendocrinology*, 49, 3–7.

Kerautret, L., Dabic, S., & Navarro, J. (2021). Sensitivity of physiological measures of acute driver stress: A meta-analytic review. *Frontiers in neuroergonomics*, 2, 756473.

Keraudren, L., Dabic, S., & Navarro, J. (2022). Detecting driver stress and hazard anticipation using real-time cardiac measurement: A simulator study. *Brain and behavior*, 12(2), e2424.

Kim, H. G., Cheon, E. J., Bai, D. S., Lee, Y. H., & Koo, B. H. (2018). Stress and heart rate variability: a meta-analysis and review of the literature. *Psychiatry investigation*, 15(3), 235.

Kleiger, R. E., Stein, P. K., & Bigger Jr, J. T. (2005). Heart rate variability: measurement and clinical utility. *Annals of Noninvasive Electrocardiology*, 10(1), 88–101.

Kogler, L., Müller, V. I., Chang, A., Eickhoff, S. B., Fox, P. T., Gur, R. C., & Derntl, B. (2015). Psychosocial versus physiological stress—Meta-analyses on deactivations and activations of the neural correlates of stress reactions. *Neuroimage*, 119, 235–251.

Kontogiannis, T. (2006). Patterns of driver stress and coping strategies in a Greek sample and their relationship to aberrant behaviors and traffic accidents. *Accident Analysis & Prevention*, 38(5), 913–924.

Koolhaas, J. M., Bartolomucci, A., Buwalda, B., de Boer, S. F., Flügge, G., Korte, S. M., ... & Fuchs, E. (2011). Stress revisited: a critical evaluation of the stress concept. *Neuroscience & Biobehavioral Reviews*, 35(5), 1291–1301.

Kotynia, M., & Stróż, A. (2023). Arousal level in driver state benchmarking – application of ECG in automotive. *Robotec.ai.* Retrieved May 25, 2023, from https://robotec.ai/arousal-level-in-driver-state-benchmarking-application-of-ecg-in-automotive/.

Lajunen, T., & Summala, H. (1995). Driving experience, personality, and skill and safety-motive dimensions in drivers' self-assessments. *Personality and individual differences*, 19(3), 307–318.

Lanatà, A., Valenza, G., Greco, A., Gentili, C., Bartolozzi, R., Bucchi, F., ... & Scilingo, E. P. (2014). How the autonomic nervous system and driving style change with incremental stressing conditions during simulated driving. *IEEE Transactions on Intelligent Transportation Systems*, 16(3), 1505–1517.

Larson, G. E., & Merritt, C. R. (1991). Can accidents be predicted? An empirical test of the Cognitive Failures Questionnaire. *Applied Psychology*, 40(1), 37–45.

LeBlanc, V. R. (2009). The effects of acute stress on performance: implications for health professions education. *Academic Medicine*, 84(10), S25-S33.

Lee, D. S., Chong, T. W., & Lee, B. G. (2016). Stress events detection of driver by wearable glove system. *IEEE Sensors Journal*, 17(1), 194–204.

Lee, J., Lee, H., & Shin, M. (2021). Driving stress detection using multimodal convolutional neural networks with nonlinear representation of short-term physiological signals. *Sensors*, 21(7), 2381.

Levine, J. A., Pavlidis, I. T., MacBride, L., Zhu, Z., & Tsiamyrtzis, P. (2009). Description and clinical studies of a device for the instantaneous detection of office-place stress. *Work*, 34(3), 359–364.

Li, F., Li, C., Long, Y., Zhan, C., & Hennessy, D. A. (2004). Reliability and validity of aggressive driving measures in China. *Traffic injury prevention*, 5(4), 349–355.

Liu, H., Allen, J., Zheng, D., & Chen, F. (2019). Recent development of respiratory rate measurement technologies. *Physiological measurement*, 40(7), 07TR01.

Lundberg, U., Forsman, M., Zachau, G., Eklöf, M., Palmerud, G., Melin, B., & Kadefors, R. (2002). Effects of experimentally induced mental and physical stress on motor unit recruitment in the trapezius muscle. *Work & Stress*, 16(2), 166–178.

Lupien, S. J., McEwen, B. S., Gunnar, M. R., & Heim, C. (2009). Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nature reviews neuroscience*, 10(6), 434–445.

Mackay, C., Cox, T., Burrows, G., & Lazzerini, T. (1978). An inventory for the measurement of self-reported stress and arousal. *British Journal of Social & Clinical Psychology.*

MacLeod, C. M. (2007). The concept of inhibition in cognition. In D. S. Gorfein & C. M. MacLeod (Eds.), *Inhibition in cognition* (pp. 3–23). American Psychological Association. https://doi.org/10.1037/11587–001.

Marin, M. F., Lord, C., Andrews, J., Juster, R. P., Sindi, S., Arsenault-Lapierre, G., ... & Lupien, S. J. (2011). Chronic stress, cognitive functioning and mental health. *Neurobiology of learning and memory*, 96(4), 583–595.

Martinussen, L. M., Møller, M., & Prato, C. G. (2014). Assessing the relationship between the Driver Behavior Questionnaire and the Driver Skill Inventory: Revealing subgroups of drivers. *Transportation research part F: traffic psychology and behaviour*, 26, 82–91.

Masaoka, Y., & Homma, I. (1997). Anxiety and respiratory patterns: their relationship during mental stress and physical load. *International Journal of Psychophysiology*, 27(2), 153–159.

Matthews, G., Jones, D. M., & Chamberlain, A. G. (1990). Refining the measurement of mood: The UWIST mood adjective checklist. *British journal of psychology*, 81(1), 17–42.

Matthews, G., & Desmond, P. A. (1998). Personality and multiple dimensions of task-induced fatigue: A study of simulated driving. *Personality and Individual Differences*, 25(3), 443–458.

Matthews, G., Joyner, L., Gilliland, K., Campbell, S., Falconer, S., & Huggins, J. (1999). Dundee Stress State Questionnaire. *Emotion*.

Mayo, L. M., & Heilig, M. (2019). In the face of stress: Interpreting individual differences in stress-induced facial expressions. *Neurobiology of stress*, 10, 100166.

Măirean, C., Havârneanu, G. M., Popușoi, S. A., & Havarneanu, C. E. (2017). Traffic locus of control scale–Romanian version: psychometric properties and relations to the driver's personality, risk perception, and driving behavior. *Transportation research part F: traffic psychology and behaviour*, 45, 131–146.

McEwen, B. S. (2008). Central effects of stress hormones in health and disease: Understanding the protective and damaging effects of stress and stress mediators. *European journal of pharmacology*, 583(2–3), 174–185.

McEwen, B. S. (2017). Neurobiological and systemic effects of chronic stress. *Chronic stress*, 1, 2470547017692328.

Mohamad, N. A. (2022). Stress and Anxiety on the Road: The Silent Victims and Their Sufferings. *Journal of Advanced Vehicle System*, 13(1), 57–69.

Moses, T. E., Gray, E., Mischel, N., & Greenwald, M. K. (2023). Effects of neuromodulation on cognitive and emotional responses to psychosocial stressors in healthy humans. *Neurobiology of Stress*, 100515.

Mou, L., Zhou, C., Zhao, P., Nakisa, B., Rastgoo, M. N., Jain, R., & Gao, W. (2021). Driver stress detection via multimodal fusion using attention-based CNN-LSTM. *Expert Systems with Applications*, 173, 114693.

Munla, N., Khalil, M., Shahin, A., & Mourad, A. (2015, September). Driver stress level detection using HRV analysis. In *2015 international conference on advances in biomedical engineering (ICABME)* (pp. 61–64). IEEE.

Nagasawa, T., Takahashi, R., Koopipat, C., & Tsumura, N. (2020). Stress estimation using multimodal biosignal information from RGB facial video. In P*roceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (pp. 292–293).

Natvig, G. K., Albrektsen, G., Anderssen, N., & Qvarnstrøm, U. (1999). School-related stress and psychosomatic symptoms among school adolescents. *Journal of school health*, 69(9), 362–368.

Nilsen, K. B., Sand, T., Stovner, L. J., Leistad, R. B., & Westgaard, R. H. (2007). Autonomic and muscular responses and recovery to one-hour laboratory mental stress in healthy subjects. *BMC Musculoskeletal Disorders*, 8, 1–12.

Noordewier, M. K., Scheepers, D. T., & Hilbert, L. P. (2020). Freezing in response to social threat: a replication. *Psychological Research*, 84(7), 1890–1896.

Ogorevc, J., Podlesek, A., Geršak, G., & Drnovšek, J. (2011, May). The effect of mental stress on psychophysiological parameters. In *2011 IEEE international symposium on medical measurements and applications* (pp. 294–299). IEEE.

Ohmi, M., Tanigawa, M., Yamada, A., Ueda, Y., & Haruna, M. (2009). Dynamic analysis of internal and external mental sweating by optical coherence tomography. *Journal of Biomedical Optics*, 14(1), 014026–014026.

Owsley, C., McGwin Jr, G., & McNeal, S. F. (2003). Impact of impulsiveness, venturesomeness, and empathy on driving by older adults. *Journal of Safety Research*, 34(4), 353–359.

Özkan, T., & Lajunen, T. (2005). Multidimensional Traffic Locus of Control Scale (T-LOC): factor structure and relationship to risky driving. *Personality and individual differences*, 38(3), 533–545.

Pauzié, A. (2008). A method to assess the driver mental workload: The driving activity load index (DALI). *IET Intelligent Transport Systems*, 2(4), 315–322.

Pedrotti, M., Mirzaei, M. A., Tedesco, A., Chardonnet, J. R., Mérienne, F., Benedetto, S., & Baccino, T. (2014). Automatic stress classification with pupil diameter analysis. *International Journal of Human-Computer Interaction*, 30(3), 220–236.

Pekruna, R. (2020). Commentary: Self-Report Is Indispensable to Assess Students' Learning. *Frontline Learning Research*, 8(3), 185–193.

Pettersson, I., Lachner, F., Frison, A. K., Riener, A., & Butz, A. (2018, April). A Bermuda triangle? A Review of method application and triangulation in user experience evaluation. *Proceedings of the 2018 CHI conference on human factors in computing systems* (pp. 1–16).

Picard, R. W., Fedor, S., & Ayzenberg, Y. (2016). Multiple arousal theory and daily-life electrodermal activity asymmetry. *Emotion review*, 8(1), 62–75.

Piranveyseh, P., Kazemi, R., Soltanzadeh, A., & Smith, A. (2022). A field study of mental workload: conventional bus drivers versus bus rapid transit drivers. *Ergonomics*, 65(6), 804–814.

Porcelli, A. J., Cruz, D., Wenberg, K., Patterson, M. D., Biswal, B. B., & Rypma, B. (2008). The effects of acute stress on human prefrontal working memory systems. *Physiology & behavior*, 95(3), 282–289.

Posada-Quintero, H. F., & Chon, K. H. (2020). Innovations in electrodermal activity data collection and signal processing: A systematic review. *Sensors*, 20(2), 479.

Pourmohammadi, S., & Maleki, A. (2013). An automatic approach to continuous stress assessment during driving based on fuzzy c-means clustering. *The Modares Journal of Electrical Engineering*, 13(1), 9–17.

Raggatt, P. T., & Morrissey, S. A. (1997). A field study of stress and fatigue in long-distance bus drivers. *Behavioral medicine*, 23(3), 122–129.

Raven, J. C., & John Hugh Court. (1998). *Raven's progressive matrices and vocabulary scales* (pp. 223–237). Oxford: Oxford Psychologists Press.

Ren, P., Barreto, A., Huang, J., Gao, Y., Ortega, F. R., & Adjouadi, M. (2014). Off-line and on-line stress detection through processing of the pupil diameter signal. *Annals of biomedical engineering,* 42, 162–176.

Reimer, B. (2009). Impact of cognitive task complexity on drivers' visual tunneling. *Transportation Research Record*, 2138(1), 13–19.

Reimer, B., Mehler, B., & Coughlin, J. F. (2016). Reductions in self-reported stress and anticipatory heart rate with the use of a semi-automated parallel parking system. *Applied ergonomics*, 52, 120–127.

Romero-Martínez, Á., Hidalgo-Moreno, G., & Moya-Albiol, L. (2020). Neuropsychological consequences of chronic stress: the case of informal caregivers. *Aging & mental health*, 24(2), 259–271.

Ryu, M. (2022). Hyundai Mobis M.VICS And Smart Cabin Secures Safe And Sound Autonomous Driving. *Hyundai Motor Group*. Retrieved May 30, 2023, from https://www.hyundaimotorgroup.com/story/CONT0000000000043965.

Sakamoto, K., Aoyama, S., Asahara, S., Mizushina, H., & Kaneko, H. (2009). Relationship between emotional state and pupil diameter variability under various types of workload stress. In *Ergonomics and Health Aspects of Work with Computers: International Conference, EHAWC 2009, Held as Part of HCI International 2009*, San Diego, CA, USA, July 19–24, 2009. Proceedings (pp. 177–185). Springer Berlin Heidelberg.

Salas, E., Kozlowski, S. W., & Chen, G. (2017). A century of progress in industrial and organizational psychology: Discoveries and the next century. *Journal of Applied Psychology*, 102(3), 589.

Sandi, C. (2013). Stress and cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 4(3), 245–261.

Saxby, D. J., Matthews, G., Hitchcock, E. M., Warm, J. S., Funke, G. J., & Gantzer, T. (2008, September). Effect of active and passive fatigue on performance using a driving simulator. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 52, No. 21, pp. 1751–1755). Sage CA: Los Angeles, CA: Sage Publications.

Schubert, C., Lambertz, M., Nelesen, R. A., Bardwell, W., Choi, J. B., & Dimsdale, J. E. (2009). Effects of stress on heart rate complexity—a comparison between short-term and chronic stress. *Biological psychology*, 80(3), 325–332.

Shackman, A. J., Maxwell, J. S., McMenamin, B. W., Greischar, L. L., & Davidson, R. J. (2011). Stress potentiates early and attenuates late stages of visual processing. *Journal of Neuroscience*, 31(3), 1156–1161.

Shaffer, F., & Ginsberg, J. P. (2017). An overview of heart rate variability metrics and norms. *Frontiers in public health*, 258.

Sharit, J., & Salvendy, G. (1982). Occupational stress: Review and reappraisal. *Human factors*, 24(2), 129–162.

Sharma, N., Dhall, A., Gedeon, T., & Goecke, R. (2013, September). Modeling stress using thermal facial patterns: A spatio-temporal approach. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction* (pp. 387–392). IEEE.

Sharma, M., Kacker, S., & Sharma, M. (2016). A brief introduction and review on galvanic skin response. *Int J Med Res Prof*, 2(6), 13–17.

Shields, G. S., Sazma, M. A., & Yonelinas, A. P. (2016). The effects of acute stress on core executive functions: A meta-analysis and comparison with cortisol. *Neuroscience & Biobehavioral Reviews*, 68, 651–668.

Shimizu, T., Shimizu, S., Higashi, Y., & Saito, M. (2021). Psychological/mental stress-induced effects on urinary function: Possible brain molecules related to psychological/mental stress-induced effects on urinary function. *International Journal of Urology*, 28(11), 1093–1104.

Shinji, A., Teruhisa, K,. (2018). Stress relieving device for vehicle (Application number CN20168065846 20161021). *Espacenet*. Retrieved May 30, 2023, from https://worldwide.espacenet.com/publicationDetails/biblio?II=28&ND=3&adjacent=true&locale=en_EP&FT=D&date=20180717&CC=CN&NR=108290477A&KC=A#.

Siddiqui, A. N., Greenberg, J. A., Ignaczak, B., Holub, P. K., Van Wiemeersch, J. R., Strumolo, G. S., Mitra, P., Melcher, D., A Cuddihy, M., Simonds, C., Hassani, A. (2017). Incapacitated driving detection and prevention (Application number GB20160018185 20161027). *Espacenet*. Retrieved May 30, 2023, from https://worldwide.espacenet.com/publicationDetails/biblio?II=32&ND=3&adjacent=true&locale=en_EP&FT=D&date=20170614&CC=GB&NR=2545317A&KC=A#.

Skippon, S. M., Reed, N., Luke, T., Robbins, R., & Chattington, M. (2008, September). Questionnaire measures of attitudes and driving behaviour: their relationships to driving behaviour measured in a driving simulator. In 한국자동차공학회 *Symposium* (pp. 111–118).

Slavich, G. M., & Shields, G. S. (2018). Assessing lifetime stress exposure using the Stress and Adversity Inventory for Adults (Adult STRAIN): An overview and initial validation. *Psychosomatic Medicine,* 80(1), 17.

Smorti, M., & Guarnieri, S. (2016). Exploring the factor structure and psychometric properties of the Manchester Driver Behavior Questionnaire (DBQ) in an Italian sample. *Test. Psychom. Methodol. App. Psychol,* 23, 185–202.

Soni, M. (2022, September 15). Harman Launches Car Driver Fatigue Detection System For OEM. *Thrust Zone*. Retrieved May 25, 2023, from https://www.thrustzone.com/harman-launches-car-driver-fatigue-detection-system-for-oem/.

Spielberger, C. D. (1979). State-trait personality inventory.

Spielberger, C.D. (1988). Professional Manual for the State–Trait Anger Expression Inventory. Odessa, FL: *Psychological Assessment Resources*.

Stock, M. (2015, September 23). Intelligent car seat detects driver's stress level. *Reuters*. Retrieved May 30, 2023, from https://www.reuters.com/article/us-car-technology-idUSKCN0RN11P20150923.

Suess, W. M., Alexander, A. B., Smith, D. D., Sweeney, H. W., & Marion, R. J. (1980). The effects of psychological stress on respiration: a preliminary study of anxiety and hyperventilation. *Psychophysiology*, 17(6), 535–540.

Thijs, J. A. A., Muhlsteef, J., Linter, R. (2010) Method and system, for monitoring vital body signs of a seated person (Application number US20080679316 20080919). *Espacenet*. Retrieved May 30, 2023, from https://worldwide.espacenet.com/publicationDetails/biblio?II=50&ND=3&adjacent=true&locale=en_EP&FT=D&date=20100902&CC=US&NR=2010222687A1&KC=A1#.

Tinella, L., Caffò, A. O., Lopez, A., Grattagliano, I., & Bosco, A. (2021). The impact of two MMPI-2-based models of personality in predicting driving behavior. can demographic variables be disregarded?. *Brain sciences*, 11(3), 313.

Torres-Salomao, L. A., Mahfouf, M., & El-Samahy, E. (2015, October). Pupil diameter size marker for incremental mental stress detection. In *2015 17th international conference on e-health networking, application & services (HealthCom)* (pp. 286–291). IEEE.

Tronstad, C., Amini, M., Bach, D. R., & Martinsen, Ø. G. (2022). Current trends and opportunities in the methodology of electrodermal activity measurement. *Physiological measurement*, 43(2), 02TR01.

Tsigos, C., & Chrousos, G. P. (2002). Hypothalamic–pituitary–adrenal axis, neuroendocrine factors and stress. *Journal of psychosomatic research*, 53(4), 865–871.

Turner, A. I., Smyth, N., Hall, S. J., Torres, S. J., Hussein, M., Jayasinghe, S. U., ... & Clow, A. J. (2020). Psychological stress reactivity and future health and disease outcomes: A systematic review of prospective evidence. *Psychoneuroendocrinology*, 114, 104599.

Useche, S. A., Cendales, B., Lijarcio, I., & Llamazares, F. J. (2021). Validation of the F-DBQ: A short (and accurate) risky driving behavior questionnaire for long-haul professional drivers. *Transportation research part F: traffic psychology and behaviour*, 82, 190–201.

Van Wijk, C. H. (2014). The use of Spielberger's State-Trait Personality Inventory (trait anxiety subscale) with naval subaquatic specialists. *International journal of occupational medicine and environmental health*, 27, 959–966.

Wang, X., Duan, H., Kan, Y., Wang, B., Qi, S., & Hu, W. (2019). The creative thinking cognitive process influenced by acute stress in humans: an electroencephalography study. *Stress*, 22(4), 472–481.

Warner, H. W., Özkan, T., & Lajunen, T. (2010). Can the traffic locus of control (T-LOC) scale be successfully used to predict Swedish drivers' speeding behaviour?. *Accident Analysis & Prevention*, 42(4), 1113–1117.

Wehrwein, E. A., Orer, H. S., & Barman, S. M. (2016). Overview of the Anatomy, Physiology, and Pharmacology of the Autonomic Nervous System. *Comprehensive Physiology*, 6(3):1239–78.

Wemm, S. E., & Wulfert, E. (2017). Effects of acute stress on decision making. *Applied psychophysiology and biofeedback*, 42, 1–12.

Wickens, C. M., & Wiesenthal, D. L. (2005). State Driver Stress as a Function of Occupational Stress, Traffic Congestion, and Trait Stress Susceptibility 1. *Journal of Applied Biobehavioral Research*, 10(2), 83–97.

Wiesenthal, D. L., Hennessy, D., & Gibson, P. M. (2000). The Driving Vengeance Questionnaire (DVQ): The development of a scale to measure deviant drivers' attitudes. *Violence and Victims*, 15(2), 115–136.

Wijsman, J., Grundlehner, B., Penders, J., & Hermens, H. (2013). Trapezius muscle EMG as predictor of mental stress. *ACM transactions on embedded computing systems (TECS)*, 12(4), 1–20.

Yamakoshi, T., Yamakoshi, K. I., Tanaka, S., Nogawa, M., Park, S. B., Shibata, M., ... & Hirose, Y. (2008, August). Feasibility study on driver's stress detection from differential skin temperature measurement. In *2008 30th annual international conference of the ieee engineering in medicine and biology society* (pp. 1076–1079). IEEE.

Zhai, J., Barreto, A. B., Chin, C., & Li, C. (2005, April). Realization of stress detection using psychophysiological signals for improvement of human-computer interactions. In *Proceedings. IEEE SoutheastCon*, 2005. (pp. 415–420). IEEE.

Zhang, W., Hashemi, M. M., Kaldewaij, R., Koch, S. B., Beckmann, C., Klumpers, F., & Roelofs, K. (2019). Acute stress alters the 'default'brain processing. *Neuroimage*, 189, 870–877.

Zontone, P., Affanni, A., Piras, A., & Rinaldo, R. (2021). Skin potential response for stress recognition in simulated urban driving. *Acta IMEKO*, 10(4), 117–123.

About Robotec.ai

Robotec.ai is a software company with three pillars of activity – Robotics, Machine Learning, and Human Factors. For the automotive industry, Robotec.ai provides services at different stages of testing and prototyping, with simulations for sensors and autonomous vehicles, contributions to open-source projects, such as the ROS2 library, and expertise in the field of Human Factors and Human-Machine Interaction. The Human Factors team—which we, the authors, are a part of—helps our customers in the development of Advanced Driver Assistance Systems, with emphasis on Interior Sensing and Driver Monitoring Systems. Through our expertise in the regulatory space, psychophysiology, data science, and hardware engineering, we can provide full assistance in testing and validating of ADAS, specifically, DMS.

Anna Stróż is a Senior Scientist at Robotec.ai. She completed second cycle degree studies at University of Warsaw in two fields: Cognitive Science and Applications of Physics in Biology and Medicine, with a major in Neuroinformatics. Anna is experienced in psychophysiological study protocols, with an emphasis on bioelectrical data acquisition and analysis. Since 2019, Anna has been involved in the projects for automotive industry, with particular expertise in drowsiness recognition among drivers. She is the leading author of the handbook.

Olga Zdzienicka is a DMS Researcher at Robotec.ai. She completed second cycle degree studies at University of Warsaw in two fields: Biology and Cognitive Science with a specialty in neurocognitive science. Her studies gave her experience in both behavioral research and research design and have provided her with insight into the world of AI and ML. Olga joined Robotec.ai forces in 2022, where she pursues her creative interests and fierce research. Olga is the contributing author responsible for the sections about subjective assessment of stress and market overview.