TECHNISCHE UNIVERSITÄT WIEN

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„REALSTRESS - A tool for stress measurement with real-time feedback“

verfasst von / submitted by

Hanna-Helene Steiner, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2022/2023 / Vienna, 2022/2023

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 453
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Master's programme Biomedical Engineering
Betreut von / Supervisor:	Vincze, Markus; Ao.Univ.Prof. Dipl.-Ing. Dr.techn.
Mitbetreut von / Co-Supervisor:	Kaniusas, Eugenijus; Univ.Prof. Dipl.-Ing. Dr.techn.

Acknowledgements

I would like to take this opportunity to thank Dr. Lara Lammer for here consistent words of encouragement and advice throughout the process of working on this thesis. Furthermore, I express my gratitude to Prof. Markus Vincze and the Automation and Control Institute (ACIN) at the Technical University of Vienna for giving me the opportunity to realize this topic in form of my master thesis and financing this research. Furthermore, I would have never taken on this topic without the most inspiring lectures about biomedical sensors and instrumentation given by Prof. Eugenijus Kaniusas.

Many thanks go out to all participants that agreed to be part of the experimental trials of this thesis, who were essential for answering the research questions presented.

Abstract

Human physiological response to stress, designed to protect the body in face of danger, presents a serious health threat in today's society as stressors persist over longer periods of time, preventing the human body from recovering properly. Thereafter, identifying biosignals associated with the early detection of stress response, and designing stress measurement tools, has been in the focus of many research papers in recent years. However, only a small group of researchers concerned themselves with real-time stress measurements or low-cost biomedical sensors in this context.
This thesis examined the usability of low-cost biomedical sensors for the detection of human stress response in a real-time feedback implementation called REALSTRESS. Previous research in this field showed mostly self-fabricated, non-customizable or expensive systems in their work. Therefore, this work aims to present results with a system that is easily reproducible with a lower budget.

For the first part of this thesis, a selection of biomedical sensors were tested on five participants in a protocol based on the trier social stress test. Thereafter, the live-feedback stress detection REALSTRESS was implemented and reviewed within a laboratory setting based on the results of the thesis's first part. Statistical analysis of the data showed that two out of the five low-cost sensors could be used for stress detection within a laboratory setting (p-value < 0.05). The two sensors in question, photoplethysmography (PPG) and electrodermal activity, were then implemented using a sliding window approach to realize the real-time feedback sensing. The final implementation of REALSTRESS was successful in differentiating between the phases of the stress test protocol within the control group. The averaged stress score showed a difference of 30% between resting and stress phase for an arithmetic exercise. Between the two sensors, PPG provided a higher score divergence (50%).

The presented tool, REALSTRESS, has the potential to be used for simple stress detection in educational and research fields. The next step for this system would be the design of an appropriate visualization of the stress scale.

Kurzfassung

Die physiologische Stressreaktion des Menschen, die den Körper vor Gefahren schützt, stellt in der heutigen Gesellschaft eine ernsthafte Bedrohung für die menschliche Gesundheit dar, da Stressoren über längere Zeiträume hinweg bestehen bleiben und den menschlichen Körper daran hindern, sich wieder zu erholen. Die Identifizierung von Biosignalen, die mit der frühzeitigen Erkennung von Stressreaktionen in Zusammenhang stehen, und die Entwicklung von Instrumenten zur Stressmessung, standen in den letzten Jahren im Mittelpunkt zahlreicher Forschungsarbeiten. Allerdings hat sich nur eine kleine Gruppe von Forschern mit Echtzeit-Stressmessungen oder kostengünstigen biomedizinischen Sensoren in diesem Zusammenhang beschäftigt.
In dieser Arbeit wurde die Verwendbarkeit kostengünstiger biomedizinischer Sensoren für die Erkennung menschlicher Stressreaktionen in einer Labor- und Live-Feedback-Implementierung untersucht. Frühere Forschungsarbeiten aus diesem Gebiet haben vorwiegend eigen-hergestellte, nicht selbst konfigurierbare oder teure Systeme vorgestellt. Diese Arbeit zielt darauf ab, Ergebnisse eines Systems zu präsentieren, welches auch mit einem geringeren Budget leicht reproduzierbar ist.

Im ersten Teil dieser Arbeit wurden ausgesuchte biomedizinische Sensoren an fünf Teilnehmern in einem Protokoll getestet, das auf dem Trier Social Stress Test basiert. AnschlieSSend wurde die Live-Feedback-Stresserkennung, REALSTRESS, implementiert und in einer Laborumgebung auf der Grundlage der Ergebnisse des ersten Teils der Arbeit überprüft. Die statistische Analyse der Daten zeigte, dass zwei der fünf Low-Cost-Sensoren für die Stresserkennung in einer Laborumgebung verwendet werden konnten, indem die Nullhypothese gleicher Mittelwerte durch einen p-Wert unter 0,05 im Welch's t-Test verworfen wurde. Die beiden betreffenden Sensoren, Photoplethysmographie (PPG) und elektrodermale Aktivität, wurden anschlieSSend mit einem Sliding-Window-Ansatz implementiert, um die Echtzeit-Feedback-Erfassung zu realisieren. Die endgültige Implementierung von REALSTRESS war erfolgreich bei der Differenzierung zwischen den Phasen des Stresstestprotokolls innerhalb der Kontrollgruppe. Der durchschnittliche Stress-Score unterschied sich um 30% zwischen der Restphase und der Stressphase bei der arithmetischen Übung. Das PPG zeigte dabei den gröSSten Score-Unterschied von fast 50%.
Das vorgestellte Tool, REALSTRESS, hat das Potenzial, für die einfache Stresserkennung in Bildungs- und Forschungsbereichen eingesetzt zu werden. Der nächste Schritt für dieses System wäre die Entwicklung einer geeigneten Visualisierung der Stressskala.

Contents

Acronyms

ANS Autonomic Nervous System. 7, 13

CPT Cold Pressure Test. 11, 12

ECG Electrocardiogram/Electrocardiography. 6, 9, 13, 21, 24, 25, 27, 32, 33, 47, 48, 53

EDA Electrodermal Activity. xiii, 6–9, 13, 14, 20, 24, 26–29, 32, 35, 36, 40, 41, 47–49, 51, 53

EMG Electromyogram/Electromyography. 6, 9, 10, 13, 20, 21, 24, 25, 27, 32, 36, 47–49, 53, 54

GSR Galvanic Skin Response. 7, 13, 14, 28

HPA axis Hypothalamicpituitaryadrenal Axis. 7, 8, 11

HR Heart Rate. 6–9, 11, 13, 14, 20, 21, 28–30, 33, 40–42, 50–52

HRV Heart Rate Variability. 7, 11, 13, 14, 52

MAST Maastricht Acute Stress Test. 11, 12

PPG photoplethysmogram/photoplethysmography. 6, 9, 14, 20, 21, 24–28, 30, 32, 33, 36, 40, 47–53

PST Perceived Stress Scale. 12, 31

SCL Skin Conductance Level. 7, 8, 32, 33, 35, 40, 47, 51, 52

SCR Skin Conductance Response. 7–9, 11, 13, 29, 32, 33, 40–42, 47, 50–52

SNS Sympathetic Nervous System. 5, 7, 8

ST Skin Temperature. 10, 13, 14, 20, 27, 32, 47–49, 53, 54

TSST Trier Social Stress Test. 1, 10–12, 18

List of Tables

List of Figures

1. Introduction

The human body responds to both psychological and physiological stress in a specific way known as the human stress response. This mechanism is designed to protect the body and should release the body back into its initial state of rest once the stressor is dealt with. However, today the humans face many stressors that are more prominent, which leads to longer states of stress response in our bodies. Given that this inner system of self-protection is not designed for long-exposure, this causes a serious health concern.[30]
Psychological stress management and stress related illnesses have become a more discussed issue in recent scientific research. Their role in human health and wellness is more important than ever, and the early detection of stress related markers draws interests from many corners. In the last 10 years, many stress-related papers have been published, with the intend to find the optimal stress signal analysis [5] [52] [86] [61], produce the most efficient biosensor for stress measurements [39] [66] [95] , or design the least-biased procedure for stress-inducing experiments [1] [60].
Prominent biosensor producers in human stress research provide an overall solution, which includes read-to-use sensors and signal-processing software. Those kits are often expensive, and therefore not easily available for the general public or less-funded research-groups. Other scientists design their own sensors, which makes it even more difficult to gain access to them for related research. Most many sensor-tools presents are not customizable, which limits the possible use-cases for stress measurements. Therefore, many of their experiments are not easily-reproducible for the general scientific community. Secondly, almost all research done in terms of stress measurements focus more on in-depth signal extraction, therefore performing data-analysis after the experimental part is conducted. However, giving the user a live response to their stress level, for example in form of a stress level scale, could prove useful in educating the public about stress management techniques.
Producing a real-time stress feedback system with lower-cost sensors, which would make the finished product easily available for other research groups, has several challenges. Firstly, not all inexpensive sensors are able to present measurements of high-enough quality for live signal-processing. Secondly, designing a software environment for simultaneous signal-processing is challenging, because the different biosignal tools will interfere with each other.
This thesis ties in with the previous work done by aiming to produce a live-feedback mechanism for stress measurements with low-budget sensors named REALSTRESS. To implement this, a stress-inducing experiment based on the Trier Social Stress Test (TSST) will be performed using inexpensive biomedical sensors. This should provide insight in the usability of non-commercial, low-end sensors for human stress response detection. Based on the results of the first part, the aim is to implement a live-feedback stress

detector using a selection of the best performing low-cost sensors. This system should deliver information in form of a stress-meter with an appropriate accuracy for this type of educational material. REALSTRESS will be evaluated through several rounds of testing on a reference candidate, by adjusting the parameters of a sliding-window-based live signal-processing tool.
The signal processing steps for REALSTRESS are illustrated in figure 1.1, and include signal extraction, sliding window function and scoring functions. For the system to work properly, all of these steps have to be adapted to each other so that the interference between functions does not impair the final product.
The experimental results proved that two out of the five tested low-cost sensors could be used in classic stress detection tests. This was attested through the Welch's t-test, when comparing the signal means between different phases of the experimental stress test protocol. Furthermore, REALSTRESS could be implemented successfully, as was presented through the difference in mean stress scale between moments of stress and resting phases. For an arithmetic exercise, the control group showed a stress scale increase of 31% between rest and event. Considering only the heart rate stress scale, the increase was in order of 50%. This proves that the method of real-time stress feedback is realizable with low-cost sensors.

Following this introductory chapter, an overview of the most important theoretical background and related literature is given in chapter 2. The main contributions of this work are described in chapter 3, which includes the methods of sensor testing and the techniques for implementation of REALSTRESS. Chapter 4 presents the most important results of the testing described in chapter 3. The findings are further discussed in chapter 5. Finally, a conclusion over the work presented in this thesis is provided in chapter 6. Additional figures and material are presented in the appendix.

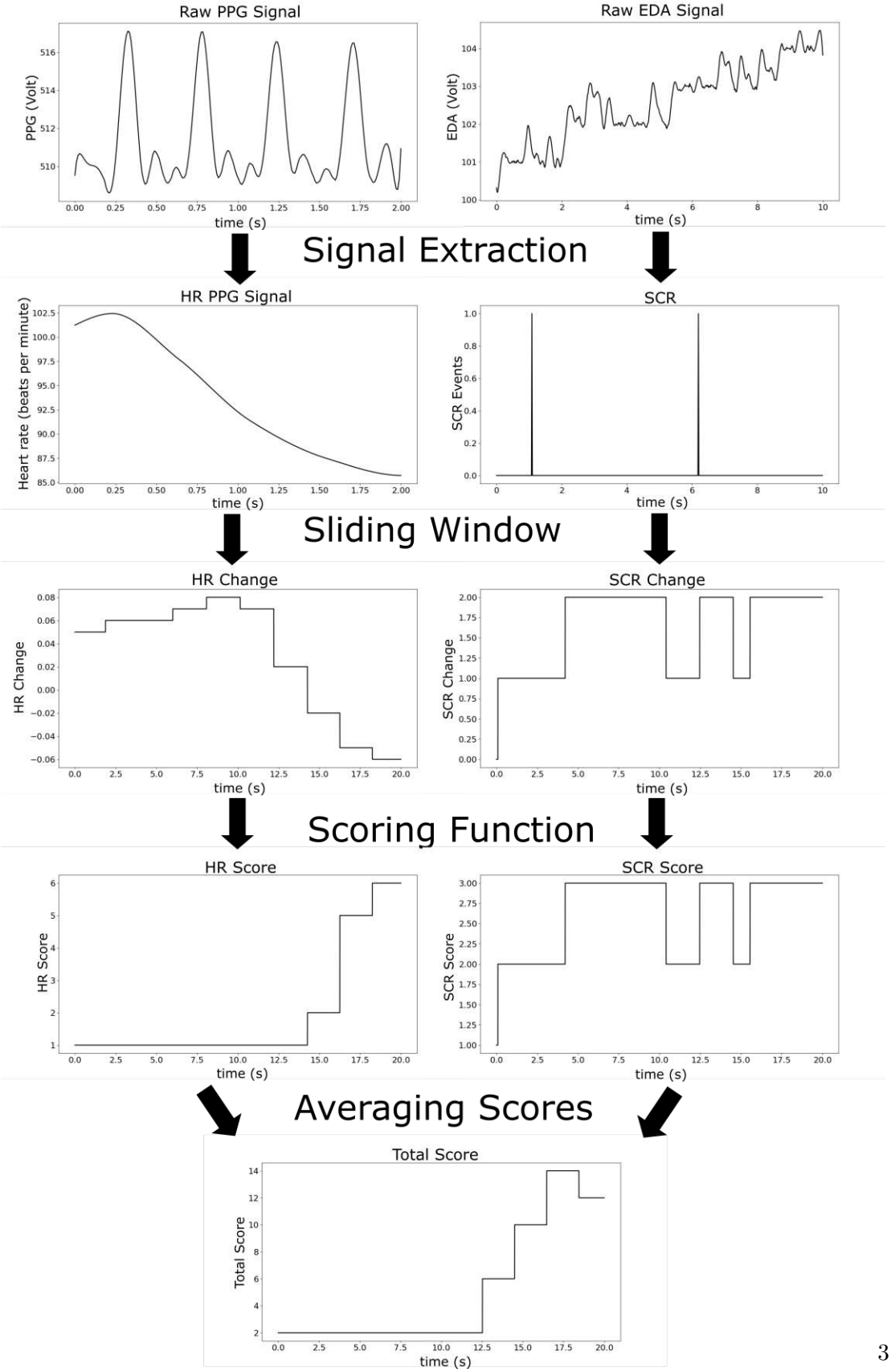

Figure 1.1.: Schematic description of signal processing steps for the implementation of REALSTRESS.

2. Theoretical Background and Literature Review

This chapter outlines the most important theoretical background for this thesis and introduces similar research done in the same field in recent years.

2.1. The Definition of Stress

In 1927, Hans Selye gave "they syndrome of just being sick" the name stress, in reference to the term *stress* in classical physics. Selye described stress as a response or physiological reaction, which is independent of its source. Furthermore, he named the stress trigger *stressor*. [30]
Stressors can be *psychosocial* or *biogenic*. Psychosocial stressors depend on the cognitive interpretation of a situation. What may cause some people psychosocial stress, may not be a stress trigger for others. Examples would be a traffic jam, grief or guilt. On the other hand, biogenic stressors are inherently triggers of stress and do not need cognitive appraisal, because they possess a sympathomimetic (Sympathetic Nervous System (SNS) activating) characteristic. Examples of biogenic stressors are coffee, cocaine, extreme heat and cold, and physical exercise. In general, psychosocial stressors are more common than biogenic stressors in affected patients.[30]
Stress is often viewed negatively, but can be beneficial to humans, as Selye described: *Eustress*, a quality of life improving positive stress arousal can be differentiated from *distress*, which has mostly negative consequences. When an individual is exposed to a stressor, first eustress is experienced. With increasing stress level, health and performance increase until they reach a maximum. When this peak is surpassed, a decline in both effects happens and eustress becomes distress. The optimal stress level, at which health and performance are at their maximal potential, is different for every individual and may depend on inherent and acquired properties.[30]
There is an distinction between stress response and target-organ effects. The psycho-physiological arousal of the individual is the stress response, while the effect of this arousal are pathologies or target-organ effects.[30]

2.2. Biosignals

Biosignals are the basis of the stress measurement research presented in this thesis. In this section, biosignals with the implications for this thesis are discussed.

2.2.1. Definition and Types of Biosignals

In context of biomedical research, physiological phenomena can be described by biosignals, which can be categorized into different classes: depending on their existence (induced or permanent), behavior (static or dynamic), origin (electrical, magnetic, mechanic, optic, acoustic, chemical, thermal, ...), among others.[48] In this subsection, a short overview about the most important types of biosignals for this thesis is given.

- Biosignals according to their *existence*:
  - **Permanent** biosignals have a signal source within the body, and do not depend on external excitation. One examples would be the Electrocardiogram/Electrocardiography (ECG) signal.
  - In contrast, when the signal has to be triggered artificially, it is an **induced** biosignals, whose duration is approximately the length of excitation. Optical oximetry (photoplethysmogram/photoplethysmography (PPG)), where light is used as a external signal source to determine local pulsatile blood volume, would be an example of induced biosignals.
- Biosignals classified by their *dynamic behavior*:
  - (Quasi) **Static** biosignals experience only little change over time and transfer information in a stationary state. While the core body temperature does change within its circadian rhythm, it does so in a slow manner over 24 hours, and is therefore considered a (quasi) static biosignal.
  - On the other side of the spectrum, Heart Rate (HR) changes with each beat and is an example of a **dynamic** biosignal. It is categorized by a high level of time domain changes.
- Finally, biosignals can be grouped by their *origin*. This label includes many different groups, therefore, this list will only describe those most relevant for this thesis. For further information about origins of biosignals, please refer to other literature (e.g., Biomedical Signals and Sensors I, E. Kaniusas, Springer 2012 [48]).
  - When the signal occurs due to **electrical activity** of muscles or neurons in the brain, they are called electrical biosignals. An example of this type would a Electromyogram/Electromyography (EMG) or ECG signal.
  - **Optical** biosignals are related to light phenomena such as absorption or scattering, such as the signal received in PPG.
  - The core body temperature changes (slowly) because the body losses or absorbs heat. This is an example of a **thermal** biosignal.

2.2.2. Heart rate and Electrodermal Activity

HR and EDA are prominently used terms within this thesis and are described in more detail in this section.

The amount of which the heart beats within a time-unit is defined as the HR. It is not constant and varies naturally, and this change in HR is defined as the Heart Rate Variability (HRV).[75] HRV happens due to the quick reaction to physiological state changes of the autonomic nervous system's regulatory mechanisms, which includes exercise and emotional activity.[49]
EDA refers to electrical potential changes of the skin. Furthermore, if no external current is applied during recordings of EDA, it is sometimes referred to as endosomatic EDA. Thereafter, the EDA signal can be parted into a tonic (Skin Conductance Level (SCL) and a phasic component (Skin Conductance Response (SCR) or Galvanic Skin Response (GSR)[27]).[21] While SCL shows only small variation, SCR is fast-changing and has been associated with Autonomic Nervous System (ANS) activity in particular of the SNS. High sympathetic activity leads to a decreased SCL signal as the skin's conductivity increases.[77] SCR is characterized by a quick rise towards the peak, followed by a creeping decrease of amplitude towards the baseline.[8]

2.2.3. The human stress response and related biosignals

This section gives an overview over the stress response of the human body and the related biosignals.
To begin with, the phenomena of homeostasis in the context of medicine has to be defined. It describes the regulating tendencies of an organism to preserve the internal environment's vital variables against non-equilibrium states. This is often achieved with negative feedback loops, where the start of one regulator system is sensed by another internal system, which works against it (usually with some delay).[6]
When a stressor is detected by the body, the system is no longer in homeostasis (or equilibrium), and the human stress response is induced with the goal to restore homeostasis.[69] In case of a psychological stressor, it must first be interpreted in order to develop from a simple stimuli to a stressor, which is called cognitive appraisal. This is followed by the neurological triggering mechanism, which initiates the multiaxial stress response. Finally, the actual stress response happens, which can be divided into the neural axes, the neuroendocrine axis and the endocrine axes.[29] More prominent is the division into the neuroendocrine axis (and here particularity the sympathoadrenal medullary (SAM) axis) and the Hypothalamicpituitaryadrenal Axis (HPA axis).[69] The neural axes are activated first, because their pure neural pathways are the quickest. Neurotransmitters, such as norepinephrine released by SNS pathways, are responsible for the change in organ-behavior.[29]
While the neural axes is the first to display a stress response, it does not lead to chronic stress arousal. Instead, the neuroendocrine axis is able to provide a stress response for a long period of time (chronic). The neuroendocrine axis is better known by the term *Fight-or-Flight Response*. The adrenal medulla is the main stress response organ in the neuroendocrine axis, which is a neural (through the autonomous nervous system) and endocrine axis. Upon neural activation, the adrenal medulla releases two hormones, the adrenal medullary catecholamines: norepinephrine (noradrenaline) and epinephrine

(adrenaline), which are the endocrine parts of the neuroendocrine axis. Out of the entire human medullary catecholamine activity, approximately 80% is due to epinephrine. The medullary catecholamines lead to effects functionally identical to those experienced after activation of the SNS stress response. However, the neuroendocrine stress response is delayed by 20-30 s. Some effects of the stimulation of the adrenal medullary axis are increased arterial blood pressure, increased HR and cardiac output, increased skeletal muscle stimulation, and decreased blood flow to skin. [29]
The endocrine axes need a more intense stimulation to trigger their response, but their somatic response is the most chronic and prolonged. Here, the adrenal cortical axis (also know as HPA axis) is of great prominence.[29] With the brain increasingly producing corticotropin-releasing hormone (CRH) and arginine vaseopressin (AVP), the pituitary glads are activated and start to secret adrenocorticotropic hormone (ACTH). The ACTH triggers the adrenal cortex to release corticosteroid hormones, which influence the behavior of almost every part of the human body.[89]
Table 2.1 presents a selection of biosignals that change their behavior due to the human stress response.[29]

Biosignal	Change in behavior	Origin
Arterial blood pressure	increase	adrenal medullary axis
HR and cardiac output	increase	adrenal medullary axis
Blood flow to peripheral organs	decrease	adrenal medullary axis
Blood flow to gastrointestinal system	decrease	adrenal medullary axis
Skeletal muscle stimulation	increase	adrenal medullary axis

Table 2.1.: Biosignal behavioral change due to human stress response.

2.2.4. Methods of measuring and processing biosignals

In this subsection, different methods used to measure and process biosignals are described. Since there is a wide variety of options, only those relevant for this thesis will be mentioned.

To measure EDA, two electrodes are placed on the hand (usually on the palm or on two fingers of the non-prominent hand) between which a weak electrical current is applied. Due to sympathetic arousal [78] the resistance through which the current must go, varies. This resistance can be measured indirectly using Ohm's Law by measuring the electrical potential difference between the two electrodes.[4]
The first step in processing raw EDA data, is extracting the phasic SCR and tonic SCL components of the signal.[20] This can be done with deconvolution techniques, for example using the non-negative deconvolution algorithm proposed in [9] or a standard deconvolution algorithm presented in [8]. The phasic signal is analyzed using latency, amplitude, rise time, and recovery time and has a typical form: A steep elevation followed by the peak and slow deterioration to the resting level. However, when several SCR peaks are following each other, the signals will overlap. For stress related research, SCR

activity is assessed through SCR amplitude and peak count detection. For analyses, it is important to discriminate between cases where the EDA is allowed to fully recover and cases where two or more SCR are superimposed. In case of the latter, signal processing has to be adjusted accordingly.[20] On way to achieved superimposed signal processing is proposed in [37], who was able to present an analysis technique, which was comparable to manual scoring methods. Tonic signal components are often not considered in stress related analyzes.[20]

HR can be determined using different measurement techniques, but the most common are ECG and PPG.[36] ECG measures the electrical excitation that origins from the sinoatrial node in the heart. When those electrical impulses travel through the heart to induce a heart beat, the changes in electrical signal can be measured using surface electrodes. The ECG records these biosignals at different parts of the body simultaneously.[34] The potential difference between two ECG electrodes is refereed to as one ECG lead, which displays the hearts electrical activity for one differential electrical axis. In general, more than one lead are used to improve the signal quality (better signal-to-noise ratio). There are different placement strategies for a various amount of leads used. For this project, whenever an ECG was placed, they were placed in accordance with the Triangle of Einthoven. The HR can be calculated using consecutive R-peaks in the electrocardiograph.[70]
The ECG signal processing consists of noise filtering, QRS detection, wave delineation, and data compression. Further processing steps might include feature extraction, clustering algorithms, and rhythm analysis among others.[83]

PPG is an optical biosensor that measures the change in microvascular tissue's blood volume. The system consists of two main components: a light emitting sensor source and a photodetector. The light source emits light in the range of red or near infrared.[3] As the light signal travels through the tissue it is attenuated differently depending on the blood volume at this tissue. With each pulse, the blood flow varies, which is visible in PPG as the AC component of the signal. The photodetector is used to measure the strength of signal in relation to the original signal emitted by the light source. Depending on the instrument, the signal is either detected as a reflection or transmission signal.[71] PPG is mainly used to detect the pulse rate and pulse rate variability (PRV).[58] Depending on where the PPG sensor is placed, the resulting signal will appear differently. The main features of the signal include a pulse peak and the dicrotic notch[47] and are usually present in all signals. Signal processing can be difficult, because movement artifacts remain a big issue in PPG data acquisition.[3] Signal processing usually consists of filtering, motion artifact removal, and pulse wave analysis. For real-time signal processing, FIR filters (e.g., Moving Average (MA) filter) are a suitable choice.[57]

The electrical excitation of muscles can be measured using a EMG, where the electrical potential difference between two electrodes is measured. Muscle activity can be determined through the amplitude peaks of the EMG signal.[48] Some recommended

EMG surface electrode placements can be looked up through the SENIAM (Surface ElectroMyoGraphy for the Non-Invasive Assessment of Muscles) project, which is a part of the EU program BIOMED II.[81]
The most important feature of the surface EMG signal is the amplitude since it indicates muscle force. Signal processing consists of noise reduction (whitening filter), signal filtering (forth order butter-worth filter), feature extraction through amplitude and power spectral analysis of the EMG signal can be used for feature extraction.[82]

Depending on where the sensor is applied, body temperature can be considered closer to core or shell temperature (Skin Temperature (ST)). While temperature readings near the axilla are closer to the core temperature, measurement on lower extremities, such as fingers or toes, are ST related. Furthermore, the body temperature depends on the environmental temperature due to the bodies heat regulation. The most common instruments to asses body temperature are infrared thermometer, thermistor, and thermocouple sensors.[25] Since body temperature is a quasi static biosignal [48], the sensors accuracy needs to be smaller than 1 °C. A thermistor sensor can be used to measure temperature due to its temperature-depending resistance change. Current calibrating methods allow for such precision to be reached using thermistors.[62][50] ST measurements can be processed using spectral analysis tools, if the data resolution allows it.[79] Usually, not more processing is done for this biosignal.[63]

Mechanical respiration signal acquired using a mechanorespirogram, provides a periodic wave as output signal. The signal can be analyzed considering its amplitude and shape.[47]

2.3. Stress test designs

Research in the field of stress measurements relies on certain testing strategies to produce reproducible results. This second provides an overview over the most common stress test designs used in research.

2.3.1. Trier Social Stress Test

In 1993, Kirschbaum et al. introduced one of the most popular experimental tool for psychobiological stress research: The TSST. It aims to induce a moderate amount of acute stress by exposing participants to psychosocial stressors. Although the exact testing protocol was altered in numerous ways throughout years of research, the test has three fundamental parts:

1. An anticipation period without an active stressor, which is needed to find the participants stress baseline. The usual duration is 15 minutes.
2. A public speaking task, most commonly a mock job interview, in front of a live audience. This part is sometimes filmed, to increase the pressure on the participant. This part is timed at 5 minutes.

3. An arithmetic task, which can be any moderately difficult mathematical exercise (e.g. counting downwards in steps of 13). It is participants are usually not aware that where will be an arithmetic task following the mock job interview. The time limit for this part of the test is 5 minutes.

Throughout all trials, participant will not receive any form of feedback from the research conductors.[2] After the arithmetic task, there is often an additional resting period.[53] The TSST will prominently be supported by the *State-Trait Anxiety Inventory* or similar psychological self-estimation forms, which are filled out by the participant itself before and after the TSST experimental task.[61]
It has been shown in various research that the HPA axis is activated as a result of the TSST. This is achieved through the combination of two main stressors: social-evaluative threat (interview and arithmetic task in front of audience, video and audio recording, no social feedback whatsoever) and loss of control (short preparation time, unannounced arithmetic task).[2]
The most common stress response biomarkers observed when performing TSST are cortisol and adrenocorticotropic hormone (ACTH), which are associated with the HPA axis. Other biomarkers that have been shown to be activated are HR, which peaks when stressors are introduced, and HRV. The use of the latter is still being debated in context of TSST. Similarly unclear is the role of blood pressure during the TSST. An increase in galvanic skin response has also been observed during the TSST [42].[1]

2.3.2. Cold pressor test

To test for hypertension in their patience, M.D. Hines and M.D. Brown came up with the Cold Pressure Test (CPT) in 1932. The test aimed to produce a universal stimuli to induce a consistent change in blood pressure. They figured out that patients experienced vasopressor effects (increase in blood pressure) after extremities were immersed in ice water. This thermosensory stimulus lead to the expected outcome for 99 % of their patients. The test starts with a period of rest (20 to 60 minutes) during which regular blood pressure readings are performed to find the signal's basal value. In the second phase, the subjects are asked to place an extremity into ice water (approximately 4 °C cold). Blood pressure is taken 30 and 60 seconds after the initial exposure to the stressor . When the extremity is no longer emerged into ice water, blood pressure is taken in regular intervals until it recovers back to basal value. They found that subjects with hypertension needed a longer period of time to return to basal blood pressure than subjects, who did not suffer from hypertension.[38] Furthermore, the CPT is frequently used in stress research [72], suggesting that a universal sympathetic response (in form of SCR can be provoked, while HPA axis responses (e.g., cortisol response) are not uniform.[22] Therefore, CPT is suggested to not be suitable for research with the aim to activate the HPA axis.[72]

T. Smeets and his research group combined the TSST and CPT with the aim to induces a response from both the sympathetic axes and HPA axis. They published their findings in a paper in 2012, in which they described their testing schema named Maastricht Acute

Stress Test (MAST). They found that the MAST outperformed the CPT when it comes to salivary cortisol response, but showed a similar response to the TSST. Their final conclusion highlights the MAST's ability to provoke a strong subjective, autonomic, and glucocorticoid stress response.[80]

2.3.3. Perceived Stress Scale

As described before, stressors are perceived differently between individuals. Therefore, it is not always clear which objective stressors will actually lead to stress-related illnesses. Furthermore, it is not possible to predict how individuals will react to stressors, because secondary factors, such as social support, may play an important role in stress coping mechanism. The Perceived Stress Scale (PST), as suggested by the team of S. Cohen in their paper of 1983 [26], was presented as a tool to measure a subjective stress level in relation to objective stressors, coping mechanism, and an subject's personality among others. The test is designed to ask the user simple questions which should be answered on a scale of zero to four depending on how stressful a situations was perceived by the individual. S. Cohen and team found the PST to be a reliable tool to measure perceived stress, especially compared to similar instruments used at that time. The research team suggested the PST to be used in research with aim to find connection between disease risk and stressors, and for measurements of experienced stress levels.

2.4. Electronic background

Due to their simplicity and affordability, micro controllers of the kind Arduino Uno are a popular choice for smaller electrical engineering projects. The company "Arduino" was founded by Massimo Banzi and David Cuartielles in 2005, and stands for affordable and easy to understand micro controller boards. The "Arduino Uno" was the original design, today, however, many more options of micro controllers are available.[51] The programming language used to write on the micro controllers was created by David Mellis [87] and is based on the already existing programming languages C and C++ [51].
The connections of sensors or actuators to the Arduino boards are established via pins. The Arduino Uno board uses an ATmega328 chip, and hosts 13 digital as well as 12 analog connection. All digital pins can be used for both input and output, while the analog pins are separated into pins which are exclusively used for either input or output. In this context, input refers to data that was collected by sensors and should be send forward, while output represents information send from the micro controller to an actuator to perform some form of action.[7]
The digital pins only recognize two states (HIGH or LOW), while the analog input pins will measure analogue input by its voltage level with a maximal resolution of 1024 states. Similar, the analog output pins will provide voltage output on 1024 different levels.[7]
The board is powered either via USB port or AC adapter. Additionally, the USB port allows to establish a connection between the board and a computer.[7]

2.5. Literature Review

In recent years there has been a lot of development in the field of human stress measurement with the use of biomedical sensors. Recent studies showed a combination of different sensors to achieve more accurate results.
Peripheral skin temperature is assumed to be a measure of physiological stress as it varies due to vascular system changes (vasodilatation and constriction), which in turn is correlated to sympathetic activity because it is controlled by the ANS) [31]. Therefore it has been part of recent research in terms of stress measurement, where skin temperature sensors have been used in combination with other biosensors [52] [67] [95] [88]. Skin conductance or galvanic skin response (GSR) is another popular biomarker for the human stress response and is often included in scientific research [43] [95] [60] [52] [5] [61] [67] [88]. HR and HR variability detected using the signal of ECG [60] [43] or PPG [95] [52] [5] [86] [61] [67] [88] sensors is used in modern stress research frequently. Other, less common sensors include EMG of the trapezius muscle [43], EEG [5], respiration rate sensors [67], sociometric sensors [61], IMU sensors [52] [43] [60], pulse oximeter [67] and blood pressure sensors [88].

In 2016 S. Yoon et al. [95] published a paper in which they introduced a wearable stress monitor in form of a small patch. The custom design allowed to measure skin temperature, skin conductance and pulswaves simultaneously with greater wearers comfort due to its small size and durability (life time estimated at 9 days). The patch was worn on the skin and consisted of several layers including a more flexible piezoelectric pulswave sensor compared to similar designs. A singular vector machine algorithm was used to identify four human emotions based on the physiological data.
In [60] a chest-worn Polar H7 HR monitor was used to observe participants stress levels in both a lab setting and a free-living field study. The sensor captured HR and R-R interval values, an additional wrist-worn sensor tracked activity data and ecological momentary assessment (EMA) prompts. The measurements were processed through feature computation. Both support vector machine (SVM) and random forest (RM) algorithms were able to detect windows of stress with a 87% accuracy in the lab setting. Furthermore, when they added a custom made GSR sensor to the lab setting, the accuracy reached 94%.
The researchers in [43] published a proof-of-concept stress monitoring equipment suitable for surgeons to wear in the operating room while performing procedures. The device recorded ECG, EMG, EDA and IMU sensor data and was positioned on the upper torso. This allowed for simultaneous monitoring of HRV and HR through the ECG, trapezius muscle contractions through the EMG, SCR through the EDA and respiration rate (RR). The device was tested in a lab setting and the recorded data was processed using MATLAB. The goal of the study, which was to record and store physiological data of several sensors simultaneously, was achieved.
K. Kyriakou et al. [52] published a rule-based algorithm which used GSR and ST to identify moments of stress (MOS) in participants. They used a wrist-worn commercially

available sensor (Empatica E4) to record the needed physiological data both in a lab setting and real-world field study. The rule-based algorithm was first developed using the lab setting recordings and afterwards used to classify the real-world data, where it achieved an average accuracy of 84%.
In [5] the human stress response was recorded during public speaking task with the use of commercially available electroencephalography (EEG), GSR and PPG sensors. Apart from frequency domain EEG features and time-domain GSR features, HR and HRV features were extracted from the PPG signal. The features were classified with machine learning algorithms, such as SVM and RF, to differentiate between situations of different stress level. With 96.25 %, the best accuracy was achieved by including features from all three sensors and feeding the feature vectors to an SVM classifier.
In a 2021 paper by A. Tazarv et al. [86] the authors measured HR and HRV using a commercially available Sportwatch equipped with a PPG sensor over a period of one to three months in a real-world study. The stress monitoring data was accompanied by self-report questionnaires and later processed using various machine learning approaches (e.g., SVM, k-Nearest Neighbors (kNN), random forest (RN)). The highest accuracy of 76 % was reached using a random forest algorithm.
M. Mozos et al. [61] combined physiological and sociometric sensors to measure stress during a lab setting experiment. The hand-and-wrist-worn physiological sensor collected EDA, HRV and PPG signal, while the sociometric sensor, which was worn as a conference badge was a IMU sensor with integrated speack recording function. The data was classified into stressful and non-stressful periods using the machine learning approaches SVM, AdaBoost and kNN. Using a combination of both sensor systems and classifying the data with Adaboost, the accuracy and precision rates of 0.94 were reached for this study.
In [67] measured physiological signals with an Arduino Uno board and related shield, designed by COOKING HACKS to detect biosignals. The research team selected four sensors for their study: ST sensor, GSR sensor, pulse oximeter, and breath-rate sensor. They classified the data using different classifiers and found that kNN provided the highest accuracy (95.98 %) with features of all four sensors were included.

In conclusion, there have been many research studies dedicated to measuring human stress response in laboratory or real-world scenarios. Most researchers focus on detecting relevant stress biomarkers using expensive sensor kits [52] [86] [5], some not widely available out of academic circles [61], or costume made sensor devices, which are not easily reproducible [95]. While some studies did use inexpensive, micro-controller compatible sensors [43] [67], some systems are no longer available for purchase [67].
Almost all stress research projects from recent years processed their data after the stressor experiment setting. There are almost no stress research projects that focus on direct-feedback loops in terms of stress detection. To achieve this effect, test subjects must see how their biosignals change due to stressful events or relaxation tasks.

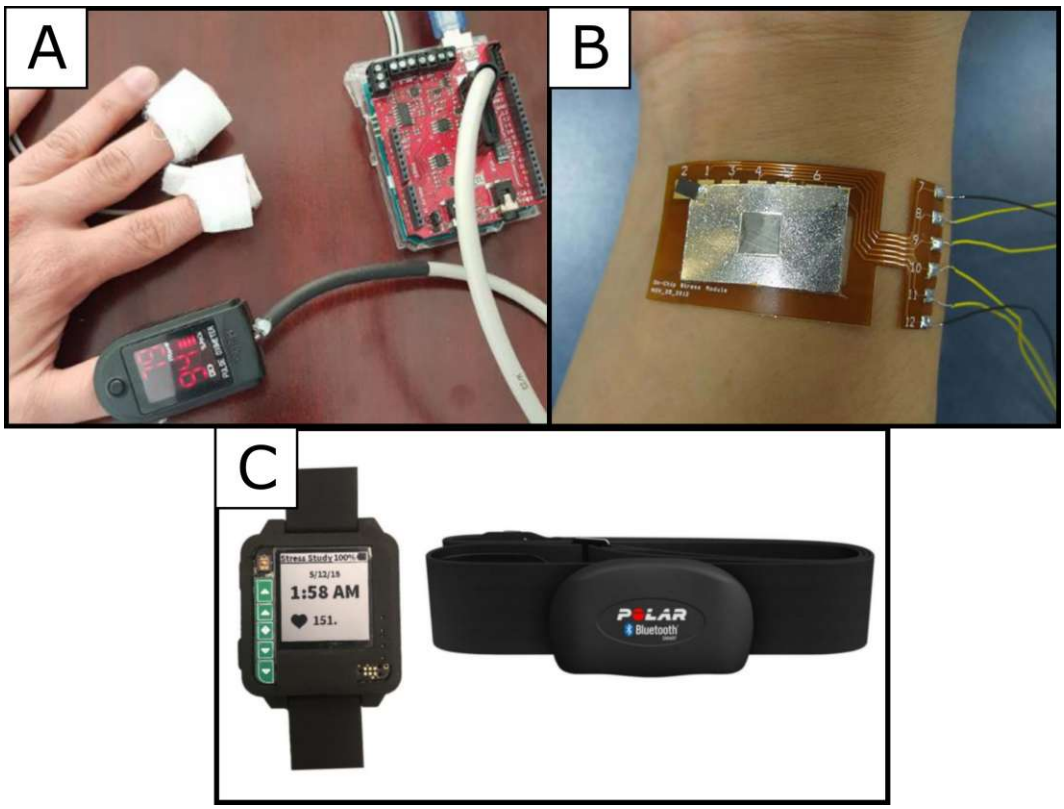

Figure 2.1.: Sensor systems used for stress response measurements in related papers. (A) GSR and oximeter sensor tested by J. Rodríguez-Arce et al. in [67]. (B) Stress sensing patch designed by S. Yoon et al. in [95], consisting of ST, GSR and pulsewave sensor. (C) Sensing devices used by V. Mishra et al. in [60] for continuous stress measurements. Image sources: (A) [67], (B) [95], (C) [60].

3. Experimental Testing of Low-Cost Sensors And Implementation of REALSTRESS

In this master's thesis two main research questions are aimed to be answered. Firstly, the question is raised whether inexpensive, broadly available biosensors are able to produce comparable results to related work in terms of stress detection. Following the first objective, the second research question asks, if it is possible to use such sensors for live-feedback sensor measurements in the form of stress detection. Both questions were explored through in-person experimental trials in which participant's biological signals were measured using several biomedical sensor devices during phases of varying subjective stress levels. This chapter provides an overview over the methodology utilized in this thesis.

3.1. Experimental Testing of Low-Cost Sensors

In the first section of this chapter, the methodology for answering the initial research question is presented. This includes the design of the experimental protocol, the sensor selection process, which signal processing steps were taken, and which statistical tools were used for data analysis.

3.1.1. Methods of Data Collection

To answer the first research question, several stress inducing tasks were performed on three days (20.07., 23.07., 24.07.) in July of 2022. Five young adult participants were recruited internally on voluntary basis to be part of the experiment. All participants were informed about the experimental procedure and agreed in written form that their data was to be analyzed and presented anonymously in form of this thesis. There was no discrimination regarding gender, age or medical history. The measurements took place over several days due to the participants schedule, but each participant performed the protocol alone (without other participants present) within an hour. The experiments took place in a laboratory setting, where the participants knew about the aim of the experiments. Biomedical data was collected during active tasks in which the participants had to perform some form of stressful exercise, as well as during resting phases before and after the tasks.

Experimental Protocol

The experimental protocol to answer the first research question was based of existing literacy in the field of human stress response. Based of the popular TSST, the main idea was to alternative between resting phases and moments of stress (MOS) while collecting biological data. Given that public speaking and arithmetic tasks were very commonly used in stress experiment protocols, they were included in the protocol for this part of the thesis. Public speaking was often realized in form of a presentation in front of viewers or a camera. However, to keep the task length short, the participants were simply asked to read a text out load. To increase the stress reaction, the text was not presented in their native language, but in a secondary language that they were still familiar with (English). The arithmetic task was taken strictly after universal formula using a subtraction task. Participants were asked to start again, if they made a mistake during the subtractions. This was done in accordance with comparable research papers.
For the final two tasks, new technology in form of a gaming console (Nintendo Switch Console) was introduced. The first task consisted of several rounds of find the error within three almost-identical images. This task could have been performed with pen and paper. However, by introducing the technology during the first new technology task, the participant were able to make themselves familiar with the system in a less rushed manner before presenting the second video game task. In addition, a more stimulating auditory and visual experience was presented. A visible countdown was shown during the entire task to introduce an additional element of stress for the user.
The final task of the protocol was performed with the same gaming console as the previous task. Participant was asked to reach the end of a platforming video game level, which challenged them with different obstacles. The level was beatable in five minutes for a trained player.
A visual overview over the protocol can be seen in figure 3.1.

Sensor Selection

Sensor selection for this part of the thesis was performed with regards to previous papers on the same topic. The selected sensors had to be affordable, well-adjustable for measurements, and quick and easy to apply and remove. However, the sensors were chosen due to their signals promising results in preceding research papers. The original sensor selection only included Arduino-compatible inexpensive sensors, but was extended by BITalino-based sensors, which were borrowed from the "Institut für Biomedizinische Elektronik" at Technical University of Vienna. Finally, a low-budget fitness tracker was included for blood pressure measurements. The Arduino-based sensors as well as the fitness tracker were provided by the Technical University of Vienna through the Automation and Control Institute (ACIN).
The first section describes the sensors that operated with an Arduino Uno R3 board. The sensors were used in accordance to their data sheet and through additional information given on the manufactures website. The sensor output was recorded using the software CoolTerm (Copyright (c) 2007-2021 Roger Meier).

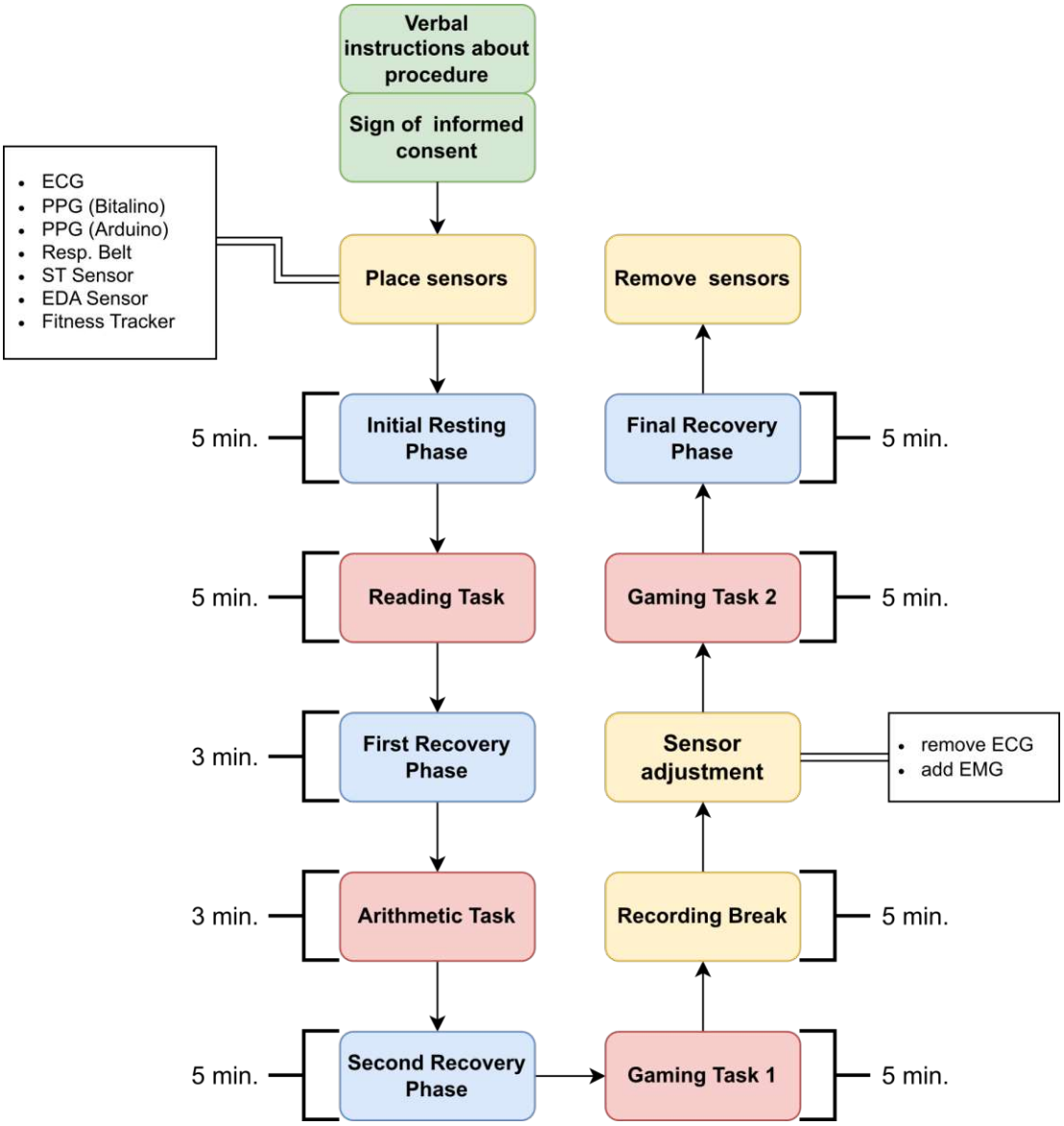

Figure 3.1.: Test procedure to answer the first research question.

1. One skin resistance sensor (EDA-sensor) by "SEEED" with grove male universal 4 pin connector and a LM324 operational amplifier. This sensor could be connected directly to an Arduino Uno board if equipped with a female grove 4 pin connector. Optionally, a grove shield could be purchased to connect the sensor to the micro controller board. The sensors are available in most online electronic stores for under ten euros. The grove shield was be purchased for five euros.[74]
   The sensor itself came with a data sheet of the operational amplifier and general instructions. Additionally, there was an online tutorial available on the sensor's company website.[74]
   The EDA sensor module consists of two nickle electrodes, which can be placed on fingers using included glove-like mounts. The electrodes measure electrical resistance between them. This signal is amplified using the op amp LM324, which is an operational amplifier for low input offset voltage (typical 3 mV) [40].
   The online guide on seeed studios website offers a tutorial on how to use the EDA sensor. It was recommended to calibrate the sensor to mid-range analog output voltage (512) before use and to average over ten data points each. Additionally, the tutorial provided an equation to translate the voltage output signal to human resistance.[74] Using this equation, the output signal for skin resistance measurements was in range of a few kΩ, which is consistent with similar measurements in literature. [33] [84] [44]

2. One HR sensor (HR-sensor) manufactured by the company "Frei" for less than three Euro per piece. The sensor did not come with a data sheet, but the retailer provided some basic information. According to their specifications, the HR-sensor could be used with an Arduino Uno board. The amplification factor was specified at 330, and the LED wavelength of the sensors was 609 nm.[10] By researching similar sensors, it can be assumed that this electrical component worked as a reflection PPG sensor. [59] [28] There were several Arduino libraries available for signal analysis and real-time HR calculation. One example of such a library was "PulseSensor Playground" by Joel Murphy, Yury Gitman, and Brad Needham, which was able to track live heart beats or calculate BPM, pulse transit time, among others. [92]

3. For measurements of ST the decision fell on "LM35 Precision Centigrade Temperature Sensor" produced by Texas Instruments. This sensor was chosen, because it is very inexpensive (under 5€ per piece), already calibrated, compatible with Arduino boards, and had a good accuracy (0.5 °C accuracy at room temperature) compared to sensors of similar price range. [41] Another advantage of this sensor was the possibility of placing it directly on the skin due to its compact size.

4. Another sensor by seeed studio was the grove EMG detector, compatible with Arduino boards, which was available for approximately 30€ [64]. Seeed studio provided a tutorial sheet with basic information, user guide, and example Arduino IDE code. The sensor data was provided via analog pins, with a maximum output voltage of 3.3 V. It used two OPA333 for zero drift amplification and one difference amplifier (INA331IDGKT). The product had three electrode connectors and came

with six single-use adhesive surface electrodes. The tutorial recommended for the output signal to be averaged over 32 data points each. [73]

The next section describes the sensors that were borrowed from the "Institut für Biomedizinische Elektronik" at Technical University of Vienna. These sensors were generally more expensive and were not compatible with original Arduino boards straight of the box, but can be connected to an Arduino device with special equipment.[55] Generally, they used their own environment "BITalino", and were designed for laboratory and general educational purposes.[16] To record the BITalino sensor output, a software created during a student project at TU Wien by Andreas Mayer and Klaus Zeiner was used. Optionally, BITalino provided their own software solution under the name "OpenSignals". [17] The following sensors from the BITalino environment were used:

1. The BITalino ECG sensor could be purchased as a bundle (HeartBIT) in combination with the PulseSensor UC-E6 (PPG) for around 240€. [18] The bundle provided a data sheet for the entire bundle as well as for each sensor individually. [12] The ECG sensor used three electrode connectors, and was suitable for chest or hand palm positioning with pregelled or dry electrodes. [13]

2. The BITalino Pulse Sensor, which was included in the HeartBIT bundle, came with both a velcro fastening strap for finger positioning and an ear lobe clip. The sensor operated as a reflective PPG device with a light signal at 520 nm wavelength. [14]

3. Lastly, the BITalino Piezo-Electric Respiration Sensor which could be placed at thorax- or abdomen-height to measure respiration through expansion of chest/abdomen-circumference. [15] It was available for purchase at 144€ per piece, and came with instructions and a data sheet. [19]

Lastly, for the non-continuous measurement of blood pressure, a fitness tracker (XD P330.741) was purchased for approximately 30€. This device was chosen due to its inexpensiveness and availability. The sensor was able to measure blood pressure as well as many other health related signals, including HR, SPO2, and caloric intake. The fitness tracker was compatible with most smartphones, however, it required a smartphone application. [65] A manual was included with purchase, which provided a user guide in several languages. It was not disclosed by the manual, how the blood pressure is measured by the sensor. [94] However, as the wrist-band of the fitness tracker gets tighter, when the blood pressure function is activated, it is likely that the tracker used some form of miniature inflatable cuff.
The final sensor selection includes eight sensors from which seven were used for measurements simultaneously at a time. The sensors were placed in accordance with figure 3.2. Gima 33371 Universal Electrode Ekg/Diameter 48 x 50 mm were used for both ECG and EMG. The electronic architecture of the experimental setup can seen in figure 3.4.

3.1.2. Signal Processing

The raw signal data is often not conclusive, whereby signal processing has to be performed after signal acquisition. This post processing differs between the various biosignals and is

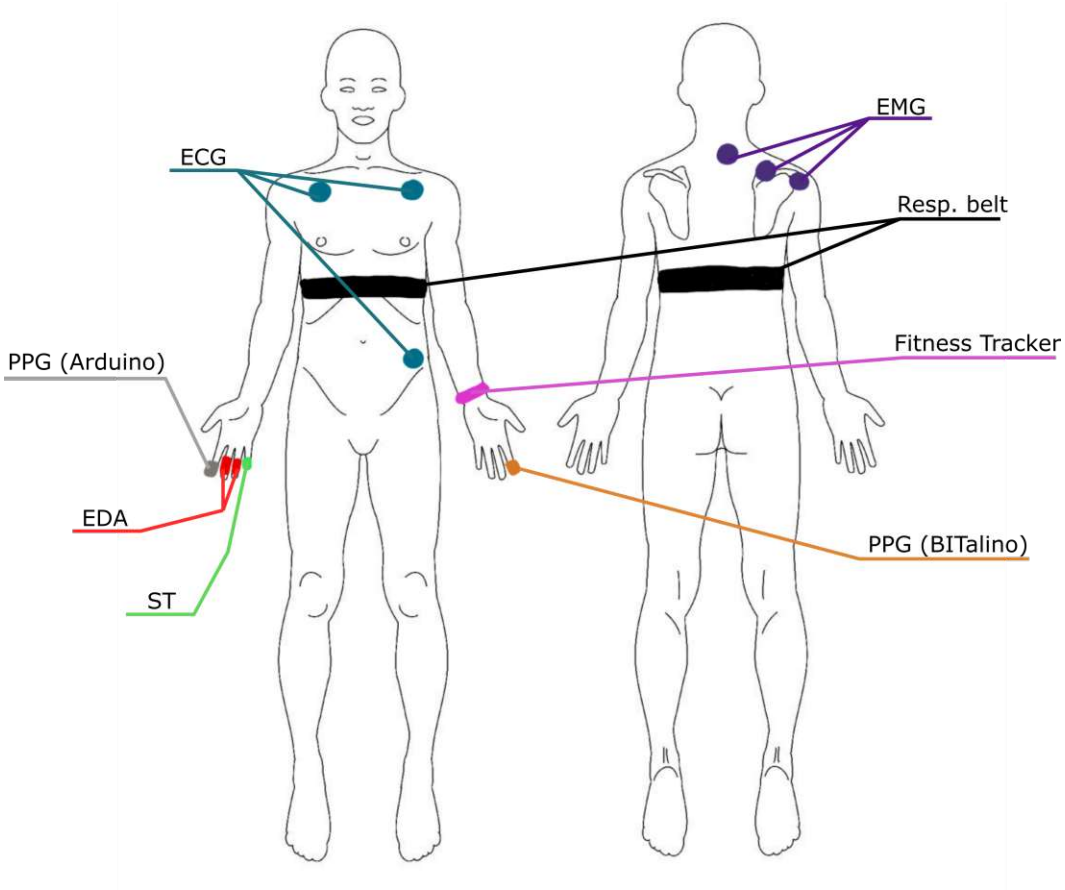

Figure 3.2.: Schematic placement for all used sensors for the experimental protocol of the first research question.

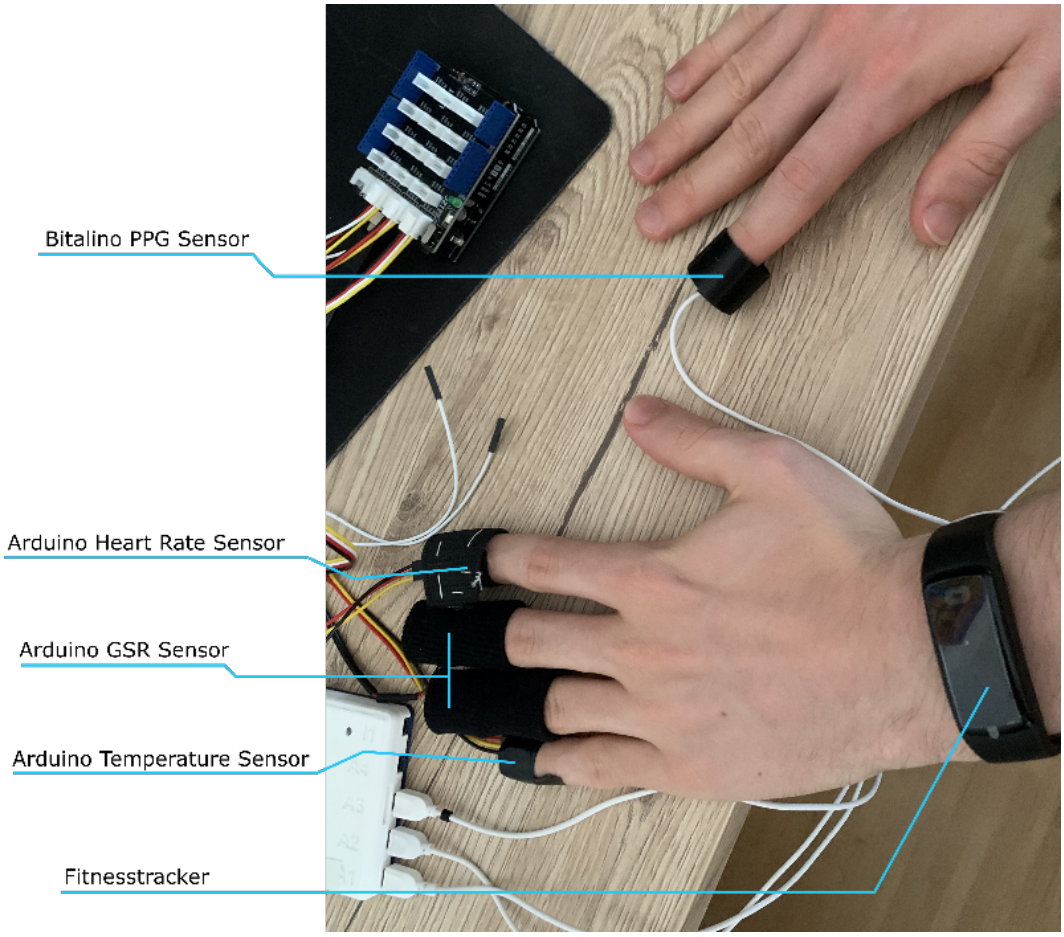

Figure 3.3.: Part of the biomedical sensors placed on upper extremities during the experiment.

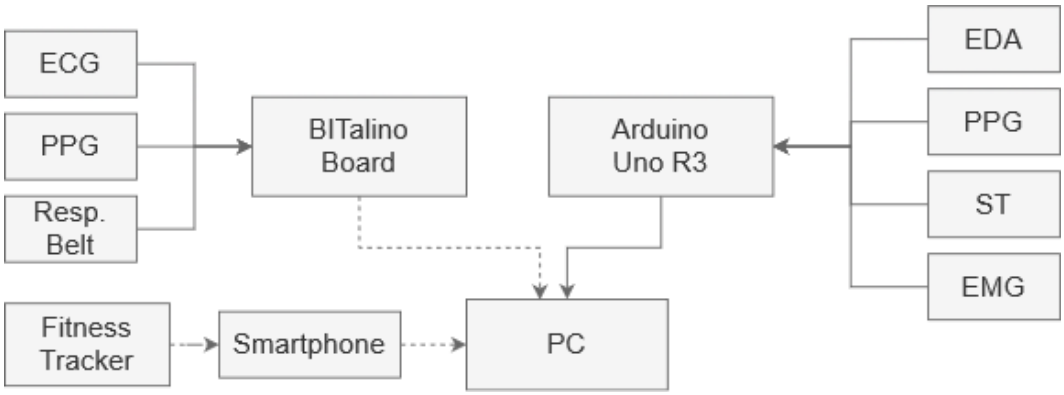

Figure 3.4.: Schematic placement for all used sensors for the experimental protocol of the first research question. Dotted lines present wireless connections.

described in this section.
Before the data was further processed, recording artifacts were removed, and the data was split by sensor type. The measurements were labeled in their time-axis to assign the corresponding protocol phases to the data. The labeled and grouped data was processed using several signal acquisition tools to extract biomedical signals.

The main computational libraries used that provide tools for bio-signal processing are BioSPPY [24] and Neuorki2 [56]. Both provide signal processing tools for different biomedical sensors, including EDA, ECG, PPG, and EMG. Due to their similarities (some Neurokit2 features are based on BioSPPy), both libraries were used for signal processing in this thesis. The results of both tools were than compared, and data extracted by the better suited library was further statistically analyzed.
When signal extraction through BioSPPy and Neurokit2 was not possible, the sensor data was processed using the methods described in 2.2.4. For EMG this included pre-processing through a fourth order butterworth filter and by-hand peak amplitude marking.
For signals were a count or change of signal feature was of interest, a sliding window approach was performed either on the raw data or on top of the already processed data through BioSPPy or Neurokit2.
After signals were extracted, the processed data was normalized and outliers were removed. Both operations were performed in accordance with data processing described in [60], where outliers trimming and z-score normalization on the data presented the best final results. Finally, the processed data of all participants was sampled and added to a big data frame for statistical analysis.
All processing steps were performed using python and various python libraries. Sampling was performed randomly, but each participant added the same number of data points to the final data frame.

3.1.3. Methods of Analysis

The processed and sampled data was statistically analyzed using the Welch's t-test in accordance with [60] and [5]. This t-test was preferred over Wilcoxon-Mann-Whitney test, because the former is recommended for larger sample sizes and is still robust for heavily skewed data [32]. Pair plots and feature importance extraction using extra tree classifier were used as additional statistical tools to analyze the data. The methods were performed with python and python libraries such as scipy.stats, pandas and seaborn.

3.1.4. Summary and retrospective evaluation of method

The protocol was designed to be a mixture of established stress inducing exercises and alternative tasks using new technology. Additionally, the latter two tasks were designed to be mostly non-verbal, because they required less active communication by the participant. This presented an opportunity to compare biosignals which would normally be effected by changes in breathing patterns, such as talking. The most prominent example for such a sensor is the respiration belt. Given that video games are very present in the modern

society, but to my knowledge are still not included in stress assessment studies, this first approach could provide more insight into possible future stress test designs. However, by keeping the well-established design of TSST for half of the experimental protocol, it was ensured that the results were comparable to previous work. One obvious downside of the video game tasks was the use of a hand-held controller to perform them. Given that many sensors were placed on the participants hands, these two tasks were more prone to movement artifacts. Nonetheless, since the fingers used to press the required buttons on the controller were sensor-free, the artifacts should not be deal-breaking.
Another popular stress inducing task, the cold pressure test, was not considered for this protocol, because the time of felt comfort, which in turn produces a measurable human stress response, may vary between participants. This would make the results between participants less comparable.
The sensor selection showed a good variety in methods and biomedical signal. With the PPG sensors, it was even possible to compare two sensors of the same type in different price ranges directly. The only sensor that did not meet expectations was the fitness tracker for blood pressure measurements. Since it was not possible to perform continuous measurements, and because the technical information presented with the tracker was sparse, the sensor proved to not be a good fit for such an experiment. But even with exclusion of the fitness tracker, it was possible to use seven sensors (six simultaneously) to measure biomedical signals during stress sensing, which was high compared to related work, where the focus was often set on fewer selected sensor types. This allowed for a new insight in multi-modal sensor measurements using micro controller based sensors.
In the development phase of the experimental protocol, the decision was made to not measure ECG and EMG simultaneously, because the electrode positioning of both sensors was in close range. Therefore, to not introduce any interference between the two signals, the ECG was measured during the first three tasks, while the EMG was only active for the last task and the final recovery phase. In retrospect, this step was unnecessary, and should not be repeated in future protocol designs. This decision reduced time of data collection for EMG considerably, making it less intuitive to compare the results.
In contrast to much of the related work, the processed data was not processed to create an algorithm for stress detection using machine-learning. Instead, the data was compared statistically, to find which sensors produced signals of well-enough quality to show a significant differences between the phases fo the experiment.

3.2. Implementation of REALSTRESS

The second section of this chapter describes the workflow that was performed to answer the second research question. It starts by describing how the experimental sessions were conducted and how sensor selection was finalized based on the results of the first part. Afterwards, signal processing and data analysis methods are presented.

3.2.1. Methods of Data Collection

The second research question required more fine-tuning of coding resources and parameters to be answered. Therefore, the method differed from the first question by having to perform many small experimental tests compared to one extensive procedure. The testing for the system design for this research question took place over three months. Most of the testing was performed on one reference candidate, because many tests had to be repeated several times under the same conditions. After the implementation of the research solution was finished, and tuned to the best of the sensors ability, a final experimental testing round was performed on seven participants. These last measurements took place on the 10.03.2023 in a laboratory environment. The participants were chosen internally beforehand, without discrimination due to age, gender or medical history. Before the experimental part, all participants were informed about the experimental tasks expected from them and signed a form of consent so that the measured data could be processed and presented anonymously as part of this thesis.

Experimental Protocol

During the implementation of the live feedback system, testing was performed on the same candidate for all rounds of experiments. To make the tests reproducible, the method of stress induction was chosen to be different breathing patterns. In more detail, the protocol consisted of a sequence of relaxed breathing and breath-holding. It has been suggested that breath holding produces an autonomic response [85] [11], which is comparable to the stress response induced by stressors. This method of testing should insure an autonomic response on the same candidate even after several rounds of testing, which was needed for this protocol.
In the final testing round of the implementation, a cold pressure test was performed in several rounds on the same candidate to test the system on a proven stress inducing task. For the testing with participants after the implementation was finished, two tasks were considered. First, the same type of breathing pattern test performed during implementation was tested on all participants. The second test was the same arithmetic task performed in the protocol of the experimental part of the first research question. Before and after each task, a resting or recovery phase was included. Both tasks were done back-on-back, with a short recovery phase between them. Each task was performed in four minutes, which adds the total measuring time for each participant up to 10 minutes including sensor placing, removing, and verbal information.

Sensor Selection

A few things had to be adjusted when working with the biosensors for a live-feedback system. Firstly, the sensor selection was cut down, since it needed to meet certain prerequisites. These include wear-ability, availability, and how promising the sensor-results were. This left only two sensors: the PPG Arduino sensor and the EDA sensor.

Even though most sensors performed well enough to be considered for the project, not all sensors are practical for use. This included two factors: First, is the sensor easy and fast to apply and remove from the participant= Secondly, is the post-processing possible in real-time feedback loop measurements?
The first condition was not met by the ECG and EMG sensors tested, since they relied on single-use electrodes. Therefore, it would not have been feasible to use these sensors in the environment of project. All the other sensors were easy to apply, and could be disinfected between use, without having to buy a new piece of equipment for each participant.
The post-processing is a bit more difficult. For once, some sensor did not output their data directly to the main secondary device used for processing, which disqualified them from the project. This was the case for the fitness tracker. Other sensors did not have data of good enough quality to post-process them automatically, which included the EMG sensor.
Keeping the two conditions in mind, the remaining sensors were the two PPG sensors, the respiration belt, the temperature sensor, and the EDA sensor. One of the PPG sensors and the respiration belt were both produced by Bitalino and disqualify due to their price and availability. The temperature sensor measured a slow-changing biosignal (ST), which was not suitable for a live feedback system. Therefore, the temperature sensor was not considered.
Finally, this left the two sensors: the EDA sensor and the PPG Arduino sensor. They were chosen due to their promising results, availability, inexpensiveness, practicality, and post-processing possibilities. For the given constraints, these sensors were expected to provide the wanted results, and were very customizable due to their compatibility with the Arduino system which itself allows for high customizability.

3.2.2. Signal Processing

The software and code to work with the sensors had to be adjusted from the version created for the first research question, since the testing in July was based on recording the data separately from processing it. A live processing is needed for the interactive wall, therefore, several Arduino libraries were considered. In more detail, the following libraries and code-resources were tested for their abilities:

- PulseSensorPlayground.h [93]: This library offered a collection of code projects by World Famous Electronics Ilc. for pulse sensors (PPG sensor) in the Arduino environment. The first version of the library was launched in 2018 and is still updated as of now (February 2023). The library came with many build-in example programs, which allowed for quick and easy work with pulse sensors. The threshold for the peak detection of the pulse signal was customizable, and there were example projects to find the ideal threshold for the sensor used. Furthermore, beside the option to calculate BPM (beats per minute), the library had many other features such as providing projects to calculate the PTT (pulse transit time) or connecting two pulse sensors at once.

- BioData [35]: Erin Gee owns the copyright to this Arduino library which was first released in 2018. It allowed for the interpretation of sensor data which includes pulse signal, respiration signal, and GSR. In contrast to PulseSensorPlayground.h, the code for the pulse sensor did not provide an easy way to change the threshold for the peak detection. However, the sampling rate of the sensor was adjustable by the user. The same was true for the provided EDA code. The library was more limited concerning the project examples compared to PulseSensorPlayground.h, even though it presents options for various sensor types. The library was last updated in September of 2022 (as of February 2023).

- engineersgarage.com [23]: In this article about stress level measurements using EDA, U. Butt uses the same Arduino based EDA sensor as this project. The article provides their example code which detects changes in signal using a self-adjusting threshold function. The strength of change qualifying for an alert can be adjusted. There is no information about when the article was published.

Each selected library and code resource was tested in a simple Arduino IDE environment for stability and reliability during resting phase and with breathing patterns. This allowed to filter out the appropriate material for further implementation.
After the initial testing, the system had to be implemented in a coding environment for later use. The system was first created by students of TU Wien during an outreach project at ACIN the group participated in for a educational material named *interactive wall*. The system is based on i2c connections between several Arduino Uno boards. Both sensor were placed on one microcontroller board for simultaneous sensor measurements to save equipment and simplify the project.
After integration of the sensors in the new environment, their output was compared between single measurements (sensor alone on one device Arduino) and dual measurements (both biosensors on one device Arduino). This procedure was chosen, because the PPG best performing processing library (PulseSensor Playground) used an interrupt pin, which was subjected to be interfering with other coding libraries when used simultaneously with them. Furthermore, both biomedical sensors were connected via analog pins on the Arduino board, and analog inputs were known to influence each other. To avoid the latter issue, for each sensor the analog input was measured twice, and the first input was neglected.
To avoid the issue of the interrupt function interfering with other libraries, two solutions were considered: Either using two different device Arduinos so that the sensors analog input cannot influence each other and the interrupt function can be used for the HR calculations. The second possibility was to use the HR function without interrupt, but have both sensors on just one device Arduino. The second solution was chosen to be further executed, to limit the amount of needed hardware for this project.
To allow for continuous signal processing with live feedback, a sliding window function was implemented for both sensors. Given that the final implementation should not rely on a designated resting phase to be recorded first for comparison, the windowed data needed to speak for itself. To achieve this, both signals needed to present a change

in signal rather than an absolute value. BioData's SCR function was able to present fast-changing EDA signal during live measurements. Using the sliding window approach and the SCR function, the amount of SCR changes per window could be presented in real time. Different window sizes were compared on several rounds of breathing pattern tests to find initial parameters for the sliding window and SCR functions.
Next, PulseSensorPlayground's non-interrupt HR function was implemented with a sliding window approach. Analog to the EDA sensor, the HR change per window was presented instead of the sole HR. The parameter selection was done through dual measurements of both HR and SCR. The testing was performed using breathing pattern tests with different window sizes for the HR change function.
After the testing for the two chosen sensor functions, a scoring function was implemented on grounds of the results of the previous tests. The general idea was for both sensors to provide a stress scale value from one to ten, which would be averaged for the final stress scale value. Given that the SCR changes per window size did almost never exceed ten, the only adjustment for the scoring function of the EDA signal that had to be done was a cap of ten.
The HR change testing results showed step drops after breath holding, which could be specific to that event. Therefore, a (stressful) event could be detected with decrease in HR. For the breathing pattern exercise, this detection was trigger after the event has already occurred, which made the response delayed. Nevertheless, it still provided the possibility to give direct feedback to the user in terms of receiving a response to a change in breathing within the testing session. The decrease in HR change was mapped on a scale from one to ten by comparing with the results presented during the rounds of testing.
To find the best combination of window sizes for both sensors, numerous parameters were tested in pairs during breathing patterns. The best combined results were chosen for the final implementation.

3.2.3. Methods of Analysis

After implementation, the system was tested on the candidate that was used for testing throughout all previous implementation exercises performed using breathing patterns and a cold-pressure test. For each test, several repetitions were performed. To check for over-fitting, the implementation was tested on different participants both for breathing patterns and an arithmetic task.
The tests done during the implementation process were assessed visual to see if there was a noticeable difference in stress score between resting and event phases. For the final implementation tests, the collected data was split into signals and phases, and measurement artifacts were removed. The data was sampled and summarized in one big data frame, which was used for statistical analysis.
Each phase was analyzed using the samples mean, variance and median. The stress score needs to show a visible difference for the user using the function. Therefore, no further statistical tests were performed, because a statistical significant difference presented through tests such as the Welsh test are not always visible by looking at the data directly.

Mean, median and variance provide a more intuitive answer to the research question as they more closely present the result that a user would see.
All analysis steps were formed in python using the libraries scipy.stats, numpy and pandas.

3.2.4. Summary and retrospective evaluation of method

The methods for the second research question differed significantly from those presented for the first. This had to do with the workflow needed for both questions, which was in a way opposite. While the first question needed a lot of data to test and analyze the sensors for their potential, this job was already done for the second question. The sensors were ready to be used and had already proven themselves in the first experimental part of this thesis. The second research question required way more fine-tuning to answer. This was done through repetitions of the same tests with slightly changed parameters. Both sensors now needed to present their signals simultaneously in a way which was intuitive for a future user. Only doing one round of tine-tuning with one larger sized experimental session would have been more of a guessing game. By improving the code step-by-step, with many repetitions, the component of randomness was reduced.
The way the workflow was created, by testing only one candidate for most of the testing rounds, the risk of over-fitting the data was high. In future implementations of this kind, it would be practical to have a small number of participants that can be used for several testing rounds. However, this is not always feasible. The biggest issue of over-fitting was with the HR function presented through PulseSensorPlayground.h, because the function required to set a threshold value manual, which varies naturally between individuals. However, the problem of the HR threshold could be avoided when choosing an appropriate PPG processing library with autonomous threshold finding algorithm. Such a library was not found while working on the implementation for this thesis.
Additionally, testing with one candidate numerous times in short time frames meant that many classic stress test designs were not suitable as they would fail to produce the same response after a few repeated session. The breathing pattern test provided the possibility achieve reproducibility with minimal effort.
This method provides a new perspective in human stress response. Because the window-wise data is processed immediately, the method allowed for live feedback in a way that was not presented to my knowledge. Even though some biomedical sensor manufacturer provide some form of live processing using their software, those solutions are most often hidden behind a larger pay-wall. This method provided a system that is simple and inexpensive to recreate, and which can be extended easily in future research.

4. Results

This chapter presents the most important results found during the testing phases to answer both research questions of this thesis. The first section concerns itself with the results of the stress test performed in July of 2022 using low-cost sensors. The findings of the testing done when using REALSTRESS, the real-time stress scale, are presented in the second part of this chapter.

4.1. Experimental Testing of Low-Cost Sensors

During the experimental test, the participants were asked to rate their the subjective stress level (similar to PST) after each task. The self-assessment of personal stress level during the four stress tasks made by the participants from 1 to 10, where 1 is Not stressed and 10 is Very stressed is also visible in table 4.1. When compared to the two traditional stress test tasks (task 1 and 2), the tasks using new technology (task 3 and 4) were overall received on the same level of stressful as the traditional versions.

The data collected during the experimental protocol in July of 2022, was analyzed using the Welch's t-test for statistically difference between the different phases of protocol, with the hypothesis that the mean of distribution of both compared phases is equal. The null hypothesis is rejected for a p-value smaller than 0.05, concluding that there was a significant difference between the means with 95% confidence. The statistical test should answer the research question, if inexpensive sensors are usable for stress detection tasks, by providing insight in what sensors were able to record data of well enough quality to be discriminate in protocol phases. Therefore, the biomedical signals recorded with given sensors were the predictor variables, whereas the different phases were the outcome variables.

The Welch test was performed with and without discrimination between the different tasks. For both analysis the task data was compared against the data of the initial

Table 4.1.: Personal stress level evaluation.

Personal stress level during each task					
	P1	P2	P3	P4	P5
Task 1	4	5	2	1	7
Task 2	6	6	8	4	5
Task 3	5	6	5	2	3
Task 4	5	6	10	3	6

resting phase in the beginning of the protocol. Given that some sensors were not active through the entire protocol (ECG, EMG), those signals are excluded from statistical results presented for all tasks as one.
Considering only the difference between summed stress inducing task phase and resting phase, a significant difference between the outcome variables could be found with all predictors but ST. The t-statistic showed that the signals extracted from the EDA sensor (SCL an SCR) decreased in task phases compared to the resting state, while cardiovascular and respiration signals increased with increasing stress induction. These results are summarized in table 4.2. The results of the Welch test for each task individual

Table 4.2.: Significant Differences between Resting and Testing Phases for Selected Features (Welch statistical t-test). The window size for the SCR was 5 seconds, and for the HR change was 40 seconds.

Features	***t-Stat***	***p-Value***
SCL	-27.020	<0.001
SCR	-2.201	0.0287
PPG (Bit) HR	5.256	<0.001
PPG (Ard) HR Change	4.909	<0.001
PPG (Ard) HR	4.909	<0.001
Respiration rate	2.150	0.033
Temperature	**0.031**	**0.975**

compared to the resting phase are shown in table 4.3. The first two tasks, representing the classic stress testing design, could be statistically distinguished from the resting phase in all but one instance, in which SCR failed to divide between the phases for the public speaking task. The behavior of the data during the established tasks showed an increase in signal value for cardiovascular, respiration and ST signals during stress phases, but an decrease for the EDA features. The two video game tasks resulted in a different cardiovascular signal behavior compared to the established tasks, while still being distinguishable from the resting phase. Most prominent, heart rates extracted from the two PPG sensors showed a negative t-static for one of the two technology tasks each. Furthermore, for the Arduino-based PPG sensor, the extracted heart rate had a twice as high t-static for the final task compared to the first two tasks. Similarly but different, the BITalino based PPG sensor showed a decline in heart rate compared to the resting phase during the last task.
Considering all four tasks, SCL decreased in value over time, which can be seen in the decrease of t-statistic with each task. The respiration rate increased slightly with each task performed. ST showed good distinguishability with low p-values and relatively high

t-statistic for the first three tasks, but failed for the final task. Heart rates extracted from both BITalino systems (ECG and PPG) showed similar t-statistics and low p-values, while the Arduino-based PPG sensor provided a HR signal which varied heavily in t-statistic from the previous two.
The pair plot 4.1 shows the strength of feature combinations clustering the phases in task phase and resting phase. The pair plot for separated tasks can be found in the appendix in figure A.3. Both figures support the results presented using the Welch's t-test. The best clustering could be achieved for feature pairs that included SCL or SCR as a feature. Other feature combinations performed worst in clustering the data into two phases. The success of SCL in behavior is visible in the extra tree classifier results in figure 4.2 as well, which presents the most important features for clustering the data in the labeled phases. The classifier determined SCL and HR change as the best features for clustering this data set, with SCL showing importance of almost 50% within all features. Both HR change and SCR needed a sliding window approach to be determined for the data set. Different window sizes for both signals were compared in their ability to produce results that can be used to distinguish between the experimental phases. The results of the corresponding Welch's t-test for each window size can be seen in table 4.4. The subsequent signal for all window sizes tested for the HR change showed significant differences between the phases with the exception of the first video game task. Opposite results were seen for the SCR window sizes, that all failed to distinguish between the the resting phase and the first and last task respectively. The first video gaming task could be differentiated from the resting phase for all SCR window sizes. The arithmetic tasks was able to produce a SCR signal that could be distinguished for almost all window sizes for both signals.

The comparison of normalized heart rate produced through data of three different sensors (ECG, PPG (Arduino) and PPG (BITalino) for four participants during resting phase and task 1 can be seen in table 4.5. For participant 1 the mean normalized heart rate for both phases showed agreement between ECG and PPG (Arduino), but differed noticeably from PPG (BITalino). For the other participants, the HR measures of all three sensor types was comparable, apart from some divergence during the task phase of participant 3.

Neurokit2 vs. BioSPPy

Signal processing was done using both neurokit2 and bioSPPy. This allowed for a direct comparison of the two signal processing tools, since they were given the exact same data. Several observations could be made about the two tools:

- The processed output data size was lower compared to the original data size for BioSPPy, while neurokit2 outputed data in size of the input. For example, the data output for the HR extracted from ECG for the first task period counted 445 entries for BioSPPy and 280000 entries for neurokit2, which is the size of the unprocessed input data.
- The first couple hundred entries of the output data of neurokit2 are identical.

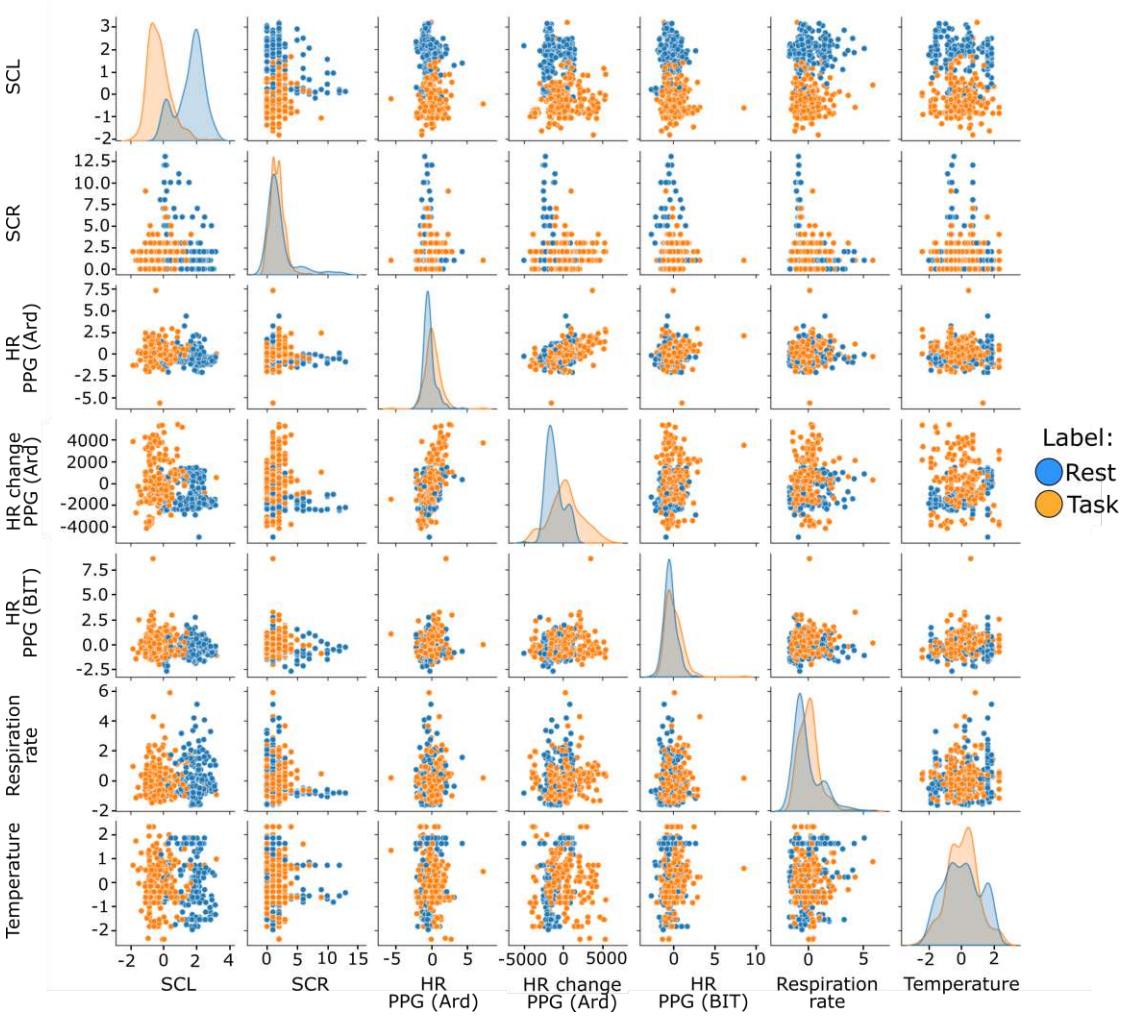

Figure 4.1.: Pair-Plot of Selected Features for all Task Phases together in comparison to the initial Resting Phase. The window size for the SCR is 10 seconds, and for the HR change is 40 seconds.

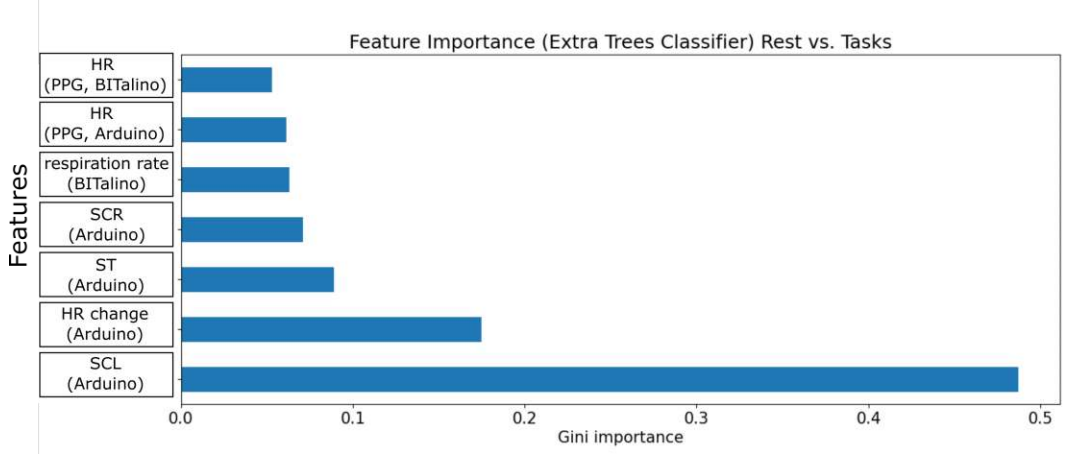

Figure 4.2.: Most important features of feature selection determined through extra-trees classifier for all Task Phases together in comparison to the initial Resting Phase. The window size for the SCR is 10 seconds, and for the HR change is 40 seconds.

Overall, repetitions in signal are more common for neurokit2 processed data, while BioSPPy mostly provides data output that varies within entries.

- BioSPPy tools performed tasks faster than neurokit2. For example for ECG processing, BioSPPy finished processing all four datasets (Rest 1, Task 1 - 3) in 1.27 seconds, while neurokit2 needed 22.8 seconds for the same task.
- Both tools delivered similar results concerning the extracted data of most sensors (one exception is peak finding of EDA signal).
- The peak finding algorithm of Neurokit2 is more accurate in its results compared to BioSPPy. An example of the difference in peak finding between the two python tools can be seen in figure A.1.

Additional remarks to the results presented above:

- All portrait results that include data that was processed with biomedical sensor tools used Neurokit2 as the main signal processing tool. In the context of sample size for data processed with either Neurokit2 or BioSPPy, it is reasonable to assume that Neurokit2 would show better results in statistical tests. For larger sample sizes statistical tests provide a higher precision because the power of a test increases with increasing sample size [68] (p. 235). Therefore, the results of statistical tests that used Neurokit2 data can be interpreted as more precise.
- As mentioned in the methodical chapter of this thesis, the EDA sensor used measures skin resistance instead of skin conductance. Therefore, the results of SCL have to be seen in reverse: When the results describe a low SCL value, the sensor measures

a low resistance between the electrodes, or in other words, a high conductance and so on.

- Both python EMG processing tools (BioSPPy and Neurokit2) were not able to work with the data due to the low data sampling frequency of approximately 70 Hz. Therefore, the signal was analyzed manually by comparing the number of events (spikes in amplitude). The result can be seen in figure A.4. It can be concluded that for participant two, three and four, there were significantly more EMG events during the task part of the experiment than in the resting period. Participant one and five did not show clear differences between the two phases.

- The temperature sensor (LM 35 DZ) used for this test had an ensured accuracy of 0.5 °C at room temperature. When testing for its fluctuations in temperature readings with no skin contact for approximately 170 seconds, the sensor signal varied by 0.03 °C over the total measurement period. After approximately 100 seconds, the sensors temperature value rose by around 0.4 °C within 30 seconds and stayed at this new value for the final 40 seconds. Jumps in temperature in the experimental datasets were usually in order of 0.5 °C.

- The Arduino PPG sensor was misplaced for some time for participant 4, therefore the data for task 2 and 3 were not used in any calculations.

- Participant 4 was an EDA non-responder and therefore excluded from sampled statistical tests from all participants regarding EDA extracted signals.

- Blood pressure measurements taken had a very low sampling rate of two to three measurements per phase. Therefore, the data was not usable for statistical analysis and was not included in the results or discussion.

Table 4.3.: Significant Differences between Resting Period and Tests for Selected Features (Welch statistical t-test). The window size for the SCR was 5 seconds, and for the HR change was 40 seconds.

Features	**Reading Task**		**Arithmetic Task**		**Technology Task 1**		**Technology Task 2**	
	t-Stat	***p-Value***	***t-Stat***	***p-Value***	***t-Stat***	***p-Value***	***t-Stat***	***p-Value***
SCL	-30.170	<0.001	-41.570	<0.001	-62.880	<0.001	-75.270	<0.001
SCR	**1.471**	**0.143**	-2.376	0.018	-3.503	<0.001	**-0.732**	**0.465**
ECG HR	19.070	<0.001	23.200	<0.001	13.090	<0.001	/	/
PPG (Bit) HR	19.640	<0.001	19.530	<0.001	12.180	<0.001	-2.663	0.008
PPG (Ard) HR Change	19.210	<0.001	9.170	<0.001	-4.648	<0.001	19.450	<0.001
PPG (Ard) HR	11.270	<0.001	11.010	<0.001	-4.006	<0.001	27.520	<0.001
Respiration rate	2.590	0.010	8.749	<0.001	9.008	<0.001	14.410	<0.001
Temperature	15.350	<0.001	8.848	<0.001	10.100	<0.001	**-1.769**	**0.077**

Table 4.4.: Significant Differences between Resting Period and Tests for HR Change (Arduino) and SCR in different sliding windows (Welsh statistical test).

Features	Window Size	Reading Task		Arithmetic Task		New Medium Task 1		New Medium Task 2	
		W-Stat	***p-Value***	***W-Stat***	***p-Value***	***W-Stat***	***p-Value***	***W-Stat***	***p-Value***
SCR	3 sec.	**0.326**	**0.745**	**-1.506**	**0.133**	-2.885	0.004	**0.405**	**0.686**
	5 sec.	**1.471**	**0.143**	-2.376	0.018	-3.503	<0.001	**-0.7325**	**0.465**
	10 sec.	**0.703**	**0.482**	-2.308	0.022	-3.728	<0.001	**-0.139**	**0.890**
	15 sec.	**-0.774**	**0.440**	-2.968	0.003	-4.849	<0.001	**-0.320**	**0.749**
	20 sec.	**-0.231**	**0.817**	-2.113	0.036	-5.090	<0.001	**-1.129**	**0.260**
HR	6 sec.	13.350	<0.001	8.307	<0.001	**-1.863**	**0.064**	18.790	<0.001
	10 sec.	15.110	<0.001	7.952	<0.001	-2.545	0.012	19.250	<0.001
	20 sec.	17.610	<0.001	9.822	<0.001	**-1.804**	**0.073**	20.350	<0.001
	30 sec.	18.220	<0.001	8.712	<0.001	-2.761	0.006	19.530	<0.001
	40 sec.	19.210	<0.001	9.170	<0.001	-4.648	<0.001	19.450	<0.001

Table 4.5.: Mean, Variance and Median of normalized heart rate extracted from ECG, PPG (Arduino) and PPG (BITalino) in direct comparison for resting phase and public speaking tasks (task phase) for four participants of the experimental protocol from July 2022.

Resting phase (initial rest)

Participant	**ECG**			**PPG (Arduino)**			**PPG (BITalino)**		
	Mean	***Variance***	***Median***	***Mean***	***Variance***	***Median***	***Mean***	***Variance***	***Median***
P1	-0.457	0.716	-0.518	-0.434	0.594	-0.432	-0.293	0.703	-0.330
P2	-0.694	0.575	-0.647	-0.593	0.293	-0.647	-0.429	0.496	-0.418
P3	-0.545	0.265	-0.619	-0.648	0.079	-0.599	-0.333	0.201	-0.429
P5	-0.119	0.905	-0.357	0.029	1.224	-0.148	-0.063	1.054	-0.312

Task phase (public speaking)

Participant	**ECG**			**PPG (Arduino)**			**PPG (BITalino)**		
	Mean	***Variance***	***Median***	***Mean***	***Variance***	***Median***	***Mean***	***Variance***	***Median***
P1	0.427	0.540	0.345	0.402	0.392	0.469	0.750	0.596	0.792
P2	-0.074	0.474	-0.159	-0.159	0.235	-0.235	-0.075	0.719	-0.189
P3	0.183	0.935	0.081	-0.111	0.467	0.050	0.332	0.609	0.305
P5	0.147	1.161	-0.092	0.161	0.792	-0.079	0.174	1.657	-0.105

4.2. Implementation of REALSTRESS

For implementation of the scoring function for SCR and HR change, different libraries and code-resources were tested for their signal acquisition and processing skills by recording breathing pattern test and visually checking how the signal changes during the test phases. Each library and sensor type was tested separately.
The results forHR calculations with the PPG (Arduino) sensor showed that *PulseSensorPlayground.h* performed best when using the implemented interrupt function. In particular, the resulting signal was stable, but changed with breath holding events, and produced a heart rate value which was in expected range of human heart rate. The alternative function provided by the library authors, which avoids the interrupt function, performed comparably well in terms of the signal changing in time of event phases. However, in contrast to the first version, the amplitude of signal changes was unrealistically high. Additionally, the portrait heart rate value was generally a little bit higher compared to the previous calculations. The results for *PulseSensorPlayground.h* can be seen in figure A.5. *Heart.h*, the library by BioData for heart rate calculations, produced a signal that showed strong fluctuations even during resting phase, with high amplitudes of change (see figure A.2. Consequently, the Heart.h performed worst compared to both PulseSensorPlayground.h's functions.
Next, the EDA sensor processing was tested with *BioData*'s EDA library *SkinConductance.h*. The calculated SCR and SCL signals both showed signal changes throughout in-between phases of the breathing pattern test, but stayed relatively stable within single phases. Results of the initial breathing pattern test for this library can be seen in figure A.6. Furthermore, SCR produced results consistent with the first tests for repeated trails of breathing pattern tests. In contrast, the code resource provided by *engineersgarage.com*, which was designed to find changes in raw EDA signal, performed unsatisfying, triggering constantly. Between the two reviewed assets, *SkinConductance.h* proved to be a stable option for SCR signal detection.

The two biosensors were tested simultaneously and separately after integration into the *Interactive Wall* coding environment. The results for the implementation using the interrupt function for the HR measurements can be seen in figure 4.3 (B) and show HR and raw EDA signal during a 90 second resting period. It can be seen that the heart rate measured separately from the EDA sensor stabilizes itself, after an initial 20 seconds of adjusting period, to a reasonably value. In contrast, the heart rate measured simultaneously with the EDA signal, inhibits strong fluctuation and never really stabilizes. The two raw EDA signals shows a familiar data pattern, with smaller local changes and bigger global drops and peaks. However, there is a difference in their baseline value.
The testing was repeated for the no-interrupt version of the heart rate measurement and the results can be seen in figure 4.3 (A). For this version of the code, there is still a difference in amplitude of the heart rate baseline, but the dual measurement heart rate signal is much more stable than its interrupt heart rate counter part in 4.3 (B). Note that the single measurement heart rate signal in 4.3 (A) shows a signal artifact in the

first 10 seconds of measurement.
The EDA raw data inhibits a lower amplitude for the single measurement than for the dual measurement for both tests.

Any further results were produced using the SCR function provided by BioData

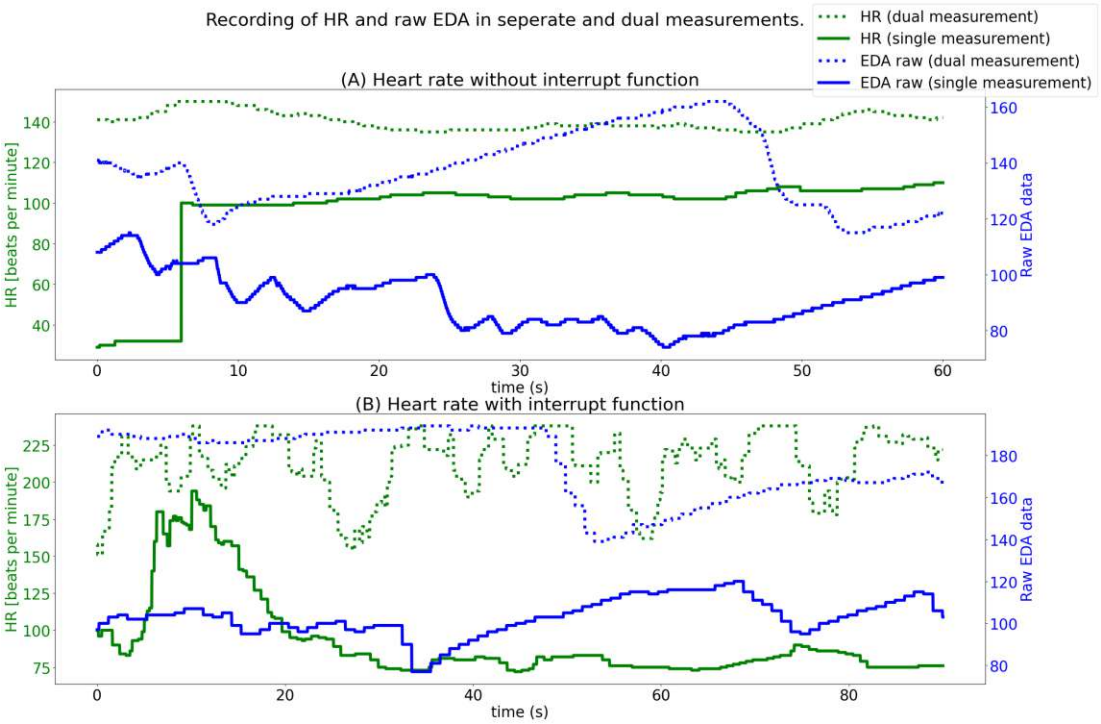

Figure 4.3.: PulseSensor Playground's heart rate calculation (green) and raw EDA data (blue) measured either simultaneously (dotted) or separated (solid). The measurements were performed using PulseSensor Playground's library either with (B) or without (A) interrupt function. The raw EDA signal was averaged over 50 data points each.

in *SkinConductance.h* and the non-interrupt heart rate calculator by *PulseSensorPlayground.h*.
Different binning sizes for the sliding window implementations for both signals were tested through single and dual measurements of both sensors. The testing protocol followed a breathing pattern change with time, during which the signals were recorded and later compared against the protocol phases. Through several rounds of pre-testing the scoring function for the HR measurements was determined to show an averaged negative change in signal as response to events. This method proved most stable within trials. The SCR scoring represented the number of detected events within a given window size. Given this implementation of the scoring function, a summary of different binning sizes for dual measurements of both signals is portrait in figure 4.4. The window size combination that

showed the best results is framed green, which was a binning size of 30 seconds for HR and 10 seconds for SCR. That the HR score increased only after the event had passed for most measurements, which was implemented this way. Smaller HR binning sizes showed more random changes throughout the protocol, with no strong response to events (breath holding). Within the bigger binning sizes (20, 30, 40 seconds), the difference between the trials was small, making all three sizes good contenders for the implementation.
The SCR score performed worst compared to the HR score in all versions of SCR binning sizes. Within the signal, larger binning windows provided a better score within events of the protocol. However, the SCR score spikes much more random compared to the heart rate score.

On base of the results presented in figure 4.4, a final round of testing was performed on the reference candidate that was repeatably tested throughout the implementation, and external participants. Both groups performed the breathing pattern test that had been used to create the functions to be tested. Additionally, the external participants completed the arithmetic task from the stress test in July of 2022, while the reference candidate participated in several rounds of the cold pressure test. The test was performed to check if there was a difference in mean score within the phases.
During all tests both implemented functions were tested with the sliding window binning size parameters chosen beforehand (HR: 30 seconds, SCR: 10 seconds). For each test, mean, variance and median of SCR Score, HR score, HR change and total score (combination of SCR and HR score) were calculated over the average within the participants or trails (reference candidate). They can be found in table 4.6 for both breathing pattern tests and in table 4.7 for the arithmetic and cold pressure test.
When comparing the breathing pattern test results in table 4.6 between the external participants and the reference candidate, it can be seen that the mean total score behaved as expected by rising in the recovery phases, after the event had happened. The SCR score showed better results for the reference candidate compared to the external participants, since the latter presented a lower mean score for event and recovery phase set side by side to the resting phase, while the former provided a more event-related mean score. The HR reached a score that was in accordance with the phases for the external participant and the reference candidate. In addition to that, the HR change showed a decreasing tendency for both non-resting phase for external participants, which was only achieved for the recovery phase by the reference candidate. Given those results, it can be said that the difference in total score was reached for both groups within the breathing pattern test. Furthermore, the external participants that were not tested throughout the implementation process, were able to prove that the implementation can produce results comparable to the reference candidate. By comparing both score, the HR score is to be preferred, especially when considering the results of the external participants.
The arithmetic task in table 4.7 showed good results for HR score and total score, but unsatisfactory mean SCR scores through the different phases. These results are similar to those presented for the breathing pattern test, which showed that the implementation testing with breathing patterns instead of established stress tasks still produced a final system that was able to detect different stress phases in classic stress test design. Similar

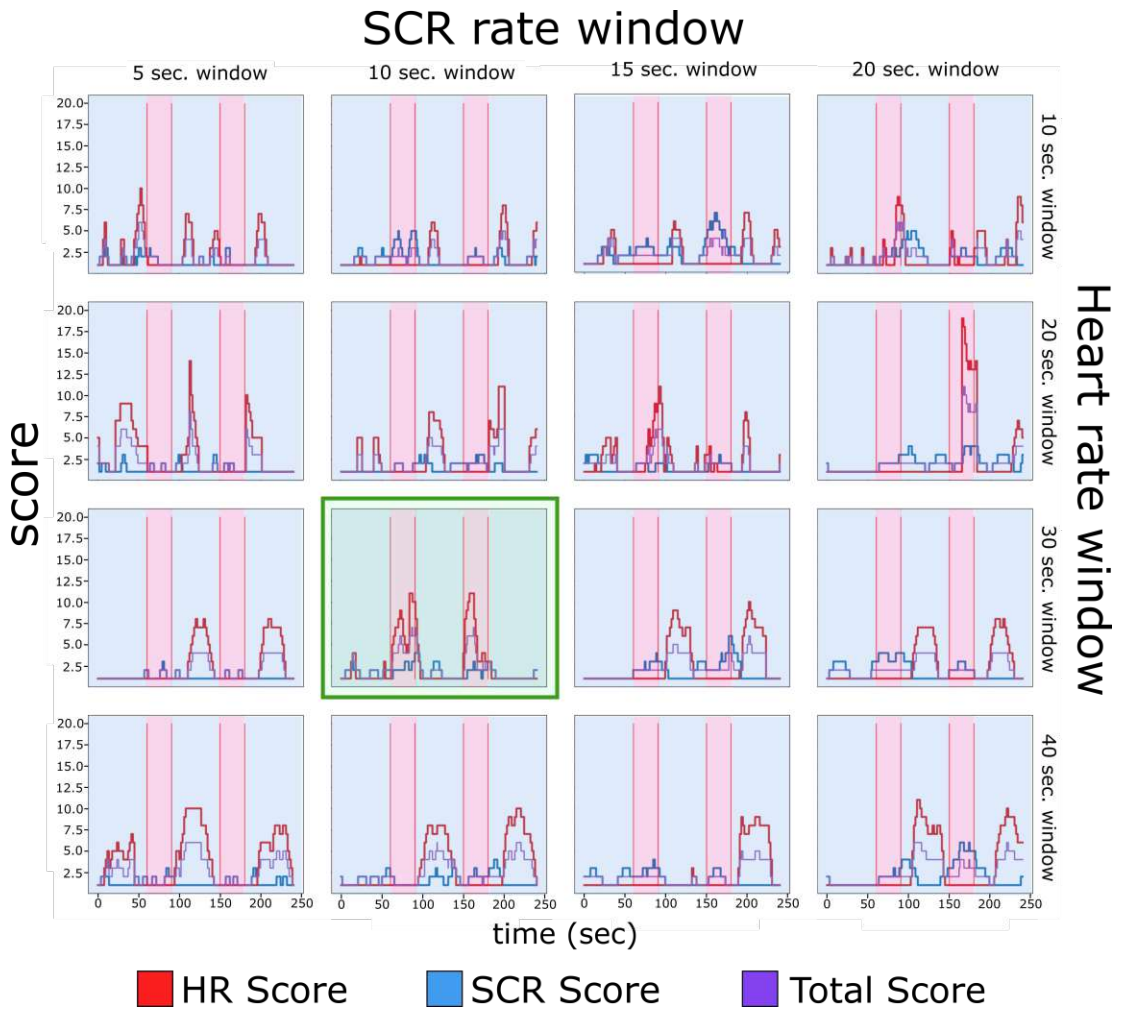

Figure 4.4.: Different binning sizes for sliding window algorithm in dual measurements of SCR and heart rate change. The implemented score for both signals individual and the averaged score is represented graphically. Additionally, the different phases of protocol for the breathing pattern test are marked by color (normal breathing - blue; breath holding - pink).

can be said about the cold pressure test presented in the same table. The mean total score of the recovery phase for this test is clearly increased compared to the resting phase, which is what was expected from the implementation. As was said for all previous results for this testing round, the heart rate score performed much better than the SCR score. However, for the cold pressure test, SCR shows a stronger increase in score for the event phase, especially compared to the other tests performed.
What is evident in all four tests is that the SCR score and the heart rate score are out of sync for most of the results. While the SCR score showed a peak mean value during events, the heart rate score increased more towards recovery. Thus, the total score, being the average of the both single scores, is generally lower in comparison.
The presented results therefore support the hypothesis that the implemented scoring function is able to deliver a score, whose averaged value is distinct between different phases. However, the size of difference in scoring is of smaller scale than expected.

Table 4.6.: Mean, Variance and Median of SCR Score, HR Score, and HR Change in Participants (7 measurements) and Reference (10 measurements) during Breathing Pattern Test in three Test Phases (Initial Rest, Event, Recovery). The HR window size is 30 seconds, the SCR window size is 10 seconds. The Total Score is averaged 1:1 over the HR and SCR score.

Participants (Breathing Pattern Test)

Features	**Rest**			**Event**			**Recovery**		
	Mean	***Variance***	***Median***	***Mean***	***Variance***	***Median***	***Mean***	***Variance***	***Median***
SCR Score	1.368	0.633	1.000	1.269	0.400	1.000	1.095	0.120	1.000
HR Score	1.500	2.409	1.000	1.641	3.280	1.000	2.360	6.958	1.000
HR Change	5.094e-03	8.747e-04	0.000	-5.912e-04	9.471e-04	0.000	-6.530e-03	1.800e-03	0.000
Total Score	1.434	0.668	1.000	1.455	0.834	1.000	1.728	1.704	1.000

Reference (Breathing Pattern Test)

Features	**Rest**			**Event**			**Recovery**		
	Mean	***Variance***	***Median***	***Mean***	***Variance***	***Median***	***Mean***	***Variance***	***Median***
SCR Score	1.149	0.211	1.000	1.663	0.394	2.000	1.184	0.236	1.000
HR Score	1.946	3.593	1.000	1.275	0.839	1.000	4.263	13.413	3.000
HR Change	-7.985e-03	9.021e-04	0.000	4.261e-02	1.511e-03	0.050	-2.581e-02	3.129e-03	-0.030
Total Score	1.548	0.904	1.000	1.469	0.314	1.500	2.723	3.232	2.000

Table 4.7.: Mean, Variance and Median of SCR Score, HR Score, and HR Change in Participants (7 measurements) during Arithmetic Test, and in Reference (9 measurements) during Cold Pressure Test in three Test Phases (Initial Rest, Event, Recovery). The HR window size is 30 seconds, the SCR window size is 10 seconds. The Total Score is averaged 1:1 over the HR and SCR score.

Participants (Arithmetic Test)

Features	**Rest**			**Event**			**Recovery**		
	Mean	***Variance***	***Median***	***Mean***	***Variance***	***Median***	***Mean***	***Variance***	***Median***
SCR Score	1.309	0.454	1.000	1.430	0.589	1.000	1.390	0.715	1.000
HR Score	1.693	3.293	1.000	2.507	6.492	1.000	2.193	7.572	1.000
HR Change	1.591e-02	2.301e-03	0.010	-6.848e-03	2.003e-03	0.000	4.047e-04	2.306e-03	0.000
Total Score	1.501	1.024	1.000	1.968	1.904	1.000	1.792	1.912	1.000

Reference (Cold Pressure Test)

Features	**Rest**			**Event**			**Recovery**		
	Mean	***Variance***	***Median***	***Mean***	***Variance***	***Median***	***Mean***	***Variance***	***Median***
SCR Score	1.152	0.221	1.000	1.709	0.465	2.000	1.200	0.243	1.000
HR Score	2.215	4.672	1.000	1.337	1.153	1.000	4.115	12.970	1.000
HR Change	-1.016e-02	1.127e-03	0.000	3.940e-02	1.632e-03	0.050	-2.203e-02	3.333e-03	-0.020
Total Score	1.683	1.177	1.000	1.523	0.412	1.500	2.658	3.138	2.000

5. Discussion

This chapter presents an in-depth discussion of the results presented in the previous chapter. Firstly, the implications of the statistical answers to the first research question are reviewed. After, the results of the implementation and testing phase of REALSTRESS are addressed.

5.1. Experimental Testing of Low-Cost Sensors

The first part of this thesis concerned itself with the hypothesis that inexpensive and open-source biomedical sensors are able to detect stress in a laboratory setting. To answer this research question, eight sensors of different price range were challenged in a small-scale stress induction experiment, which involved five participants. Through statistical analysis of the recorded data, the hypothesis was supported for two out of the five inexpensive sensors and for all three mid-priced sensors by BITalino. Both, EDA and PPG (Arduino), proved successful in distinguishing between different test phases through their signal features. The ST sensor was not able to produce similar results, and rejected the hypothesis with a p-value of 0.975 through the Welch's t-test for equal means. EMG and blood pressure (fitness tracker) signals were inconclusive, because statistical analysis of the data sets was not possible due to poor sampling rate of the signals. Finally, the three BITalino based sensors (ECG, PPG, respiration belt), who are located within a slightly higher price range, proved to meet expectations by delivering distinguishable signals between all test phases.

The results presented for the EDA and both PPG sensors are in line with previous research findings, such as [5], [61] and [60]. The outstanding performance of the EDA extracted SCL feature is apparent when visually examining the data, which showed that the SCL amplitude decreased sharply with the start of each stress tasks and recovered slowly right after (see figure A.8). When comparing SCL between the one EDA-non-responder participant and remaining candidates (see figure A.7 for an example), the signal decreased constantly in amplitude for the non-responder, while it was more distinct in amplitude distributions for the remaining (EDA responding) participants. Given those findings, the measured SCL signal can be interpreted as not random, but as a specific response to the stressor. Thereafter, SCR did not perform as well as SCL regarding the results of the statistical analysis, even though it is the preferred EDA signal within most stress research publications [77]. This might have been due to the relatively low signal sampling rate of only 70 Hz. Keeping in mind that SCR is a fast-changing signal, which is extracted through peak finding algorithms, and that SCR peaks are often not clearly

separated from each other, it makes peak finding additionally challenging in EDA signal analysis [54].
The low sampling rate was less troublesome for the data processing of the Arduino-based PPG sensor, as peak detection for this type of signal is generally easier to perform as the peaks are more distinct. This was proven by direct comparison between Arduino PPG and BITalino ECG and PPG sensors, where the latter had a sampling rate of 1 kHz. All three resulting heart rates were similar within their performance during the statistical analysis. Overall, all three cardiovascular sensors behaved as expected from literature for the given task. However, it is remarkable that both PPG senors provided comparable results given the price difference of almost 100€ between them.
The low sampling rate of the Arduino system also influenced the EMG sensor negatively. Both biomedical signal processing tools used in this thesis were not able to analyze the EMG signal properly by extracting peaks or removing noise. As a result, the peak detection had to be done manually, so that remarks can still be made about the data. First, the peak count showed an excess of events during the task phase for three out of the five participants. This speaks for the usability of the method for stress detection, and is in line with results presented in related work such as [43]. However, since the signal was not processable with external open-source tools, the sensor cannot be recommended for further research, since it limits the ability to analyze the data effectively.
The ST sensor was not able to distinguish between resting phase and the sampled mix of data for all task phases. It could be argued that this result is not representative, since the p-value of the Welch test rejected the null hypothesis of equal means between the data sets only for one out of four tasks when analyzed separately. However, other factors argue against the usability of the sensor, most prominently the test of sensor stability described in the additional remarks of the results. Even though the ST sensor showed only small fluctuation over three minutes of recording, the witnessed temperature increase of 0.4 °C within 30 seconds during the test behaved similarly to the increases seen in the measured data for the experimental protocol. Therefore, the changes in ST might exhibit some randomness due to the sensor's behavior. Another argument against the sensor is based on the relatively low sensor accuracy of 0.5 °C. Given that ST is a slow changing biomedical signal, larger shifts in ST could not be expected within the experiment. While smaller changes are possible, related research with sensor equipment of higher accuracy showed results in which ST changes were observed to be much smaller than measured in this thesis for comparable time intervals [90]. In conclusion, the chosen ST sensor is not a suitable choice for human stress response research given the reasoning provided above.
The respiration rate was expected to show a very clear difference between resting phase and task phase, since the resting phases were non-verbal, while most task phases had a least a small talking component. Furthermore, previous research, such as [91], rated respiration features as useful tools in stress detection. Even though all participants were asked to not talk during task 3 and task 4, none of them managed to perform those tasks non-verbally. When comparing the results (difference in means between phases) between verbal and non-verbal tasks, it is surprising that the difference to the resting phase is

much stronger for the non-verbal exercises. The contrary result would have been more intuitive, with a more significant difference between verbal tasks and resting phase. Even though the participants spoke during non-verbal talks, the amount of talking done was much less compared to verbal tasks. The results suggests that the respiration rate is a good indicator for distinguishing between test phases, even when the tasks are non-verbal. Not much can be said about the fitness tracker used for measuring blood pressure. Given that the sensor did not allow for customization and that the sampling rate was very low, no usable results could be extracted from the sensor.

In summary, all sensors except for the fitness tracker, the ST sensor and the EMG sensor, were able to produce data that allowed to discriminate between the resting phase and the task phases with statistical significance. The inexpensive PPG Arduino sensor was able to perform comparable to the more expensive PPG Bitalino sensor. Both the EDA system and the EMG sensor suffered due to their low signal sampling rate, which made it more difficult to post-process the data using tools specifically targeted towards such sensors (the usual sampling frequency is in the range of 1 kHz [56] [24]). However, while the EDA features could still be used for further analysis, the EMG data had to be manually analyzed which excluded the data from statistical tests, since it would have been too much effort to process the data manually in more detail.
The research question investigated in this thesis stands out compared to similar research work done for commodity and inexpensive sensor hardware, such as [60], [43] or [86], due to the extremely inexpensive PPG sensor used during experiments. Despite low price, the sensor was able to produce similar results to the almost thirty-times as expensive BITalino PPG sensor.

For future research done on this topic and following this thesis's results, certain improvements need to be made to minimize limitations encountered in this thesis.
Firstly, the introduction of less-familiar equipment during the stress test, since two tasks were performed on a Nintendo Switch, led to many questions by the participants during the tasks. This was a problem, because the tasks were originally designed to be non-verbal. In future experiments, non-verbal tasks should not be performed on for the participants non-familiar equipment to provide accurate results.
Secondly, if the budget provides the possibilities, some sensors should be replaced with more advanced sensors which deliver a continuous, high-quality signal of higher sampling rate. The sensor that would benefit most from updating are the ST sensor, the blood pressure sensing unit and the EMG sensors. The latter two sensors would benefit in form of better automatic post-processing possibilities. The former needs to be updated to ensure that the results for ST measurements are not just random sensor behavior but a response to the stressor.
Lastly, a larger number of participants for stress testing could provide a higher level of confidence for the results of the experimental part of future research work. Furthermore, shortening the length of the test protocol might be useful, since data collection would be less time consuming, which would allow for more subjects to be tested.

5.2. Review of REALSTRESS

The second research question of this thesis tied up with the results of the first, by asking if the inexpensive biomedical sensors, which have proven themselves to be sufficient in simple stress measurements through the methods of part one, were capable of performing live stress measurements with direct feedback for the to be measured participants in form of a real-time stress scale. After several rounds of testing throughout the implementation process, the final round of stress trials performed on the reference candidate, who was the main test subject for all of implementation, and external participants, who were not included in the implementation process, answered the research question through positive results. Although the mean value by which the stress scale differed between testing phases was small, the existence of this difference within all four tests showed that the implementation was successful. Between the two sensor stress scores, SCR provided a more intuitive result, since its peak score appeared during the event phases of the protocol. However, HR change score showed higher magnitudes of stressor reaction, although this response was delayed and appeared primarily during the recovery phase. Given that the final implementation was built with the delayed signal behavior in mind, the results are what was expected. Because the two score peaks are out of sync, the total score, which is the averaged score of both sensor scores, suffers in amplitude. Nevertheless, the implementation presented sufficient success, by delivering a result that answered the research question positively.

Reviewing the results in more detail, a few observations need to be made about the different code-resources considered for the real-time signal processing. The first PPG library that was tested and eventually chosen, *PulseSensorPlayground.h*, requires a manual threshold input to work. Therefore, this library was a possible choice for measurements, where the threshold is similar for all individuals or threshold-finding is performed externally. During testing with external participants, the library failed to detect a HR for some participants. Since the function's threshold was set for the reference candidate, this behavior was expected to an extent. This means in concrete terms that the implementation only worked for those users, whose PPG signal met the threshold. However, when looking at the results of the external testing, the library was successful when averaged over all participants. Furthermore, the threshold-problem was not only influenced by participant-specific signal characteristics, but also by temperature influences on the blood flow to the fingers.
Even though the threshold-problem was an inconvenience for the implementation, similar libraries, such as BioData's *Heart.h*, did not compare to *PulseSensorPlayground.h* in terms of stability and accuracy. Furthermore, the latter offered processing tools with and without interrupt functions, which provided an additional level of customizability.
One interesting discovery was made when analyzing the PPG signal further: The sensor appeared to output not a raw PPG signal, but the second derivative of it. This might have caused the problems for the BioData library, because it was expecting a different signal behavior.

The EDA sensor tested with BioData's SCR calculation functions showed expected behavior, with the signal responding to events (breath holding) and recovering after. However, it appeared that the signal was rising even before the breath was held, which might be due to anticipation. Nevertheless, the SCR functionality BioData offered was the most promising and intuitive, which might be because the EDA sensor used by BioData was very primitive, which could lead to the code being more robust to noisier signals. The SCL signal produced was not considered for the implementation, even though SCL showed excellent results for the first research question. Because the good results for SCL of part one were mainly due to the length of measurements, and the long recovery phase of the SCL signal could be accounted for since the signal did not need to respond to changes quickly, the focus was laid on SCR. Computing an implementation based on SCR is better suited for live detection of stressor responses, since signal changes are faster compared to SCL.
The code-resource published by Engineersgarage.com did not work as intended and constantly reported a change in signal. This comes especially surprising, because the authors used the exact same EDA sensor that was used in this thesis. Their code determined the change threshold automatically, which might not have worked as intended and lead to output being hyper-sensitive.
Going back to the results of the tests performed with the final implementation of the scoring function, generally, the SCR score performed worse compared to the HR score. Especially for the external participants, the score changed only slightly between the testing phases, and not always as intended. Since the SCR score showed far better results for the reference candidate, over-fitting might have been a bigger problem for SCR than for the HR score. Furthermore, the final implementation might had worked better on the external participants, if the influence of SCR score on the total score would have been significantly reduced or even removed completely. However, the SCR score provided good results for the reference candidate, which leads to the conclusion that the signal could be a valuable asset if over-fitting is able to be avoided during the implementation process. This could be done by including more reference candidates throughout implementation, and fitting the data on each candidate individually as well as on the candidates' average.

Another limitation proved to be dual measurements of both sensors: The superior method of HR calculations with *PulseSensorPlayground.h* was determined to be the PPG tool using an interrupt function. However, this function can interrupt (hence the name) other functions in doing their job, or work worse themselves when other functions are also executed in the same loop. This was proven though the results, which showed that for dual measurements using the interrupt function, the HR signal was no longer stable. Therefore, the code-version working without an interrupt function was chosen for the final implementation, since it provided better overall results for dual measurements.
Alternatively, both sensors could have been measured through separate microcontrollers, with the total score being calculated on a third device. However, one main topic of this thesis was a simple and inexpensive design, which would have not been achieved by using

more devices than necessary.

The problem of dual measurements continued throughout the implementation of sliding window functions for both sensors. Finding the best combination of binning sizes for the sliding windows showed expected results. Bigger bins for the HR function provided the best scores, which is in line with literature. Considering that most measures of HRV require bigger window sizes to be reliable [75], and the measure used for the HR changes in this thesis's trials is somewhat related to those measures (basic form of numeric deviation of HR), the results are reasonable.

Even though the results presented showed that a live stress feedback could be done with inexpensive sensors and open-source libraries, more research has to be done to present a solution that works reliably for all users. It would be advised to include more reference candidates during the implementation process to avoid over-fitting. Another improvement that should be done is the inclusion of an automatic threshold finder for the PPG signal processing. Furthermore, different sensor types or extracted signals from the presented sensors could be considered. More specifically, how a SCL-based score compares to its SCR counterpart presented in this thesis.

6. Conclusion

Finally, this chapter presents a conclusion of the thesis' work. The first section concerns itself with the conclusion of the first research question. Conclusion and outlook for REALSTRESS are presented in the second section of this chapter.

6.1. Experimental Testing of Low-Cost Sensors

This thesis extends work done in the area of human stress sensing, by hypothesizing that low-cost sensors are able to measure biomedical signals with high-enough quality to be used in stress detection experimental trials. In order to verify this research question, sensor data collected within a classic stress test design was reviewed through statistical means. Out of the five inexpensive sensors tested, two supported the hypothesis (EDA, PPG), one sensor rejected it (ST sensor), and two had too low of a sampling rate to perform statistical testing on their data (EMG, fitness tracker/blood pressure). Furthermore, the low-cost Arduino PPG sensor was able to compare to the mid-price PPG and ECG sensors by BITalino.
The sensor testing was done with unique tasks on different individuals, which included an EDA-non-responder participant. While the number of participants would need to be increased for a more generalized result, this mix of trials and person-specific stress reactions provided a good perspective on the sensors abilities. Expectations about the performance of the inexpensive sensors were kept within limits, since most of the previous work this thesis ties in with presented results acquired through higher-cost well established biomedical equipment. However, expectations were exceeded for both EDA and PPG (Arduino), which were the most prominent sensors tested within the stress research literature found. Unanticipatedly, both signal extraction tools used in this thesis did not manage to process EMG data, because of its low sampling rate. All three BITalino sensors performed as expected, which in turn made the two cardiovascular sensors (ECG, PPG) a good reference base for the other sensors to be compared against.
The work presented in this thesis about low-cost sensor usability for stress detection has extended the previous work by not only testing inexpensive sensing equipment, but also comparing it to higher-budget sensors directly. Additionally, the results present possible limitations that can be met when using cheaper equipment (e.g., issues due to low sampling rate). By focusing on a simpler and more customizable sensing system, which does not require proprietary software to be used, this research provides a reproducible low-budget solution for human stress response detection.

When following up on this thesis, future researchers should take care to invest in a

different system for EMG and ST measurements. Furthermore, testing the low-cost sensors in non-laboratory tests could be interesting, especially considering movement artifacts. Finally, engaging a higher number of participants in the experimental sessions should give the results a higher level of confidence.

6.2. REALSTRESS

The aim of the second part of this thesis was to determine the practicability of implementing a sensor-based stress scale with real-time feedback through low-cost biomedical sensor equipment. By means of statistical signal analysis of different phases within a laboratory stress test, a distinction between the mean stress scales could be demonstrated for a reference candidate and an external control group. Furthermore, the distinction between test protocol phases could not only be made for the stress inducing exercise on which the implementation was based on, but also for two disparate popular stress trials. Thereby, the implementability of the proposed system could be confirmed.
On the grounds that numerous open-source libraries for biomedical signal extraction already existed, focusing the implementation on building a sliding-window system that works for dual signal processing rather than reinventing the wheel by producing a signal processing tool from scratch, was the route chosen for this thesis. In retrospect, this decision proved to be wise, since the combination of two different signal processing tools working simultaneously verified the expectations of mutual conflicts between the sensors. However, through step-wise parameter adjustments and extensive testing, the signal processing functions could be adjusted to produce the desired results. Furthermore, since the testing during the implementation process was done based on breathing patterns, the expectations for the usability of the score for more classical stress inducing tasks were lowered. Nonetheless, the scale proved to be comparably efficient for the latter, exceeding expectations and demonstrating that the body's reaction to breath holding can be used to adjust a system for stress detection. Finally, it was anticipated that the implementation would inherit some form of over-fitting due to it only being based on one candidate. Still, the final score was less prone to over-fitting than initially thought, even though it was still noticeable.
The second part of this thesis has demonstrated that real-time feedback in stress measurements is possible with inexpensive sensor equipment. Furthermore, the implementation could be accomplished through less-conventional stress inducing methods, but still transferred well in tests, where proven stress tasks were up for testing. A real-time feedback stress scale, available in educational facilities or bigger office spaces, could improve the conception of stress throughout individuals, and might lead to improved health by reminding users to take their bodies' own stress signals more serious. This work confirms that this concept can be realized with low-cost sensors and open-source libraries, providing a foundation that can be built upon in future research.

6.3. Outlook

Based on the results, future research adding to this work might consider a more versatile approach for the implementation, including more candidates throughout the building-process. Additionally, since this thesis focused the real-time feedback implementation on two sensors that were deemed appropriate given the initial selection, including different sensors in future research might share more insight in the topic of real-time signal processing. Considering experimental settings outside the laboratory might also be an interesting follow up to this thesis.
The next step for improving the implementation would be investing in a user-friendly interface that presents the stress score graphically. This could be done through the gaming engine Unity.

Bibliography

[1] A. P. Allen, P. J. Kennedy, J. F. Cryan, T. G. Dinan, and G. Clarke. Biological and psychological markers of stress in humans: Focus on the trier social stress test. *Neuroscience & Biobehavioral Reviews*, 38:94–124, 2014.

[2] A. P. Allen, P. J. Kennedy, S. Dockray, J. F. Cryan, T. G. Dinan, and G. Clarke. The trier social stress test: principles and practice. *Neurobiology of stress*, 6:113–126, 2017.

[3] J. Allen. Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement*, 28(3):R1–39, Mar. 2007.

[4] J. A.N. Galvanic Skin Response Measurement and Analysis. *Galvanic Skin Response Measurement and Analysis*, 10:2015, June 2015.

[5] A. Arsalan and M. Majid. Human stress classification during public speaking using physiological signals. *Computers in Biology and Medicine*, 133:104377, 2021.

[6] L. Asarian, V. Gloy, and N. Geary. Homeostasis. pages 324–333. Academic Press, San Diego, Jan. 2012.

[7] M. Banzi and M. Shiloh. *Getting started with Arduino.* Maker Media, Inc., 2022.

[8] M. Benedek and C. Kaernbach. A continuous measure of phasic electrodermal activity. *Journal of neuroscience methods*, 190(1):80–91, 2010.

[9] M. Benedek and C. Kaernbach. Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology*, 47(4):647–658, 2010.

[10] berrybase.de. Storepage: optischer puls sensor, 2022.

[11] R. Bhargava, M. G. Gogate, and J. F. Mascarenhas. Autonomic responses to breath holding and its variations following pranayama. *Indian Journal of Physiology and Pharmacology*, 32(4):257–264, 1988.

[12] biosignalsplux. *BITalino (r)evolution HeartBIT Bundle Data Sheet*, 2020.

[13] biosignalsplux. *Electrocardiography (ECG) Assembled Sensor Data Sheet*, 2020.

[14] biosignalsplux. *Photoplethysmography (PPG) Sensor Data Sheet*, 2020.

[15] biosignalsplux. *Respiration (PZT) Sensor Datasheet*, 2020.

[16] biosignalsplux. Bitalino, 2022.

[17] biosignalsplux. Opensignals (r)evolution (download), 2022.

[18] biosignalsplux. Store page: Heartbit, 2022.

[19] biosignalsplux. Store page: Piezo-electric respiration (pzt) sensor, 2022.

[20] W. Boucsein. *Electrodermal Activity*. Springer US, Boston, MA, 2012.

[21] W. Boucsein. Principles of electrodermal phenomena. In *Electrodermal activity*, pages 1–86. Springer, 2012.

[22] T. Buchanan, D. Tranel, and R. Adolphs. Impaired memory retrieval correlates with individual differences in cortisol response but not autonomic response. *Learning and Memory*, 13(3):382–387, 2006.

[23] U. a. Butt. Measuring spikes in stress levels using a galvanic skin-response sensor and Arduino.

[24] C. Carreiras, A. P. Alves, A. Lourenço, F. Canento, H. Silva, A. Fred, et al. BioSPPy: Biosignal processing in Python, 2015–. [Online; accessed <today>].

[25] C. Childs. Chapter 29 - Body temperature and clinical thermometry. volume 157 of *Thermoregulation: From Basic Neuroscience to Clinical Neurology, Part II*, pages 467–482. Elsevier, Jan. 2018.

[26] S. Cohen, T. Kamarck, and R. Mermelstein. A Global Measure of Perceived Stress. *Journal of Health and Social Behavior*, 24(4):385–396, 1983.

[27] H. Critchley. Review: electrodermal responses: what happens in the brain. neurosci 8: 132-142, 2002.

[28] D. Das. How does the pulse sensor work and how to interface it with arduino?, 2022.

[29] G. S. Everly and J. M. Lating. The anatomy and physiology of the human stress response. In *A clinical guide to the treatment of the human stress response*, pages 19–56. Springer, 2019.

[30] G. S. Everly and J. M. Lating. The concept of stress. In *A clinical guide to the treatment of the human stress response*, pages 3–18. Springer, 2019.

[31] G. S. Everly and J. M. Lating. Measurement of the human stress response. In *A clinical guide to the treatment of the human stress response*, pages 129–157. Springer, 2019.

[32] M. W. Fagerland. t-tests, non-parametric tests, and large studies–a paradox of statistical practice? *BMC medical research methodology*, 12:78, June 2012.

[33] R. M. Fish and L. A. Geddes. Conduction of electrical current to and through the human body: a review. *Eplasty*, 9, 2009.

[34] I. for Quality and E. in Health Care (IQWiG). What is an electrocardiogram (ECG)?, Jan. 2019.

[35] E. Gee. BioData, Dec. 2022. original-date: 2018-01-03T23:34:12Z.

[36] K. Georgiou, A. V. Larentzakis, N. N. Khamis, G. I. Alsuhaibani, Y. A. Alaska, and E. J. Giallafos. Can Wearable Devices Accurately Measure Heart Rate Variability? A Systematic Review. *Folia Medica*, 60(1):7–20, Mar. 2018.

[37] S. R. Green, P. A. Kragel, M. E. Fecteau, and K. S. LaBar. Development and validation of an unsupervised scoring system (Autonomate) for skin conductance response analysis. *International Journal of Psychophysiology*, 91(3):186–193, Mar. 2014.

[38] E. A. Hines and G. E. Brown. The cold pressor test for measuring the reactibility of the blood pressure: Data concerning 571 normal and hypertensive subjects. *American Heart Journal*, 11(1):1–9, Jan. 1936.

[39] Z. Huang, H. Chen, H. Ye, Z. Chen, N. Jaffrezic-Renault, and Z. Guo. An ultrasensitive aptamer-antibody sandwich cortisol sensor for the noninvasive monitoring of stress state. *Biosensors & Bioelectronics*, 190:113451, Oct. 2021.

[40] T. Instruments. Lmx24, lmx24x, lmx24xx, lm2902, lm2902x, lm2902xx, lm2902xxx quadruple operational amplifiers, 2015.

[41] T. Instruments. *LM35 Precision Centigrade Temperature Sensors*, 2017.

[42] D. Jezova, A. Makatsori, R. Duncko, F. Moncek, and M. Jakubek. High trait anxiety in healthy subjects is associated with low neuroendocrine activity during psychosocial stress. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 28(8):1331–1336, 2004.

[43] N. Z. Jia, D. Mejorado, S. Poullados, H. Bae, G. Traverso, R. Dias, and N. Hanumara. Design of a wearable system to capture physiological data to monitor surgeons stress during surgery. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 4539–4542. IEEE, 2020.

[44] M. Jones. Chapter 5 - faultfinding to fettling. In M. Jones, editor, *Building Valve Amplifiers (Second Edition)*, pages 381–429. Newnes, second edition edition, 2014.

[45] L. J. Julian. Measures of anxiety. *Arthritis care & research*, 63(0 11), 2011.

[46] L. J. Julian. Measures of anxiety: State-Trait Anxiety Inventory (STAI), Beck Anxiety Inventory (BAI), and Hospital Anxiety and Depression Scale-Anxiety (HADS-A). *Arthritis Care & Research*, 63(S11):S467–S472, 2011.

[47] E. Kaniusas. *Biomedical Signals and Sensors I.* Biological and Medical Physics, Biomedical Engineering. Springer, Berlin, Heidelberg, 2012.

[48] E. Kaniusas. *Biomedical signals and sensors I: Linking physiological phenomena and biosignals.* Springer Science & Business Media, 2012.

[49] E. Kaniusas. Physiological phenomena and biosignals. In *Biomedical Signals and Sensors I*, pages 183–282. Springer, 2012.

[50] O. Kochan, K. Przystupa, A. Shulhai, V. Pohrebennyk, J. Su, and J. Koziel. Ad-hoc Temperature Measurements Using a Thermistor. In *2019 12th International Conference on Measurement*, pages 228–231, May 2019.

[51] D. Kushner. The making of arduino. *IEEE spectrum*, 26:1–7, 2011.

[52] K. Kyriakou, B. Resch, G. Sagl, A. Petutschnig, C. Werner, D. Niederseer, M. Liedlgruber, F. Wilhelm, T. Osborne, and J. Pykett. Detecting moments of stress from measurements of wearable physiological sensors. *Sensors*, 19(17):3805, 2019.

[53] I. Labuschagne, C. Grace, P. Rendell, G. Terrett, and M. Heinrichs. An introductory guide to conducting the trier social stress test. *Neuroscience & Biobehavioral Reviews*, 107:686–695, 2019.

[54] C. L. Lim, C. Rennie, R. J. Barry, H. Bahramali, I. Lazzaro, B. Manor, and E. Gordon. Decomposing skin conductance into tonic and phasic components. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, 25(2):97–109, Feb. 1997.

[55] P. B. Madalena Proença. Are bitalino sensor compatible with arduino?, 2022.

[56] D. Makowski, T. Pham, Z. J. Lau, J. C. Brammer, F. Lespinasse, H. Pham, C. Schölzel, and S. Chen. Neurokit2: A python toolbox for neurophysiological signal processing. *Behavior research methods*, 53(4):1689–1696, 2021.

[57] E. Mejía-Mejía, J. Allen, K. Budidha, C. El-Hajj, P. A. Kyriacou, and P. H. Charlton. 4 - Photoplethysmography signal processing and synthesis. pages 69–146. Academic Press, Jan. 2022.

[58] E. Mejía-Mejía, J. M. May, R. Torres, and P. A. Kyriacou. Pulse rate variability in cardiovascular health: a review on its applications and relationship with heart rate variability. *Physiological Measurement*, 41(7):07TR01, Aug. 2020.

[59] microcontrollerslabhub.com. Monitor heart rate using pulse sensor and arduino, 2022.

[60] V. Mishra, G. Pope, S. Lord, S. Lewia, B. Lowens, K. Caine, S. Sen, R. Halter, and D. Kotz. Continuous detection of physiological stress with commodity hardware. *ACM transactions on computing for healthcare*, 1(2):1–30, 2020.

[61] O. M. Mozos, V. Sandulescu, S. Andrews, D. Ellis, N. Bellotto, R. Dobrescu, and J. M. Ferrandez. Stress detection using wearable physiological and sociometric sensors. *International journal of neural systems*, 27(02):1650041, 2017.

[62] P. Narczyk, K. Siwiec, and W. A. Pleskacz. Precision human body temperature measurement based on thermistor sensor. In *2016 IEEE 19th International Symposium on Design and Diagnostics of Electronic Circuits & Systems (DDECS)*, pages 1–5, Apr. 2016.

[63] K. Palanisamy, M. M, and S. Yaacob. Multiple Physiological Signal-Based Human Stress Identification Using Non-Linear Classifiers. *Electronics and Electrical Engineering*, 19, Sept. 2013.

[64] reichelt electronik. Store page: Grv emg detector arduino - emg-detektor.

[65] reichelt electronik. Store page: Xd p330.741 fitnesstracker.

[66] P. Rice, S. Upasham, B. Jagannath, R. Manuel, M. Pali, and S. Prasad. Cortiwatch: Watch-based cortisol tracker. *Future Science OA*, 5(9):FSO416, 2019.

[67] J. Rodriguez-Arce, L. Lara-Flores, O. Portillo-Rodriguez, and R. Martinez-Mendez. Towards an anxiety and stress recognition system for academic environments based on physiological features. *Computer methods and programs in biomedicine*, 190:105408, 2020.

[68] B. Rosner. *Fundamentals of Biostatistics*. Cengage Learning, July 2015. Google-Books-ID: yn4yBgAAQBAJ.

[69] G. Russell and S. Lightman. The human stress response. *Nature Reviews Endocrinology*, 15(9):525–534, Sept. 2019.

[70] J. L. Salinet and O. Luppi Silva. Chapter 2 - ECG Signal Acquisition Systems. pages 29–51. Academic Press, Jan. 2019.

[71] I. Schelkanova, A. Pandya, A. Muhaseen, G. Saiko, and A. Douplik. 13 - Early optical diagnosis of pressure ulcers. Woodhead Publishing Series in Biomaterials, pages 347–375. Woodhead Publishing, Jan. 2015.

[72] L. Schwabe, L. Haddad, and H. Schachinger. HPA axis activation by a socially evaluated cold-pressor test. *Psychoneuroendocrinology*, 33(6):890–895, July 2008.

[73] I. Seeed Studio. Grove - emg detector, 2008-2021.

[74] I. Seeed Studio. Grove - gsr sensor, 2021.

[75] F. Shaffer and J. P. Ginsberg. An overview of heart rate variability metrics and norms. *Frontiers in public health*, page 258, 2017.

[76] N. N. Shah, M. L. Schwandt, B. Hobden, D. S. Baldwin, J. Sinclair, R. Agabio, and L. Leggio. The validity of the statetrait anxiety inventory and the brief scale for anxiety in an inpatient sample with alcohol use disorder. *Addiction*, 116(11):3055–3068, 2021.

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

[78] Y. Shi, N. Ruiz, R. Taib, E. Choi, and F. Chen. Galvanic skin response (GSR) as an index of cognitive load. In *CHI '07 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '07, pages 2651–2656, New York, NY, USA, Apr. 2007. Association for Computing Machinery.

[79] V. Shusterman and O. Barnea. Sympathetic nervous system activity in stress and biofeedback relaxation. Monitoring SNS activity with the photoplethysmographic-wave envelope and temperature-variability signals. *IEEE engineering in medicine and biology magazine: the quarterly magazine of the Engineering in Medicine & Biology Society*, 24(2):52–57, 2005.

[80] T. Smeets, S. Cornelisse, C. W. E. M. Quaedflieg, T. Meyer, M. Jelicic, and H. Merckelbach. Introducing the Maastricht Acute Stress Test (MAST): A quick and non-invasive approach to elicit robust autonomic and glucocorticoid stress responses. *Psychoneuroendocrinology*, 37(12):1998–2008, Dec. 2012.

[81] D. Stegeman and H. Hermens. Standards for suface electromyography: The european project surface emg for non-invasive assessment of muscles (seniam). 1, 01 2007.

[82] L. Sörnmo and P. Laguna. Chapter 5 - The Electromyogram. Biomedical Engineering, pages 337–410. Academic Press, Burlington, Jan. 2005.

[83] L. Sörnmo and P. Laguna. Chapter 7 - ECG Signal Processing. Biomedical Engineering, pages 453–566. Academic Press, Burlington, Jan. 2005.

[84] D. E. Tadlock. Avionics safety. In *Safety Design for Space Systems*, pages 403–474. Elsevier, 2009.

[85] I. Taneja, M. S. Medow, D. A. Clarke, A. J. Ocon, and J. M. Stewart. Postural change alters autonomic responses to breath-holding. *Clinical Autonomic Research: Official Journal of the Clinical Autonomic Research Society*, 20(2):65–72, Apr. 2010.

[86] A. Tazarv, S. Labbaf, S. M. Reich, N. Dutt, A. M. Rahmani, and M. Levorato. Personalized stress monitoring using wearable sensors in everyday settings. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 7332–7335. IEEE, 2021.

[87] C. Thompson. Build it. share it. profit. can open source hardware work. *Work*, 10(08), 2011.

[88] E. Vavrinsky, V. Stopjakova, M. Kopani, and H. Kosnacova. The concept of advanced multi-sensor monitoring of human stress. *Sensors*, 21(10):3499, 2021.

[89] H. Vedder. Physiology of the HypothalamicPituitaryAdrenocortical Axis. volume 7 of *The Hypothalamus-Pituitary-Adrenal Axis*, pages 17–31. Elsevier, Jan. 2007.

[90] C. H. Vinkers, R. Penning, J. Hellhammer, J. C. Verster, J. H. Klaessens, B. Olivier, and C. J. Kalkman. The effect of stress on core and peripheral body temperature in humans. *Stress*, 16(5):520–530, 2013.

[91] J. Wijsman, B. Grundlehner, H. Liu, J. Penders, and H. Hermens. Wearable Physiological Sensors Reflect Mental Stress State in Office-Like Situations. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, pages 600–605, Sept. 2013. ISSN: 2156-8111.

[92] N. Y. World Famous Electroncs LLC, Brooklyn. How does the pulse sensor work and how to interface it with arduino?, 2015-2017.

[93] WorldFamousElectronics. Pulsesensorplayground, 2016.

[94] XD Connect. *Manual P330.74X*.

[95] S. Yoon, J. K. Sim, and Y.-H. Cho. A flexible and wearable human stress monitoring patch. *Scientific reports*, 6(1):1–11, 2016.

A. Appendix

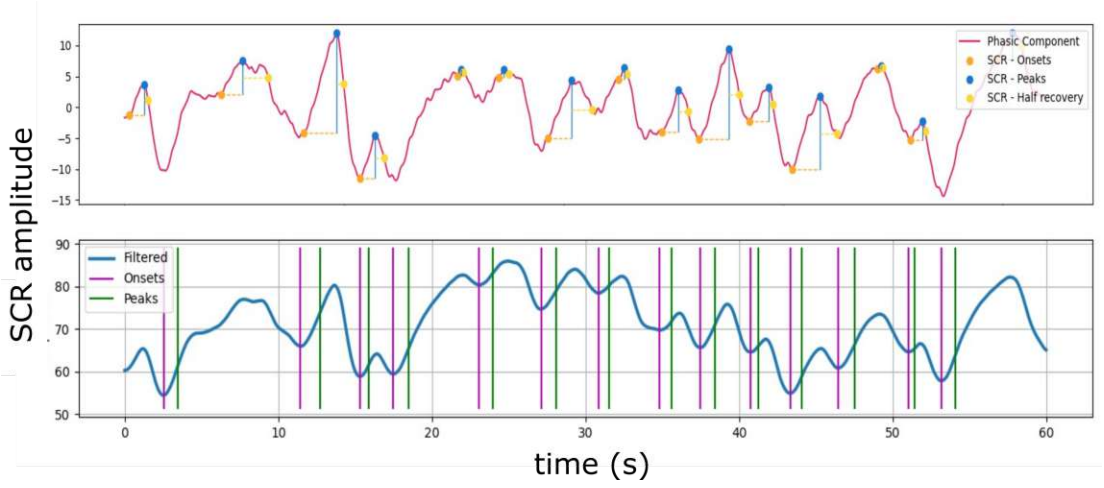

Figure A.1.: Difference in peak finding using Neurokit2 and BioSPPy seen on random cuts of task and rest phases. A clear difference in found peaks can be seen.

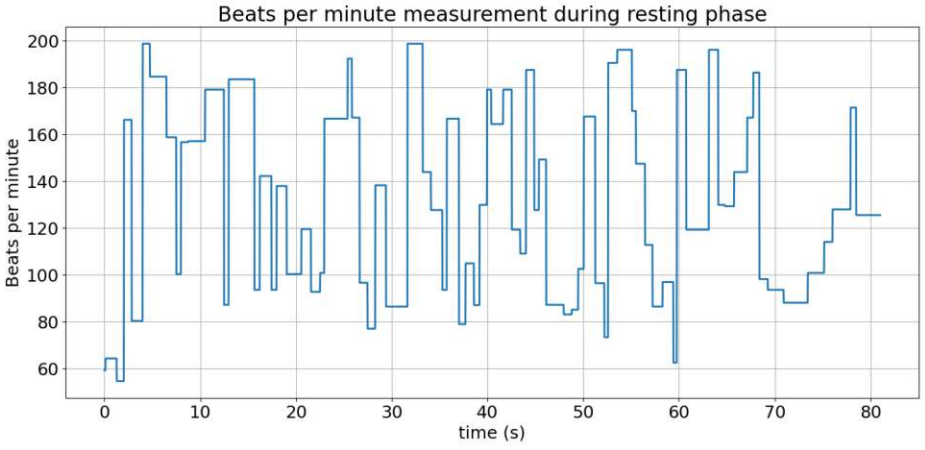

Figure A.2.: BioData's beats per minute function tested on resting state.

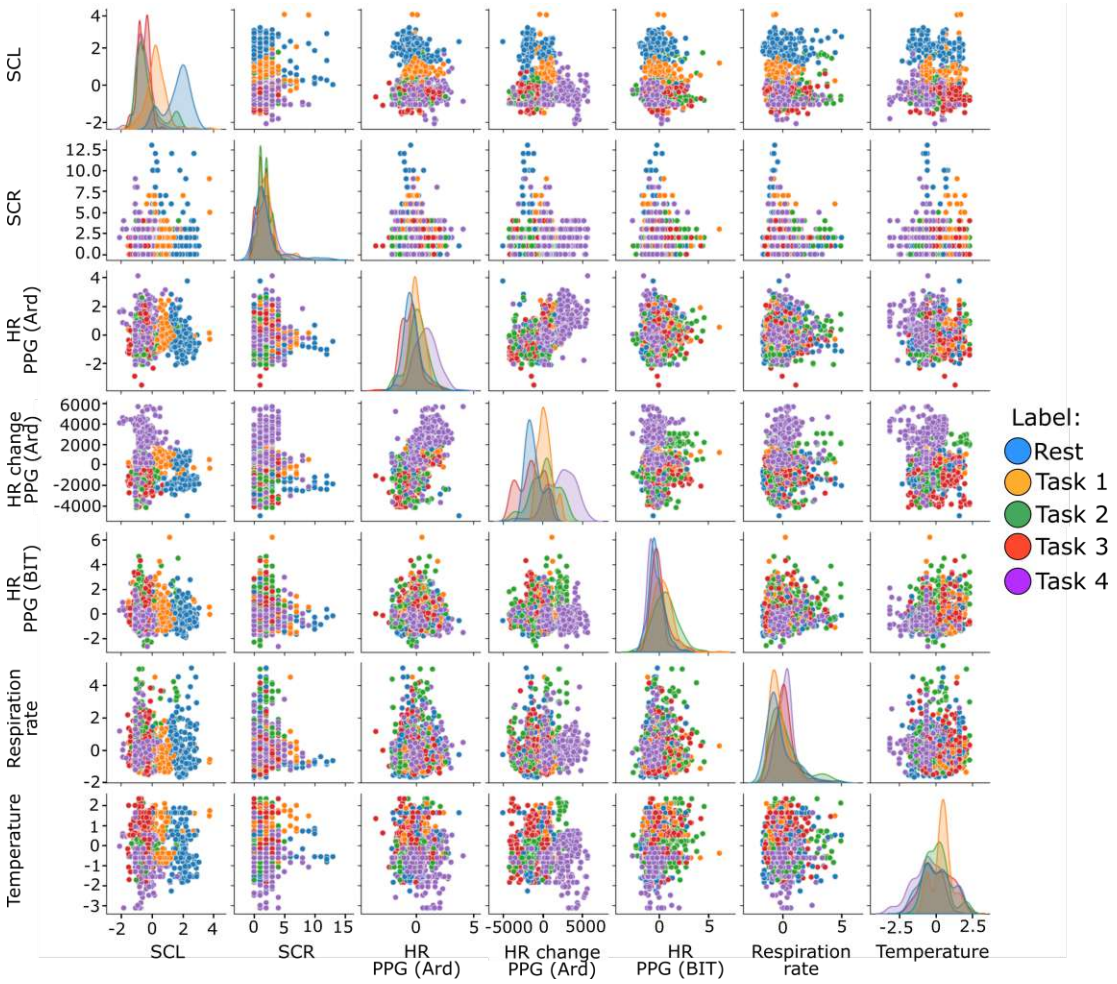

Figure A.3.: Pair-Plot of Selected Features for all Task Phases separately in comparison to the initial Resting Phase. The window size for skin conductance response (SCR) is 10 seconds, and for heart rate (HR) change is 40 seconds.

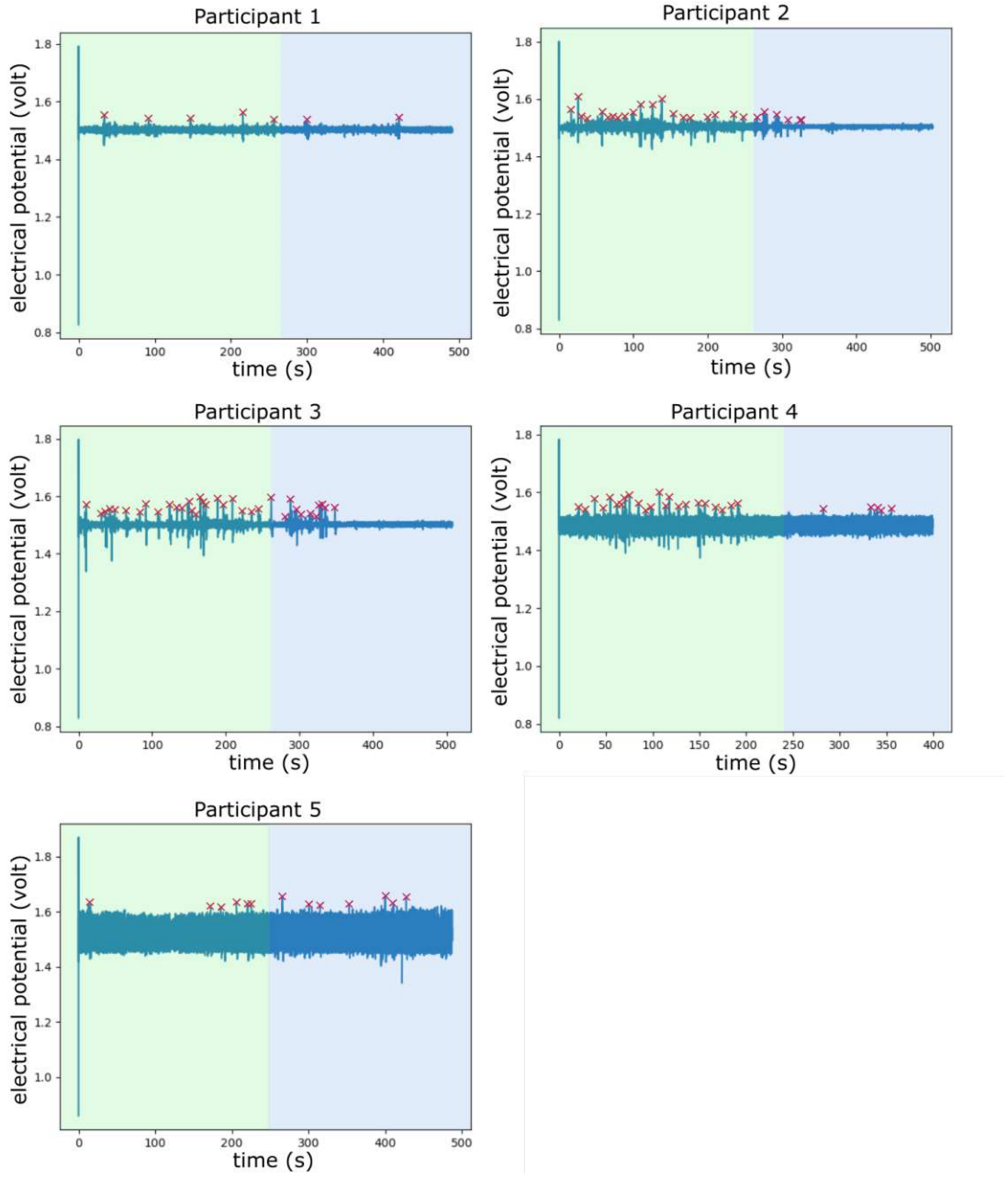

Figure A.4.: Filtered electromyography (EMG) datasets with manually marked events. The task duration is marked in green, the resting period is shadowed in blue.

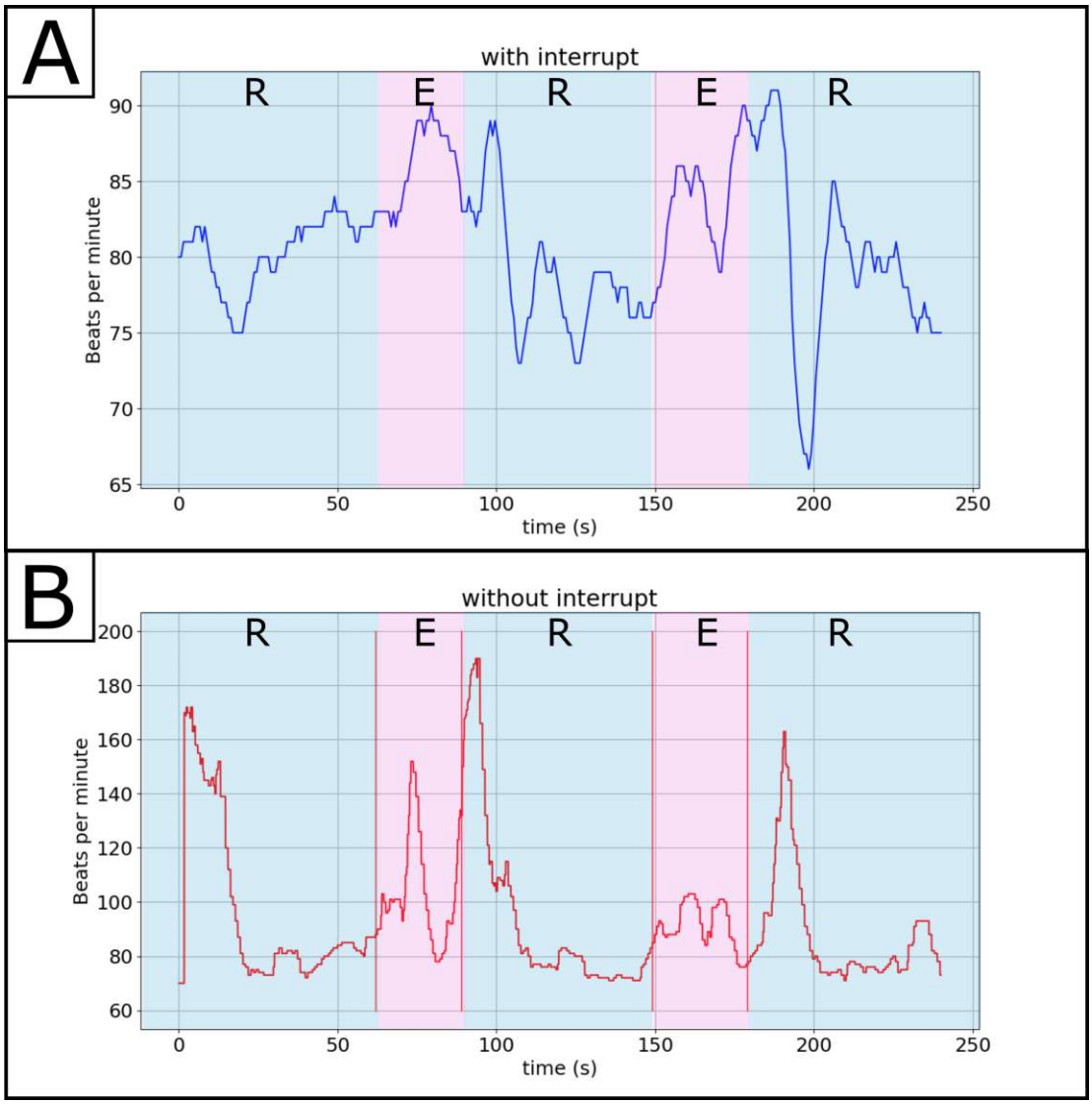

Figure A.5.: PulseSensor Playground library tested using the libraries example project with (A) and without (B) interrupt function to calculate heart rate on breathing pattern of alternating normal breathing (R, blue) and breath-holding (E, magenta).

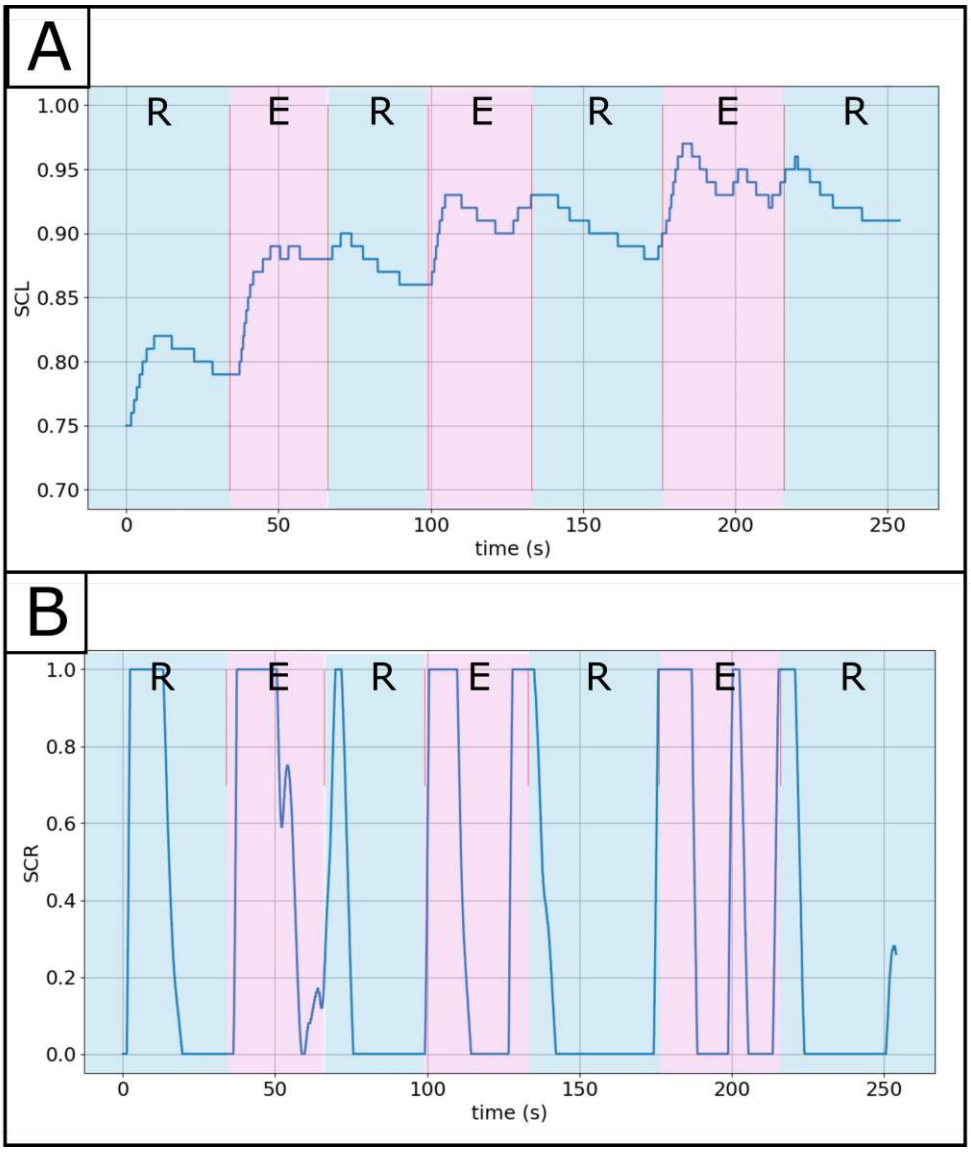

Figure A.6.: BioData's electrodermal activity functions skin conductance level (SCL) (A) and skin conductance response (SCR) (B) tested on breathing pattern of alternating normal breathing (R, blue) and breath-holding (E, magenta).

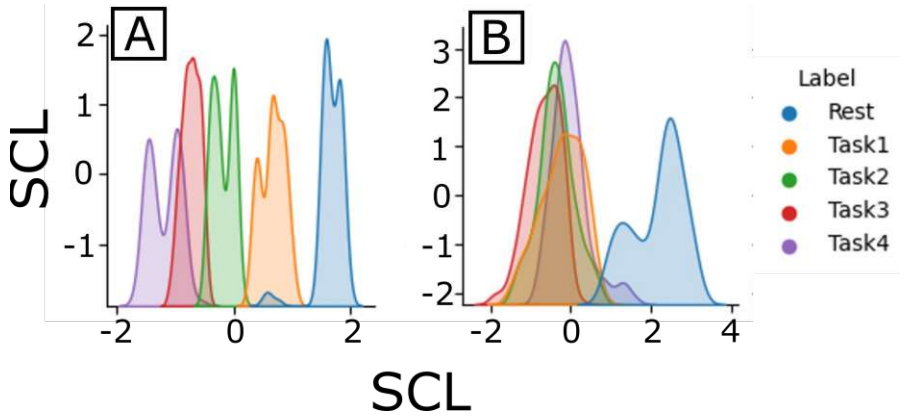

Figure A.7.: Distribution of skin conductance level (SCL) data of all testing phases for an electrodermal activity (EDA)-non-responder (A) compared to a participant with EDA response (B).

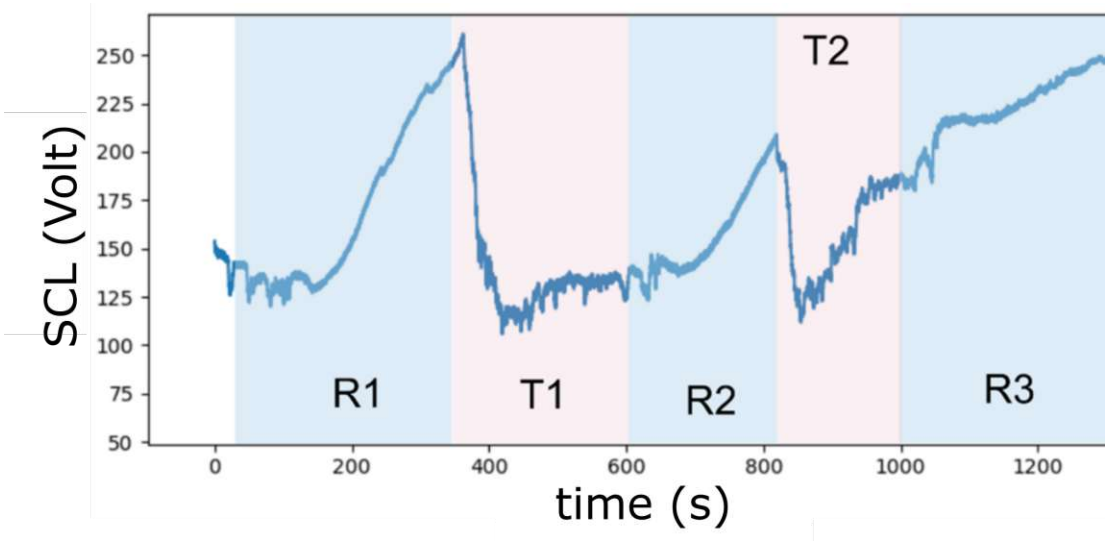

Figure A.8.: Curve progression of skin conductance level (SCL) during part of the stress testing performed on participants in July 2022.