Comparison in stress-related features between laboratory and daily-life settings in healthy subjects: A multi-sensor approach

Laura Schiphorst, BSc.

Department of Biomedical Signals and Systems
Faculty of Science and Technology
University of Twente
Enschede, The Netherlands

Supervisory committee

Chairman:	Prof.Dr.ir. P.H. Veltink
Second supervisor:	Dr. ir. S.U.Yavuz
Daily supervisor:	Carlijn Braem, MSc.
External supervisor:	Dr. M.L. Noordzij

Date: 18-08-2022

Acknowledgements

This thesis would have not been possible without the help of a number of people. Firstly, I want to thank my daily supervisor Carlijn for the support and feedback throughout my thesis. Our meetings were helpful to bring my work to a higher level. Secondly, I want to thank Dr. Utku Yavuz for his valuable feedback and his enthusiasm attitude and for being my supervisor in the Master Research Honours program. Third, I want to thank prof.Dr. Peter Veltink for being the chairman of my committee and his valuable feedback. Furthermore, I want to thank Matthijs Noordzij for being my external committee member and for his constructive feedback and enthusiasm of the project. I am very grateful that you all made time for me despite your busy schedules. Besides, I want to thank Niala den Braber for her feedback during the weekly meetings and Anne-Linde and Linde for the feedback provided on my thesis. Finally, I want to take this opportunity to thank my family and friends for the support throughout my thesis.

Samenvatting

Ambulante detectie van stress is belangrijk bij het voorkomen van chronische ziektes. In deze studie is er een raamwerk ontwikkeld voor het verwerken van fysiologische signalen en onderzocht hoe de stress gerelateerde kenmerken van de signalen variëren tussen laboratorium stressoren en stressoren in het dagelijks leven. De stressreacties werden gemeten met de draagbare sensor Empatica E4, die een hartslag (PPG), huidgeleiding (SC) en huidtemperatuur (ST) meet. De laboratorium meting bestond uit het geven van een presentatie, hoofdrekenen en de Stroop Colour Word Test. In het dagelijks leven werd stress gemeten met korte vragenlijsten (Ecological Momentary Assessment) gedurende twee weken. Lineaire modellen zijn gemaakt voor het laboratorium en het dagelijks leven voor elke deelnemer op basis van de stress gerelateerde kenmerken en de zelfrapportages. De resultaten wijzen in de richting dat de stress gerelateerde kenmerken verschillen tussen het laboratorium en het dagelijks leven en individuen. Dit suggereert de behoefte aan geïndividualiseerde modellen voor stressdetectie en om de informatie uit laboratorium- en dagelijkse metingen als complementair te bekijken. Daarnaast werd een raamwerk gecreëerd voor stressdetectie met een multi-sensor benadering, die filters en kwaliteitsindicatoren bevat om de kwaliteit van de signalen te beoordelen.

Contents

Acknowledgements 2
Samenvatting 3
List of Abbreviations 6
Chapter 1: Research article 8
I. INTRODUCTION 9
II. Methods 10
II.1 Study population 10
II.2 Physiological recording 11
II.3 Study procedures 11
II.4 Data analysis 13
III. Results 14
IV. Discussion 17
V. Conclusion 19
References 19
Appendices 22
A.1 Ecological Momentary Assessment (EMA) 22
A.2 Median vs. Butterworth filter 24
A.3 Savitzky-Golay filter 27
A.4 Quality indicators 28
A.5 Results of linear model with absolute difference with 95% CI 30
A.5.1 Laboratory results of the linear models 30
A.5.2 Daily-life results of the current stress level of the linear models 31
A.5.3 Daily-life results of the past one hour stress level of the linear models 33
Chapter 2: Background 36
2.1 Introduction 36
2.2 Definition of stress 36
2.3 Physiological mechanisms of stress response 36
2.4 Physiological signals related to stress 38
2.4.1 Photoplethysmography 40
2.4.2 Skin conductance 41
2.4.3 Skin temperature 42
2.4.4 Confounding factors to stress response 42
2.5 State of the Art 43

2.6 Scope of the thesis 44
References.. 45

List of Abbreviations

ANS	**A**utonomic **N**ervous **S**ystem
PSNS	**P**ara**s**ympathetic **N**ervous **S**ystem
SNS	**S**ympathetic **N**ervous **S**ystem
PPG	**P**hoto**p**lethysmo**g**raphy
PVN	**P**ara**v**entricular **N**ucleus
SAM	**S**ympathetic **A**drenal **M**edullary **S**ystem
CRF	**C**orticotropin-**R**eleasing **F**actor
AVP	**A**rginine **V**aso**p**ressin
ACTH	**A**drenoco**r**tico**t**ropic **H**ormone
HPA	**H**ypothalamic-**P**ituitary-**A**drenal
PPG	**P**hoto**p**lethysmo**g**raphy
HR	**H**eart **R**ate
HRV	**H**eart **R**ate **V**ariability
RMSSD	**R**oot **M**ean **S**quared of the **S**quared **D**ifferences
SDNN	**S**tandard **D**eviation of **N**ormal-to-**N**ormal intervals
pNN50	**P**roportion of the successive **N**ormal-to-**N**ormal intervals that differ more than **50** ms
MIST	**M**ontreal **I**maging **S**tress **T**est
LF	**L**ow **F**requency
HF	**H**igh **F**requency
LFHF	Ratio of **L**ow **F**requency band and **H**igh **F**requency band
SC	**S**kin **C**onductance
SCL	**S**kin **C**onductance **L**evel
SCR	**S**kin **C**onductance **R**esponse
SCWT	**S**troop **C**olour **W**ord **T**est
ST	**S**kin **T**emperature
SD ST	**S**tandard **D**eviation of **S**kin **T**emperature
BMI	**B**ody **M**ass **I**ndex
STEADY	S**t**ress m**e**asurements in l**a**boratory and **D**ail**y**-life
ACC	**Acc**eleration
EMA	**E**cological **M**omentary **A**ssessment
DSI	**D**aily **S**tress **I**nventory
PSS	**P**erceived **S**tress **S**cale
SG	**S**avitzky-**G**olay
mHR	**M**edian **H**eart **R**ate
SDNN	Interquartile range of the **N**ormal-to-**N**ormal intervals
pNN50	**P**roportion of successive **n**ormal-to-**n**ormal intervals that differ more than **50** ms
HF	**H**igh **F**requency band
LF/HF	Ratio of **l**ow and **h**igh frequency band
mSCL	**M**edian **S**kin **C**onductance **L**evel
minSCL	**M**inimum **S**kin **C**onductance **L**evel
Absdiff2	Signal power of skin conductance in second difference
SCRR	**S**kin **C**onductance **R**esponse **R**ate
SCR dur	**S**kin **C**onductance **R**esponse **dur**ation
SCR amp	**S**kin **C**onductance **R**esponse **amp**litude

SCR pow	**S**kin **C**onductance **R**esponse **pow**er
mST	**M**edian **S**kin **T**emperature
SD ST	Interquartile range of **S**kin **T**emperature
ST slope	Slope of **S**kin **T**emperature
RMSSD	**R**oot of the **m**edian of the **s**um of **s**quares of **d**ifferences in normal-to-normal intervals
IQR	**I**nter**q**uartile **R**ange

Chapter 1: Research article

Comparison in stress-related features between laboratory and daily-life settings in healthy subjects: A multi-sensor approach

Schiphorst, L.R.B., Braem, C.I.R., Yavuz, U.S, Noordzij, M.L., Veltink, P.H.

***Abstract*— Ambulatory detection of stress is important in preventing chronic diseases. In this study, we created a multi-sensor processing framework and investigated how the stress-related features vary between laboratory and daily-life stressors. The stress responses were measured with the wearable Empatica E4, containing photoplethysmography (PPG), skin conductance (SC) and skin temperature (ST). The laboratory measurement consisted of a public speaking task, a mental arithmetic task and the Stroop Colour Word Test. Stress in daily-life was measured for two weeks with Ecological Momentary Assessment. Linear models for the laboratory and daily-life stressor were designed for each participant based on the stress-related features and self-reports. The results point into the direction that the stress associated features are different between laboratory and daily-life settings and individuals. This suggests the need for individualized models for stress detection and for viewing the information from laboratory and daily-life measurements as complementary. Next to that, a framework was created for stress detection with a multi-sensor approach, which contained pre-processing steps as filtering and quality indicators to assess the quality of the signals.**

***Index Terms*—Multi-sensor, stress, physiological signals.**

I. Introduction

Chronic diseases, such as diabetes and obesity, are the main cause of mortality and are responsible for 70% of all the deaths worldwide. One major cause of these chronic diseases is an unhealthy lifestyle, characterized by for example an unhealthy diet, physical inactivity, excessive alcohol consumption and smoking [1,2]. In addition, chronic stress is a common risk factor for 75-90% of chronic diseases [3]. Therefore, early detection and prevention of the negative effects of stress are important in preventing chronic diseases and improving the quality of life of patients with chronic diseases [4].

There is no universal definition of stress [5]. The definition of stress in this study is taken from Liu et al. [3](p.1): “Stress is a state of threatened homeostasis provoked by a psychological, environmental, or physiological stressor.” This definition is chosen since it covers the three approaches for defining stress (environmental, physiological and psychological) present in literature. The stress responses occur as a reaction to stressors, which can be defined as stimuli that can disturb the equilibrium inside the body [5].

A distinction can be made between acute and chronic stress. Acute stress arises from an incident and endures for a short period of time, such as having to finish work before a deadline [5]. Chronic stress arises due to a prolonged imbalance and induces negative mental and physical symptoms, influencing blood pressure and cholesterol levels [5].

Currently, the gold standard to assess stress is by the use of questionnaires, such as the Perceived Stress Scale (PSS) [4,6]. However, questionnaires are time-consuming to fill in and provide only a momentary retrospective snapshot of the stress level of an individual, which limits the accuracy in determining and monitoring stress [4,6-7]. On the other hand, it is well established that stress activates the sympathetic nervous system (SNS), which causes particular physiological responses such as an increase in heart rate and an increase in sweat production [5]. These physiological responses can be defined as the response of the body to prepare for surviving the threat, it activates the sympathetic–adrenomedullary (SAM) system and hypothalamic–pituitary–adrenocortical (HPA) axis [8]. These responses can be continuously measured through wearable sensors in daily-life [5]. The responses reflect primarily acute stress, but it may prolong the response to a chronic stress response [8].

Previous research has measured physiological responses to capture stress in the laboratory and in daily-life [4,5]. In the laboratory setting, stress responses in physiological signals from wearable sensors are already shown to be associated with chronic and acute stress [7,10]. Also, high classification accuracies were achieved for stressed moments in the laboratory based on physiological responses [9,11]. This

shows that stress induced in laboratory settings can be captured through physiological responses.

The physiological stress responses are extensively researched in the laboratory, but the application of laboratory measurements in the real world is limited. With the rise of the wearable sensors, the focus of researching stress has shifted from laboratory towards stress detection in ambulatory settings, to improve lifestyle and prevent chronic diseases [4].

The past research in ambulatory stress detection can be divided into semi-ambulant, which is observing stress in a specific situation, and ambulantory, which is monitoring in the daily-life setting [4,7,13]. Healey and Picard for example showed that skin conductance and heart rate had the highest association with the stress level of the participants while driving [13]. The SWEET study by Smets et al., showed associations between physiological signals and stress based on self-reports in daily-life [4]. Although these results show that it is promising to use physiological signals in daily-life for stress detection, the classification accuracies in daily-life are lower compared to laboratory settings [14].

Next to the lower classification accuracies, there are other challenges to overcome in stress detection in daily-life. Examples are reduced signal quality, confounding factors, lack of gold standard and person-specific stress responses [7,11]. The signal quality is reduced in daily-life due to influences of surroundings, motion artefacts or improper use of the device. So, signal quality indicators to detect artefacts are needed to ensure the quality of physiological signals in the analysis. Previous research has created methods to detect artefacts, but one framework to assess quality of multiple physiological signals is lacking. As a multi-sensor approach was shown to surpass one signal approaches in stress detection [7]. Another important challenge is the influence of physical activity, since it activates the sympathetic nervous system like the stress response system and the quality of the signals [7]. Another challenge is that there is no gold standard in daily-life settings to compare the results of the physiological signals to. To overcome the challenge of limited gold standard, research relies for now on self-reported questionnaires [4,5]. Moreover, Smets et al. [4] showed with a large-scale research in an ambulatory setting that stress responses tend to be person-specific, so personalized approaches are needed.

The person-specific stress response suggests the need for personalized techniques, such as personal models, for stress detection in an ambulatory setting [7,15]. Smets et al., created clusters on Montreal Imaging Stress Test (MIST) in daily-life to personalize models [9]. The results imply that the type of features, characteristics of physiological signals, that are associated with a stress test in daily-life and daily-life stressors can be similar. This suggests that features may be a promising tool to improve the classification accuracy [9]. In addition, the results show that the stress test may be useful for personalized calibration, to capture individual physiological information [9]. Although the results indicate that a stress test could be used for normalization and clustering of features, the models with information from the stress test showed low classification accuracies. Hovspian et al. created a model trained on laboratory data and applied to laboratory and field data, giving a prediction accuracy of 90% for laboratory data and 72% for the field data [11]. These models built in the laboratory show lower accuracies in daily-life. Thus, more information about the relation between the stress responses during laboratory stressors, a stress test in daily-life and daily-life stressors is needed, to know how the accuracy of stress classification can be improved.

To the best of our knowledge, a framework to analyze multiple physiological signals is missing and the stress responses from laboratory stressors, a stress test and daily-life stressors have not been compared before. The wearable multi-sensor used in this study is Empatica E4, which measures blood volume pulse, skin conductance and skin temperature. The stress test used in daily-life in this study is the (computerized) Stroop Colour Word Test (SCWT), which has been shown to be associated with stress by previous studies [16,17]. The research captures two aims: (1) create a framework for a multi-sensor approach (PPG, SC, ST and ACC) and (2) compare stress-related features from the wearable sensors between laboratory stressors, SCWT and daily-life stressors in healthy individuals.

II. Methods

II.1 Study population

This study was approved by ethics committee of Behavioural, Management and Social sciences at University of Twente (220397). STEADY (STress mEasurements in lAboratory and DailY-life) is an exploratory study in which six healthy subjects were included. The STEADY study ran from 18-05-2022 and is still including participants at the University of Twente. All participants provided their consent before participation.

One inclusion criterium was being between the age of 18 and 65 years, since ageing can influence physiological signals [18]. Further inclusion criteria were being a Dutch- or English-speaker and having a job or study that primarily requires sitting during the day, to decrease the influence of physical activity [7]. Exclusion criteria of this study are: diagnoses of cardiovascular disease, diagnoses of mental disorders like a depression, use of medications that have an effect on the nervous system or cardiovascular system, Having diabetes, BMI > 25, problems with blood pressure, or pregnancy [5]. Participants received compensation for participation in the study based on the answer ratio of the EMA.

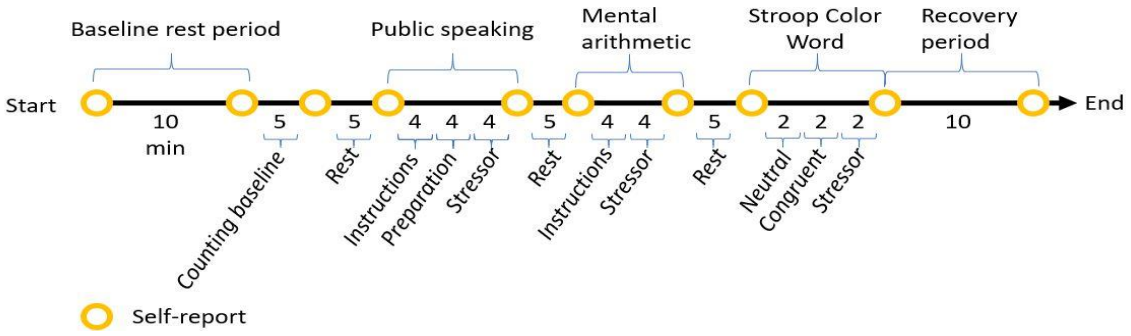


Figure 1: Laboratory protocol of STEADY, based on Plarre et al [21]. Three stressors (public speaking, mental arithmetic and Stroop Colour Word) are presented with resting moments and a baseline counting task. The sequence of the stressor was randomized for each participant.

II.2 Physiological recording

The Empatica E4 wristband (Empatica Srl, Milan, Italy) was used in this study to record multiple physiological signals. Photoplethysmography (PPG) was recorded at 64 Hz, skin conductance (SC) at 4 Hz, skin temperature (ST) at 4 Hz and 3-axis acceleration (ACC) at 32 Hz [19]. Participants were instructed to put the wristband on their non-dominant hand [20].

II.3 Study procedures

The study protocol consisted of the laboratory and daily-life measurements, in which the measurements lasted two hours and two weeks, respectively. Before the laboratory measurements started, subjects provided baseline characteristics, such as age, length, and weight. In order to prevent the influence of confounding variables on the measurement, participants were not allowed to smoke, perform any vigorous physical activity (e.g. running) or have any intake of caffeine four hours prior to the laboratory measurement. Also, the use of alcohol had to be avoided one day prior to the measurement and pain medication should not be taken 72 hours before the measurement [21]. The participants put on the Empatica E4 wristband 15 minutes prior to the laboratory measurements, in order to let sensors in the wristband adjust to the skin [22].

II.3.1 Laboratory measurement

The laboratory protocol was similar to the protocol from Plarre et al. [21]. An overview of the protocol can be found in Figure 1. At the start of the measurement, there was a ten-minute baseline period in which the participants watched a nature documentary to be in a non-stressed state [21]. After that, there was a counting baseline of five minutes in which the participants counted out-loud from one till fifty repeatedly, to capture the change in physiology due to speaking [11]. After the counting task there was a rest period of five minutes so the physiological signals returned partially to baseline [21].

The laboratory measurement contained three stressors, for which the order was randomized. One stressor was the public speaking task, in which participants had four minutes to prepare a presentation about a given topic and then had four minutes to give the presentation. The participants were told that the presentation was recorded to induce more stress [21]. Another stressor was a mental arithmetic test, in which participants had to add the sum of a three-digit number to another number [21]. The researcher intervened by saying 'wrong' when the participant answered wrongly, but the researcher also stated 'wrong' if the participant answered about five times in a row correctly to induce extra stress. The last stressor was the SCWT, that participants performed in daily-life settings as well. It consisted of three phases: neutral, congruent and incongruent [23]. In the neutral phase '###' are shown in different colors and the participants named the right color. In the congruent phase the colors will match the meaning of the text, such as 'black'. During these two phases no time or feedback was given. In the last phase incongruent words and colors are presented, with feedback and time pressure. Each phase takes two minutes.

During the laboratory measurement, the participants wore the Empatica E4 continuously. Next to that, the participants filled in self-reports before and after each task or rest period [18]. The self-reports were incorporated to capture how stressed the participants felt during the different phases of the measurement. In the self-reports, the participants had to rate their current stress level on a Likert scale of one to five (S1-S5).

II.3.2 Daily-life measurement

The day after the laboratory assessment, the daily-life measurement started. The participants wore the Empatica E4 continuously for two weeks. Next to that, during the two week period participants performed the SCWT in EthicaData (Ethica, Toronto, Canada) twice. It had the same format as in the laboratory setting. Moreover, the participants received a notification from the EthicaData app to fill in the Ecological Momentary Assessment (EMA). The EMA was designed similar to Smets et al. [4] and consisted of the Self Assessment Manikin (SAM) score, stress level and confounding factors [4]. The SAM score measures emotions linked to stress, which helps to distinguish positive stress from negative stress. The

second part of the SAM asked the current and past 1-hour stress level. With the current stress level the physiological features can be determined over a shorter time frame which may be more accurate [5,9]. The stress level over the past hour reduces the chance of missing any stressed moments. These questions were asked on a Likert scale from one to five, ranging from being not stressed to extremely stressed [4]. The third part on confounding factor was included to enable

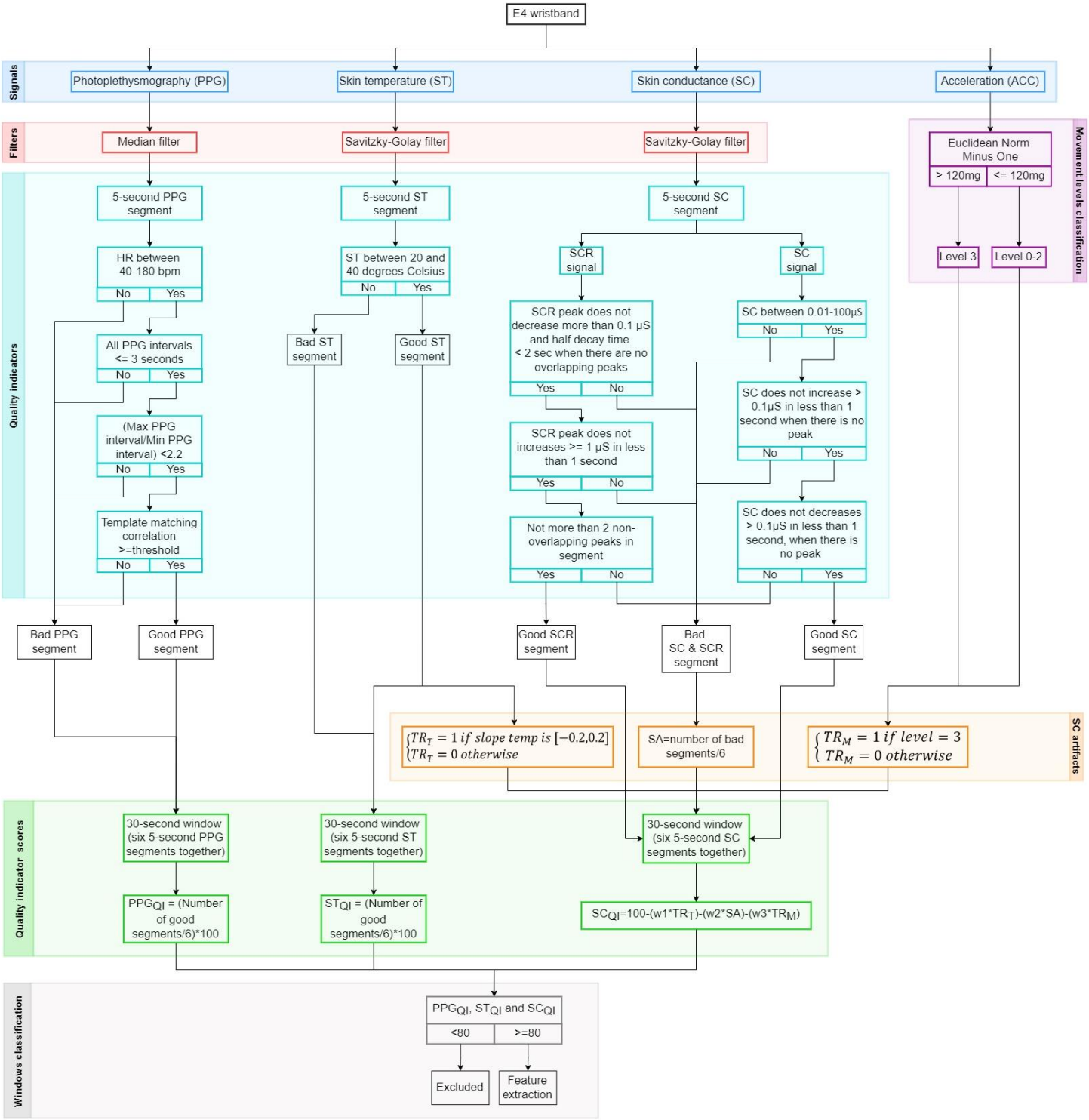


Figure 2: Pre-processing steps of the physiological signals containing signals (blue), filters (red), quality indicators (light-blue), movement levels classification (purple), SC artifact detection (orange) and quality indicators scores (green) . PPG is filtered with median filter and subsequently passed through quality indicators and template matching, leading to a quality score (PPG_{QI}). ST is filtered by SG-filter and passed through quality indicators, after which the slope is also input for SC quality (TR_T) and the quality score of ST is determined (ST_{QI}). SC is filtered by SG-filter and also passed quality indicators. Then a quality score for SC is made (SC_{QI}) based on TR_T, number of shape artefacts (SA) and norm of acceleration (TR_M). Only windows with 80% of data of the three signals are used in further steps.

indicating when the participant had any food, caffeine or alcohol intake during the past hour, if they had any [4]. Also, current and past 1-hour performed physical activities were registered with a predefined list to aid recall. The questions are presented in Appendix A.1.

The EMAs were prompted randomly in the following time blocks: 07:00 - 08:00, 09:30 - 10:30, 12:00 - 13:00, 14:30 - 15:30, 17:00 - 18:00, 19:30 - 20:30. The EMAs were prompted six times, since that has been shown to lead to compliance rates of more than 80% [24-26]. Participants had thirty minutes to answer the questions.

Next to the EMAs, there was a daily questionnaire about the types of stressors that occurred the last 24 hours called the Daily Stress Inventory (DSI) [27]. At the last day of the measurement in daily life, the participants had to fill in the Perceived Stress Scale (PSS) questionnaire [28].

II.4 Data analysis

II.4.1 Pre-processing

The raw data obtained from Empatica E4 was first synchronized with the self-reports from the lab or self-reports of the EMA questionnaire by using Unix timestamp. After that, the signals were pre-processed to filter out artifacts due to motion and external influences.

The PPG signal was filtered by a 9-sample median filter [29], and detrended. The median filter is a non-linear filter, which replaces each sample with the median of surrounding values set by a window size [29]. More information on the filter can be found in Appendix A.2. In addition, quality indicators developed by Orphanidou et al. [30] were used to assess the quality of PPG signal for every 5 second segment. Three rules were evaluated: (1) Heart rate (HR) had to be between 40 and 180 beats per minute, (2) the maximal time between two PPG pulses may not exceed three seconds and (3) the ratio of the maximal and minimal time between two PPG pulses had to be less than 2.2. Lastly, a correlation template matching procedure was performed, which classified the ten-second segment signal as good or bad quality [30,31]. The initial template was created based on the mean of the good pulses, based on the three quality indicators, during the baseline segment for laboratory measurement or on the night before (or after if night before was missing) for ambulatory measurement. The PPG sample was resampled to match the template. Both the template and PPG sample were normalized in amplitude prior to the template matching procedure [31]. The worst 25% of the correlated peaks was discarded per day. Examples of the good or bad quality segments are given in Appendix A.4.

The SC signal was first filtered by a Savitzky-Golay (SG) 5th-order filter and 11-sample window, since the SG filter smooths the SC signal without changing the signal properties [29,32]. More information on the filter can be found in Appendix A.3. Next to that, the artefact detection method was implemented on 5-second windows, for the following artifacts: (1) shape artefacts, (2) movement artefacts and (3) temperature-based artefacts. (1) were detected by quality indicators based on SC signal and peaks in the phasic component of SC signal [33]. Continuous Decomposition Analysis (CDA) of Ledalab was used to decompose the SC signal into phasic and tonic signals [34]. (2) were detected by determining the norm of acceleration for each window, resulting in four categories from inactive to vigorous [33]. (3) were determined based on the slope of the skin temperature signal of each window [32].

Every 30-second window receives a score on these three components: (1) was scored based on the ratio of shape artefacts in the window, (2) was given the score 1 if the movement is vigorous, else it is 0 and (3) was scored as 1 if the slope of temperature is outside [-0.2,0.2], 0 otherwise. This was the lowest threshold possible, since the accuracy of Empatica E4 is 0.2 degrees Celsius. Next, the weights of these components were set to 2/3, 1/3 and zero in laboratory and 2/3, 1/6 and 1/6 in daily-life. In laboratory criterium 3 was not taken into account, because all participants were in a controlled room so it is assumed that the temperature would not influence the quality of the signals. The weight of the shape artefacts was set to 2/3, meaning that if two shape artefacts were present in the window, the window was classified as bad similar to PPG. In daily-life the self-reports including moderate to high physical activity were excluded, so it can be expected that windows will not have moderate to high physical activity. In the end, this resulted in a score from 0-100 for data quality [33], only windows with a score of 80 or higher are included in further analysis. Examples of bad segments due to quality indicators are given in Appendix A.4.

The ST signal was also first filtered by a SG filter (order = 5, window = 9 samples) [29,32]. The quality of the ST signal is classified as bad if the ST is lower than 20 or higher than 40 degrees Celsius [4,32]. An example of a bad segments is provided in Appendix A.4.

II.4.2 Feature extraction

In total, 21 physiological features were determined from the recorded physiological signals: eight, ten, and three features from the PPG, EDA and ST signal, respectively (Table 1). These features were chosen based on previous research that has shown a significant association between these features and stress [4,5,7,11,12,35-45]. The median and interquartile range of the features was chosen based on the distribution of the features. A window of 30 seconds (six five-second segments) was chosen to calculate the features. Only the windows in which 80% of the data was good quality data, meaning five out of six five-second segments, according to the quality indicators and template matching were used [4]. After the features were extracted, the features were z-normalised per participant to remove variations in the data [4]. Next, a

correlation analysis was performed to determine correlated features per subject. If the features had a correlation coefficient of 0.8 or higher, the features were seen as correlated [31].

Table 1: Overview of features extracted from the three physiological signals, with references which are studies that used the features successfully for stress detection

Feature	Feature full name	References
mHR	Median heart rate	[4] [7] [35] [36]
HRV	Heart rate variability	[7] [12] [35] [36] [37] [38]
SDNN	Interquartile range of normal-to-normal intervals	[4] [5] [35] [39]
RMSSD	Root of the median of the sum of squares of differences in normal-to-normal intervals	[4] [35] [40]
pNN50	Proportion of successive normal-to-normal intervals that differ more than 50 ms	[35] [39] [40]
LF	Low frequency band (0.04-0.15 Hz)	[38]
HF	High frequency band (0.15-0.4 Hz)	[38]
LF/HF	Ratio of low and high frequency band	[4] [5] [38]
mSCL	Median skin conductance level	[41] [42]
minSCL	Minimum skin conductance level	[40]
SD SCL	Interquartile range of skin conductance level	[40]
Absdiff2	Signal power of skin conductance in second difference	[11] [42]
mSCR	Median skin conductance	[7]
SCRR	Skin conductance response rate	[5] [40] [43]
SCR dur	Total duration of skin conductance response	[7]
SCR amp	Total amplitude of skin conductance response	[7] [43]
SCR area	Skin conductance response area	[4]
SCR pow	Signal power of skin conductance response	[4]
mST	Median skin temperature	[4] [44] [45]
SD ST	Interquartile range of skin temperature	[4]
ST slope	Slope of skin temperature	[4]

II.4.2 Statistical analysis

The data from the laboratory measurement was split into rest moments and baseline speaking task. The rest moments were the rest moments during the laboratory protocol. The Mann-Whitney U test was performed to test for significant differences between rest moments and the baseline speaking task. Bonferroni correction was used to correct for multiple comparisons. Additionally, the signals were split into S1, S2, S3, S4 and S5, corresponding to the self-reported stress level of the Likert scale of self-reports during laboratory measurements. S1 means being not stressed and S5 means being very stressed. Differences in features between the five levels were investigated with linear models, with the feature as response variable and fixed effects was the stress-level. The models were individually created, since the stress-response can be person-specific. Differences in response time of SCWT were investigated with the Kruskal-Wallis H test followed by Tukey Kamer post-hoc tests. The same was done for the three phases of SCWT.

Sixty minutes before the questionnaire was prompted was used for stress level of one hour in daily-life [9] and two minutes prior to questionnaire was used for the current stress level [5]. The Mann-Whitney U test was used to test whether windows with food intake differed from windows without food intake. The same was performed for the intake of alcohol and caffeine. To investigate the differences in stress-levels in daily-life similar linear models were created as in laboratory, but now the time between the measurements was included, since the timepoints were further apart. By laboratory measurement these were not included as the timepoints were all close to each other. Linear models were corrected for multiple comparisons with Bonferroni correction.

III. Results

III.1 Population characteristics

Six participants were included of which two participants were excluded and one participant was not used in the data analysis. One participant was excluded due to diagnosis of mental disorder, one participant was excluded due to sensor failure. In one subject data was excluded, because no data was available from the EMA questionnaire and sensor data simultaneously without moderate to high physical activity. From the remaining 3 participants, one was male. The median age of the eligible participants was 28 years with interquartile range of 4.0 years. The BMI of the participants was 23.41 kg m^{-3} +/- 2.53. Two participants were college students and one participant had a fulltime employment. The average PSS score was 24 (IQR = 5.25). DSI showed that participants experienced different stressors and that these stressors can differ per day. All three participants experienced stress by mainly work related factors, such as thinking about unfinished work, hurrying to meet deadlines, being unable to complete task, being unorganized and being unable to complete all tasks for the day. However, also social factors such as not hearing from someone you expected to hear from and health factors such as not having enough sleep caused stress.

III.2 Adherence

During the laboratory measurement, 100% of the self-reports were filled in. During the daily-life measurement, on average 78.6% of the EMAs were filled in, which resulted in a total of 198 completed EMAs. The DSI and PSS questionnaires were completed 100% of the time. Furthermore, all subjects completed both SCWT in daily-life.

III.3 Data quality

During the laboratory measurement, 13 out of 30 windows (43.33%) and 94 out of 210 windows (44.76%) of the baseline speaking task and rest moments, respectively, contained good quality data. The stress levels (S1, S2, S3-S4) during the laboratory measurement had 59 out of 140 (42.14%), 48 out of 110 (43.64%) and 13 out of 50 (26%) respectively good

quality windows. The good quality windows of SCWT in daily-life ranged from 25% to 50%, most good quality windows during congruent phase (50%) and lowest good quality windows during neutral phase (25%). The current stress level in daily-life had 81 out of 352 (23.01 %), 36 out of 184 (19.57 %) and 19 out of 84 (22.62 %) good quality windows respectively. The past one hour stress level in daily-life had 1867 out of 5520 (33.82%), 1118 out of 3840 (29.11%) and 253 out of 720 (35.14%) good quality windows

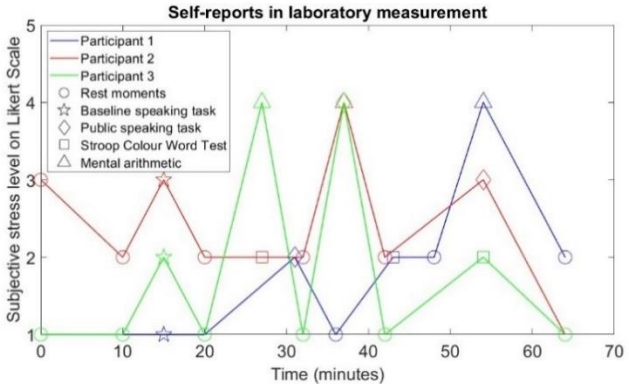


Figure 3: Self-reports in laboratory measurement for the three participants

respectively. The PPG data quality was the lowest of the three physiological signals (37.78%), SC had 99.09% and ST 100%.

III.4 Correlation analysis

High correlations (r>=0.8) were investigated amongst features in laboratory and daily-life measurement per subject. If two features were at least correlated in one of the two settings, one of the two features was excluded from further analysis. mHR was chosen over HRV and RMSSD because literature showed higher reliability for mHR with E4. HF was chosen over LF, because the window length might not be long enough to capture LF [34]. mSCL was chosen over SD SCL. SCR amp and SCR pow were chosen from correlated SCR features, because these two features are associated with stress in daily-life. An overview of the features for each participant is presented in Table 2.

Table 2: Overview of features for each participant

Feature	Participant 1	Participant 2	Participant 3
mHR	*	*	*
SDNN	*	*	*
pNN50	*		
HF	*	*	*
LF/HF	*	*	*
mSCL	*	*	*
minSCL	*		
Absdiff2			*
SCRR			*
SCR dur			*
SCR amp	*	*	*
SCR pow	*	*	*
mST	*	*	*
SD ST	*	*	*
ST slope	*		

III.5 Laboratory measurement

During the laboratory measurement, self-reports and physiological signals were obtained. Figure 3 shows that the participants responded differently to the stressors, for example the increase in stress-level is not similar and the decrease in stress-level after the stressor differed. All three participants had different trends. Participant s1 and 2 had an increase and decrease in self-reported stress, respectively. While participant 3 started and ended with no self-reported stress.

The percentage of correct answers in neutral, congruent and incongruent phase of SCWT were determined per subject. Participant 1 had 100%, 100% and 100% respectively. Participant 2 had 98.36%, 98.47% and 99.01% . Participant 3 had 99.08%, 99.19% and 97.65%. All participants showed a significant difference in response time between congruent and incongruent phase and participant 2 and 3 also had significant difference between neutral and incongruent phase (Figure 5A). Only participant 2 showed a significant difference between neutral and congruent phase.

For each participant it was tested if features were significantly different during the baseline speaking task compared to rest moments. For participants 1 and 2 there were no features that were significantly different during baseline speaking task and rest moments. Finally, two features (mHR and mSCL) showed a significant difference between baseline speaking task and rest moments for participant 3.

Linear models for each participant between each feature and the stress-levels were created. Stress levels S3 and S4 were merged together (S3-S4), since the self-reports were imbalanced [4]. Participant 1 had a significant difference ($p<0.001$) in HF between S1 and S2 and between S1 and S3-S4 (Figure 4). The absolute difference between S1 and S2 was $-1.03\cdot 10^{5}$ with 95% confidence intervals (CI) of $-1.03\cdot 10^{4}$ to $1.02\cdot 10^{4}$. The absolute difference with 95% CI between S1 and S3-S4 was $-1.02\cdot 10^{5}$ ($-1.03\cdot 10^{4}$,$1.02\cdot 10^{4}$). Additionally, the participant had a significant difference ($p<0.001$) in SCR pow between S1 and S2 and between S1 and S3-S4. The absolute difference with 95% CI between S1 and S2 was $-1.78\cdot 10^{-2}$ ($-1.94\cdot 10^{-2}$, $-1.63\cdot 10^{-2}$). Next to that, the absolute difference with 95% CI between S1 and S3 was -0.017237 ($-1.88\cdot 10^{-2}$, $-1.57\cdot 10^{-2}$).

III.6 Stress test in daily-life

The stress test in daily-life in this study was the SCWT, which was 100% completed. The percentage of correct answers in neutral, congruent and incongruent phase of SCWT were determined per subject per SCWT. By the first SCWT in daily-life participant 1 had 100%, 100% and 97.40%, participant 2 had 99.12%, 96.64% and 99.03% and participant 3 had 99.08%, 99.17% and 98.92% respectively. By the

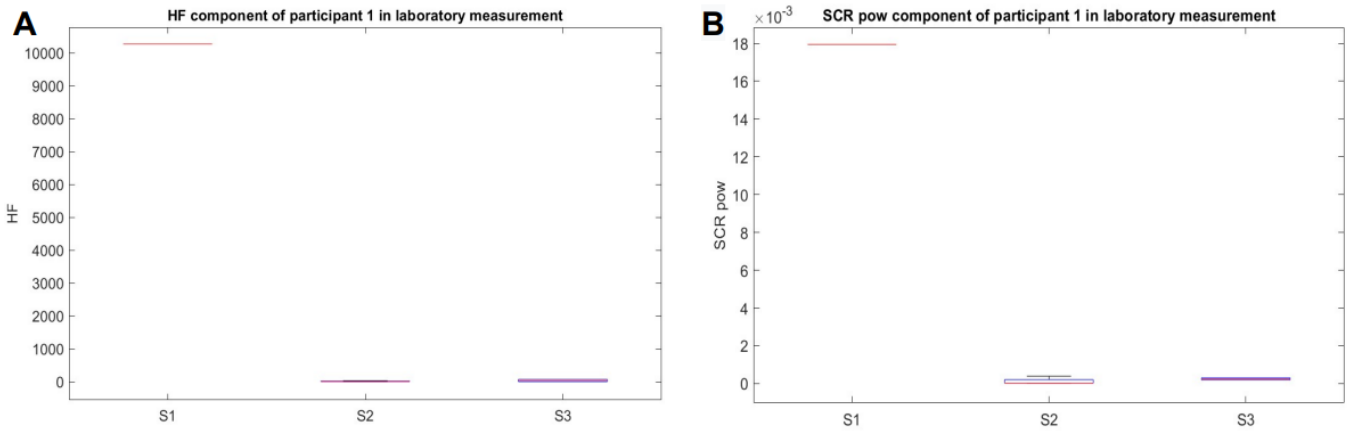


Figure 4: Results of the HF component of participant 1 in laboratory measurement (A) and SCR pow (B) between the three stress levels

second SCWT in daily-life participant 1 had 98.90%, 100% and 100%, participant 2 had 100 %, 100 % and 99.05% and participant 3 had 98.29%, 98.40% and 97.39% respectively.

For each participant it was investigated if there were significant differences in response time between the three phases of SCWT (Figure 5B). All participants showed a significant difference in response time between congruent and incongruent phase by week 1, participant 1 also showed this in week 2. Participant 1 in week 2 and participant 3 in week 1 also showed a significant difference between neutral and incongruent phase. Finally, participant 1 showed in week 1 a significant difference between neutral and congruent phase.

There were no significant differences between the first and second SCWT in daily-life in the three phases. Next, all the data was taken together and there were no significant differences between the three phases of SCWT.

III.7 Daily-life measurement

35.58% of the filled in EMAs contained moderate to high physical activity, 37.45% contained food intake and 10.11% had caffeine or alcohol intake. In daily-life participants felt 57.30%, 31.84% , 8.61% and 1.50%, not stressed, a little stressed, moderately, very and extremely stressed during the past hour. For the current stress level this was, 59.93%, 29.96% 8.24%, 1.12% of the time 0% of the time. Self-reports are shown in Figure 6. For participant 1 83.8% of the self-reports are the same for current and past one hour stress level, for participant 2 and 3 this was 88.3% and 78.6% respectively.

The Mann-Whitney U test showed that there were no significant differences in physiological features between windows with food intake compared to windows without food intake for the current stress level. However, when looking at the stress level of the past hour, there was a significant difference in features between windows with food intake and windows without food intake. For the current stress level, windows with food intake were also used for the analysis. Whereas for the stress level of the past hour, only windows without food intake were used.
A similar analysis could not be done for windows with caffeine or alcohol intake, since there were no windows with caffeine or alcohol intake and without intake that had the same stress level and were not affected by high to moderate physical

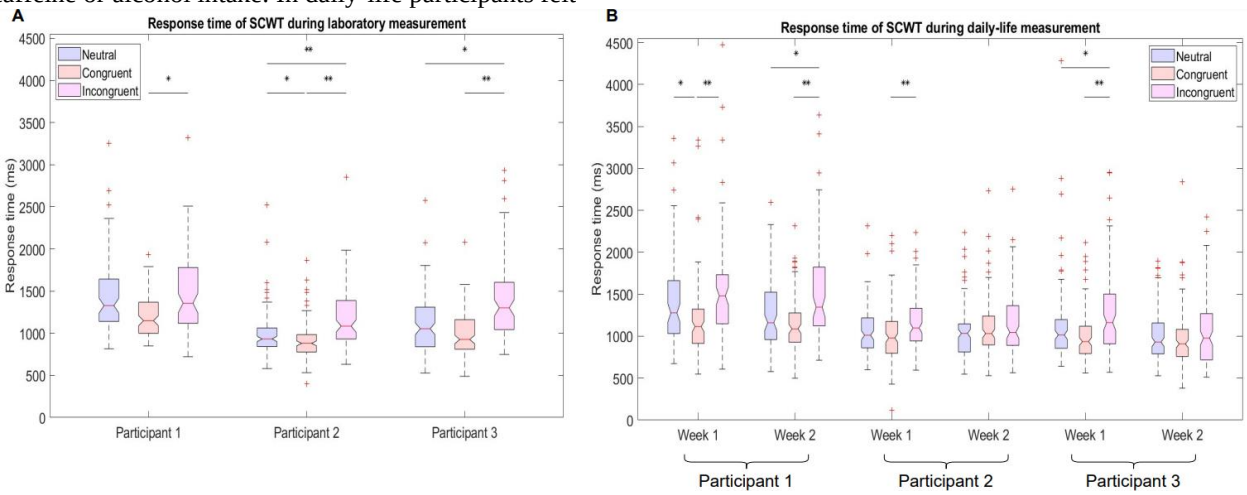


*Figure 5: Response time of SCWT during (A) laboratory measurement and (B) daily-life measurement, the colours blue, red and pink represent the three phases neutral, congruent and incongruent respectively, * indicates a significant difference of p<0.05 and ** indicates a significant difference of p<0.001*

activity. So, all windows with alcohol and caffeine intake were excluded from further analysis.

For both the current stress level and the stress level of the past hour, linear models were created for each participant and each feature. S3-S4 were merged together due to imbalanced self-reports [4]. No significant differences were found for all participants for both the current and the past one hour stress level. The corresponding absolute differences between the stress levels in feature values and the confidence intervals can be seen in Tables 4-6 in Appendix A.5. In general, it was seen that the variation in confidence intervals is larger than the absolute difference in a stress-related feature between two stress levels.

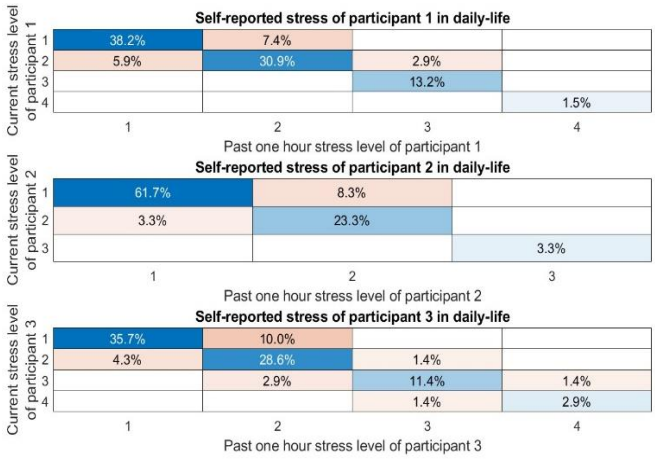


Figure 6: The self-reported stress of the three participants in daily-life of the past one hour stress level on the x-axis and the current stress level on the y-axis. 1 means being not stressed and 4 means being very stressed.

IV. Discussion

This study created a framework for capturing stress with a multi-sensor approach and compared stress-related features between laboratory and daily-life settings in three healthy individuals. The results point into the direction that stress-related features are not similarly associated to stress in laboratory and daily-life setting. Next to that, stress-related features were participant specific. Furthermore, no associations were found between stress-related features and SCWT. Future research should focus on including more subjects, to confirm these results.

The framework captures three physiological signals, which are often used in stress detection in daily-life. The framework builds on previous research by implementing not only filtering steps and quality indicators for PPG, SC and ST [4,29,30,32], but also artifact detection in SC due to thermoregulation for example, which is assumed to be a large problem in ambulatory data [33]. Since good quality data is one of the main challenges in stress detection, especially in daily-life, previous research uses different ways to ensure good quality data by using methods such as smoothing, quality indicators, artifact detection based on machine learning. Although the different methods used are successful, the difference in methods also makes it more difficult to compare results. Especially because the results about the relation between stress and physiological signals have not been consistent [7]. Therefore, it is important that one general framework is created and this study provides a first start. Further research can built on this framework by including movement artefact detection for PPG, determining a cut-off for threshold for the slope of the temperature and further analysing which weights for skin conductance score components are most optimal for example.

Comparing the results of the laboratory to the daily-life measurements showed that no stress-related feature was significantly associated with stress in both situations, for the current and past one hour stress level. This was not as expected, because in literature it was suggested that a participant would respond to stress in a similar manner in different settings [9]. An explanation for this finding is that there were not enough good quality windows to find effects in daily-life, as large part of the data had to be excluded due to low quality. Another reason for the finding is that the two minute interval, for the current stress-level, did not reflect the stress response, since no significant differences were found for the current stress level. For the past one hour also no significant differences were found which can be due to having large changes in physiology within one hour [9], masking the stress response. This makes it more difficult to detect, which might also explain the large variation in confidence intervals seen in the daily-life measurement. Another reason could be that the stress in daily-life are smaller stressors or trigger a different response, which may be more difficult to detect. Furthermore, stress in daily-life may have been short in duration [49], in which it is hard to find it in a large time window or to capture it with EMA due to recall bias. Lastly, the stress in daily-life can be very different to the stress captured in laboratory, since only mental and psychological stressors were applied in laboratory. Whereas in daily-life many different stressors were present such as work-related stress or damage of your property.

The results in literature about features associated to stress in daily-life are different. Smets et al. [4] used in an ambulatory study also the stress level of the past hour and found four features (mHR, SCR area, SCR pow and mST) to be significant different between the stress levels. These features were not found to be responsive to stress in this study, which might be explained by person-specificity of the stress response. Wijsman et al. [5] created regression models based on ambulatory data for the current stress level and for the average stress level of the past thirty minutes. Next to that, in [5] it was also seen that some subjects had the same features in both models, indicating that they had a similar stress response. However, others had different features included in the two models. This emphasizes the person-specificity of the stress response. In this study, no stress-related feature was responsive to stress for the current stress level and the stress level of the past hour, so at this point no pattern can be found in the physiological signals between the two stress references. An explanation for this might be that the references reflect different type of stress, very short-lived stress versus intermediate stress for example.

Next to that, it is important to note that the stress measurement

in daily-life could be influenced by speaking. Two participants felt little stress after the baseline speaking task in the laboratory measurement, which was included to investigate the influence of speaking on features. Additionally, two features showed significant differences between baseline speaking task and non-stressed moments (mHR and mSCL) in participant 3. The time domain heart rate feature showed significant differences, because speaking influences respiration, which in turn influences HRV [47]. In a laboratory measurement by Smets et al., it was also seen that misclassifications of baseline speaking task occurred [11]. However, it is important to note that participant 3 also experienced stress during the baseline speaking task. Therefore, the difference in features could occur due to stress or due to speaking or both. In laboratory and daily-life analysis based on stress levels these features in participant 3 did not show significant differences.

The self-reports of the laboratory measurement showed the person-specificity of the stress response, similar to results in literature [4], leading to analysis of stress levels based on the self-reports. The results showed that different features responded to stress in the participants, by participant 1 only had significant differences in HF and SCR pow. HF decreased between S1 and S2 and between S1 and S3-S4, which was similar to literature [48]. The feature SCR pow also showed a decrease, which contradicts literature [4]. A reason for this could be that the sample size. The comparison between different self-reported stress levels in laboratory has not been done before.

As this is the first time the SCWT is used with these settings, SCWT is analysed by response time and scores to make a first step towards validation. There was a significant increase in response time of the incongruent phase compared to the congruent phase in laboratory measurement for participant 1 week 1 and in week 2 in daily-life, also compared to the neutral phase. Also participants 2 and 3 showed a significant difference in response time between congruent and incongruent phase in week 1 in daily-life. Participant 3 also showed a significant difference in response time between neutral and incongruent phase in week 1 in daily-life. In addition, there was a decrease of the number of correct answers given during the incongruent phase compared to the congruent and neutral phases in laboratory by participant 3 and in daily-life measurement for most participants. This was as expected, since the incongruent phase should induce stress, leading to longer response time and lower SCWT scores [9]. However, there were also unexpected findings. In laboratory the SCWT score was higher in the congruent phase compared to neutral phase for two participants. A reason for this could be that a learning effect occurred, since the congruent phase is quite similar to the neutral phase and it is repeated after each other. All participants showed in week 2 in daily-life a lower response time for the incongruent phase compared to the incongruent phase in week 1. This was not as expected and can be due to learning effect. It is difficult to compare these results to literature, as scores and response times are not often reported and different settings are used [47].

No significant differences were found in stress-related features between the three phases of SCWT in daily-life. A reason for this could be the small sample size or due to bad signal quality for example. Another reason could be that the SCWT does not induce stress, since the self-reports in laboratory setting showed that the smallest stressor change in self-reports was by SCWT. This may indicate that the SCWT as used in this study might not be useful to induce stress. In this study, no self-reports were prompted by SCWT in daily-life, which makes the comparison to laboratory and daily-life results and interpreting the results more difficult. Further research should also incorporate self-reports when using a stress test in daily-life to assess if stress was successfully induced and further research is necessary to validate the SCWT with these settings.

The strengths of the study were that it captures stress in a fully controlled and uncontrolled setting with a stress test in daily-life as bridge in-between which is unique. Furthermore, this study provides a framework of pre-processing steps for a multi-sensor approach to reduce noise and detect artifacts in these physiological signals. Also, in laboratory setting a comparison was made between the stress-related features based on self-reports, which has not been done before. Next to that, the analysis was done on individual level with ideographic approach, which is not often done in literature. More studies on individual level is needed, since group level information is not straightforward to transfer to individual levels and the stress response is person-specific.

Despite the strengths, this study also had limitations which have to be taken into account when interpreting the results. First of all, the sample size of this study was very small. Based on only three participants no strong conclusions can be drawn from this research.
Second, there are some limitations in the pre-processing steps. The window length was 30 seconds, which may have been too short to determine features such as HF [35]. Also, the skin temperature was used as reference for ambient temperature, but the difference between skin temperature and ambient temperature may vary. For example, when the blood pressure is increased, the skin temperature may show a more rapid decline compared to ambient temperature [49]. Next to that, the PPG template was set daily, future research should look into updating the PPG template throughout the measurement to make it more compatible to good PPG segments [31]. Furthermore, excluding the windows with high physical activity, food intake and caffeine and alcohol intake may have led to bias in the results. Next to that, the threshold on the slope of the temperature was relatively high compared to other threshold used in literature [33]. This could have led to thermoregulation influences on the skin conductance signal. However, a lower threshold was not possible due to the accuracy of the E4 sensor. Finally, the way the 30-second windows were created may have influenced the results. The windows were created from start of the measurement, but it using another start point for creating the windows might lead to more good quality windows.
Third, the sensor reliability is questioned in literature. It is discussed that it seems E4 underestimates the SC features, which may be due to the wrist being less responsive to SC. This means that with E4 only larger stressors can be detected [50]. In this study, a large part of the windows had to be excluded

due to artifacts, especially in daily-life, mainly due to PPG signal. The data quality in daily-life was around 10% lower for the past one hour stress level and 20% lower for the current stress level compared to laboratory. By the current stress level and laboratory, the signal quality was reduced for higher stress levels, whereas by the past one hour stress level the signal quality was the highest for S3-S4. However, by the past one hour stress level the signal quality scores were all rather close to each other. These difference in samples available for each stress level might have influenced the results.

Fourth, the stress references in daily-life measurement are based on self-reports, which may lead to bias compared to the laboratory measurement. This may raise doubts about if stress is actually captured or rather an increase in ANS activity. This should be further examined by looking at arousal levels and control of the SAM in combination with self-reports and physiological signals [4]. Also, it should be noted that the formulation of the 5-point Likert scale may have influenced the self-reports [7]. Further research should try to locate the stress response in the hour. Moreover, the obtained stress-levels were unbalanced, which may have influenced the results of the linear models.

Fifth, the stress level did not always return to baseline after rest in the laboratory measurement, which means that the rest periods of five minutes may not have been enough. This was not reported by Plarre et al. [21], but Smets et al.[9] did report similar findings with a rest period of two minutes.

Finally, the models that are compared between laboratory and daily-life are not exactly the same. Since the laboratory measurement is only one hour, it was assumed the timepoints will not differ as much as in daily-life within a stress level. Further research should test if this assumption is valid. Also, the autocorrelation between feature values was not taken into account as most features did not show high autocorrelations. Further research should take autocorrelation within features values into account. Furthermore, the linear model could be overfitted, as the sample size in this population is small. Finally, since this study focused on single subject, the results are more difficult to generalize to a broader population.

Further research should look to validate the results that were found in this study with larger sample sizes and with other wearable sensors. Next to that, methods should be created for handling confounding variables in the data such as moderate to high physical activity, to reduce bias. Also, other stress tests in daily-life situations should be investigated, for example the Montreal Imaging Stress Task [9], to investigate if there are similarities between the stress response from the stress test and daily-life stressors. Next to that, further research could investigate if the similar stress responses occur for the same stressor and explore different timeframes for the stress level in daily-life. Moreover, as the results point into the direction that the stress-related features do not have similarities between laboratory and daily-life stressors, this should be further investigated with single subject studies.

V. Conclusion

This study provides a framework for a multi-sensor approach for stress detection in laboratory and daily-life settings. Next to that, the results point into the direction that the stress-related features that are responsive to stress are not similar between laboratory and daily-life. These findings suggest that information about a person's stress in laboratory and daily-life is complementing each other. Also, the results showed different stress-related features associated to stress for each participant, highlighting that the stress response is person-specific and the need for personalized analysis. This study suggests the need for personalized techniques to model the stress response and to see the laboratory and daily-life settings as providing different information. Further research is necessary to improve the stress detection through physiological signals to create personalized just-in-time interventions to improve health.

References

[1] Tong HL, Quiroz JC, Kocaballi AB, Fat SCM, Dao KP, Gehringer H, Chow CK, Laranjo L. Personalized mobile technologies for lifestyle behavior change: A systematic review, meta-analysis, and meta-regression. Prev Med (Baltim) 2021;148. doi:10.1016/J.YPMED.2021.106532.

[2] Chigom E. Non-communicable diseases. 1988. doi:10.5005/jp/books/11410_18.

[3] Liu YZ, Wang YX, Jiang CL. Inflammation: The common pathway of stress-related diseases. Front Hum Neurosci 2017;11:316. doi:10.3389/FNHUM.2017.00316/BIBTEX.

[4] Smets E, Rios Velazquez E, Schiavone G, Chakroun I, D'Hondt E, De Raedt W, Cornelis J, Janssens O, Van Hoecke S, Claes S, Van Diest I, Van Hoof C. Large-scale wearable data reveal digital phenotypes for daily-life stress detection. Npj Digit Med 2018 11 2018;1:1–10. doi:10.1038/s41746-018-0074-9.

[5] Wijsman J. Sensing stress: stress detection from physiological variables in controlled and uncontrolled conditions 2014. doi:10.3990/1.9789036537858.

[6] Sun F-T, Kuo C, Cheng H-T, Buthpitiya S, Collins P, Griss M. LNICST 76 - Activity-Aware Mental Stress Detection Using Physiological Sensors. LNICST 2012;76:282–301

[7] Smets E, De Raedt W, Van Hoof C. Into the Wild: The Challenges of Physiological Stress Detection in Laboratory and Ambulatory Settings. IEEE J Biomed Heal Informatics 2019;23:463–73. doi:10.1109/JBHI.2018.2883751..

[8] Everly GS, Lating JM. The Anatomy and Physiology of the Human Stress Response. A Clin Guid to Treat Hum Stress Response 2013:17–51. doi:10.1007/978-1-4614-5538-7_2.

[9] Smets E. Towards large-scale physiological stress detection in an ambulant environment 2018.

[10] Karthikeyan P, Murugappan M, Yaacob S. A review on stress inducement stimuli for assessing human stress using physiological signals. Proc - 2011 IEEE 7th Int Colloq Signal Process Its Appl CSPA 2011 2011:420–5. doi:10.1109/CSPA.2011.5759914.

[11] Smets E, Casale P, Großekathöfer U, Lamichhane B, De Raedt W, Bogaerts K, Van Diest I, Van Hoof C. Comparison of Machine Learning Techniques for Psychophysiological Stress Detection. Commun Comput Inf Sci 2015;604:13–22. doi:10.1007/978-3-319-32270-4_2.

[12] Hovsepian K, Al'absi M, Ertin E, Kamarck T, Nakajima M, Kumar S. CStress: Towards a gold standard for continuous stress assessment in

the mobile environment. UbiComp 2015 - Proc 2015 ACM Int Jt Conf Pervasive Ubiquitous Comput 2015:493–504. doi:10.1145/2750858.2807526.

[13] Healey JA, Picard RW. Detecting stress during real-world driving tasks using physiological sensors. IEEE Trans Intell Transp Syst 2005;6:156–66. doi:10.1109/TITS.2005.848368. [14] How Laboratory Experiments Can Be Exploited for Monitoring Stress in the Wild: A Bridge Between Laboratory and Daily Life

[14] Can YS, Gokay D, Kılıç DR, Ekiz D, Chalabianloo N, Ersoy C. How Laboratory Experiments Can Be Exploited for Monitoring Stress in the Wild: A Bridge Between Laboratory and Daily Life. Sensors 2020, Vol 20, Page 838 2020;20:838. doi:10.3390/S20030838.

[15] Can YS, Chalabianloo N, Ekiz D, Fernandez-Alvarez J, Riva G, Ersoy C. Personal Stress-Level Clustering and Decision-Level Smoothing to Enhance the Performance of Ambulatory Stress Detection with Smartwatches. IEEE Access 2020;8:38146–63. doi:10.1109/ACCESS.2020.2975351.

[16] Zhang T, Gong N, Jia R, Li H, Ni X. Stroop effect in smartphone addiction among college students. Medicine (Baltimore) 2021;100:e26741. doi:10.1097/MD.0000000000026741.

[17] Bajaj JS, Thacker LR, Heuman DM, Fuchs M, Sterling RK, Sanyal AJ, Puri P, Siddiqui MS, Stravitz RT, Bouneva I, Luketic V, Noble N, White MB, Monteith P, Unser A, Wade JB. The Stroop Smartphone App is A Short and Valid Method to Screen for Minimal Hepatic Encephalopathy. Hepatology 2013;58:1122. doi:10.1002/HEP.26309.

[18] Voss A, Schroeder R, Heitmann A, Peters A, Perz S. Short-Term Heart Rate Variability—Influence of Gender and Age in Healthy Subjects. PLoS One 2015;10:e0118308. doi:10.1371/JOURNAL.PONE.0118308.

[19] Iqbal T, Redon-Lurbe P, Simpkin AJ, Elahi A, Ganly S, Wijns W, Shahzad A. A Sensitivity Analysis of Biophysiological Responses of Stress for Wearable Sensors in Connected Health. IEEE Access 2021;9:93567–79. doi:10.1109/ACCESS.2021.3082423

[20] Albertetti F, Simalastar A, Rizzotti-Kaddouri A. Stress Detection with Deep Learning Approaches Using Physiological Signals. Lect Notes Inst Comput Sci Soc Telecommun Eng LNICST 2021;360:95–111. doi:10.1007/978-3-030-69963-5_7.

[21] Plarre K, Raij A, Hossain SM, Ali AA, Nakajima M, Al'Absi M, Ertin E, Kamarck T, Kumar S, Scott M, Siewiorek D, Smailagic A, Wittmers LE. Continuous inference of psychological stress from sensory measurements collected in the natural environment. Proc 10th ACM/IEEE Int Conf Inf Process Sens Networks, IPSN'11 2011:97–108.

[22] Kocielnik R, Sidorova N, Maggi FM, Ouwerkerk M, Westerink JHDM. Smart technologies for long-term stress monitoring at work. Proc CBMS 2013 - 26th IEEE Int Symp Comput Med Syst 2013:53–8. doi:10.1109/CBMS.2013.6627764.

[23] Jongstra S, Wijsman LW, Cachucho R, Hoevenaar-Blom MP, Mooijaart SP, Richard E. Cognitive Testing in People at Increased Risk of Dementia Using a Smartphone App: The iVitality Proof-of-Principle Study. JMIR MHealth UHealth 2017;5. doi:10.2196/MHEALTH.6939.

[24] Jeffers AJ, Mason TB, Benotsch EG. Psychological eating factors, affect, and ecological momentary assessed diet quality. Eat Weight Disord 2020;25:1151–9. doi:10.1007/S40519-019-00743-3/TABLES/4.

[25] Engel SG, Wonderlich SA, Crosby RD, Mitchell JE, Crow S, Peterson CB, Grange D Le, Simonich HK, Cao L, Lavender JM, Gordon KH. The Role of Affect in the Maintenance of Anorexia Nervosa: Evidence from a Naturalistic Assessment of Momentary Behaviors and Emotion. J Abnorm Psychol 2013;122:709. doi:10.1037/A0034010.

[26] Reichenberger J, Pannicke B, Arend AK, Petrowski K, Blechert J. Does stress eat away at you or make you eat? EMA measures of stress predict day to day food craving and perceived food intake as a function of trait stress-eating. Psychol Health 2021;36:129–47. doi:10.1080/08870446.2020.1781122.

[27] Brantley PJ, Waggoner CD, Jones GN, Rappaport NB. A daily stress inventory: Development, reliability, and validity. J Behav Med 1987;10:61–73. doi:10.1007/BF00845128.

[28] Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983;24:385–96. doi:10.2307/2136404.

[29] Chandra V, Priyarup A, Sethia D. Comparative Study of Physiological Signals from Empatica E4 Wristband for Stress Classification. Commun Comput Inf Sci 2021;1441:218–29. doi:10.1007/978-3-030-88244-0_21/TABLES/2.

[30] Orphanidou, C. et al. Signal quality indices for the electrocardiogram and photoplethysmogram: derivation and applications to wireless monitoring. IEEE J. Biomed. Health Inform.19, 832–838 (2014)

[31] Lim PK, Ng SC, Lovell NH, Yu YP, Tan MP, McCombie D, Lim E, Redmond SJ. Adaptive template matching of photoplethysmogram pulses to detect motion artefact. Physiol Meas 2018;39:105005. doi:10.1088/1361-6579/AADF1E.

[32] Sevil M, Rashid M, Askari MR, Maloney Z, Hajizadeh I, Cinar A. Detection and Characterization of Physical Activity and Psychological Stress from Wristband Data. Signals 2020;1:188–208. doi:10.3390/signals1020011.

[33] Gashi S, DI Lascio E, Stancu B, Swain V Das, Mishra V, Gjoreski M, Santini S. Detection of Artifacts in Ambulatory Electrodermal Activity Data. Proc ACM Interactive, Mobile, Wearable Ubiquitous Technol 2020;4. doi:10.1145/3397316.

[34] Benedek, M., Kaernbach, C.: A continuous measure of phasic electrodermal activity. J. Neurosci. Methods 190(1), 80–91 (2010)

[35] L. Salahuddin, J. Cho, M. G. Jeong, and D. Kim. Ultra short term analysis of heart rate variability for monitoring mental stress in mobile settings. In Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE, pages 4656–4659, Aug. 2007.

[36] Hickey BA, Chalmers T, Newton P, Lin CT, Sibbritt D, McLachlan CS, Clifton-Bligh R, Morley J, Lal S. Smart Devices and Wearable Technologies to Detect and Monitor Mental Health Conditions and Stress: A Systematic Review. Sensors (Basel) 2021;21. doi:10.3390/S21103461.

[37] Taelman J, Vandeput S, Spaepen A, Van Huffel S. Influence of mental stress on heart rate and heart rate variability. IFMBE Proc 2008;22:1366–9. doi:10.1007/978-3-540-89208-3_324.

[38] Lucini D, Norbiato G, Clerici M, Pagani M. Hemodynamic and Autonomic Adjustments to Real Life Stress Conditions in Humans. Hypertension 2002;39:184–8. doi:10.1161/HY0102.100784.

[39] Tharion E, Neelakantan N, Parthasarathy S. Short-term heart rate variability measures in students during examinations. Natl Med J India 2009;22.

[40] de Vries GJJ, Pauws SC, Biehl M. Insightful stress detection from physiology modalities using Learning Vector Quantization. Neurocomputing 2015;151:873–82. doi:10.1016/J.NEUCOM.2014.10.008.

[41] Jacobs SC, Friedman R, Parker JD, Tofler GH, Jimenez AH, Muller JE, Benson H, Stone PH. Use of skin conductance changes during mental stress

testing as an index of autonomic arousal in cardiovascular research. Am Heart J 1994;128:1170–7. doi:10.1016/0002-8703(94)90748-X.

[42] Wijsman J, Grundlehner B, Liu H, Penders J, Hermens H. Wearable Physiological Sensors Reflect Mental Stress State in Office-Like Situations. Undefined 2013:600–5. doi:10.1109/ACII.2013.105

[43] Setz C, Arnrich B, Schumm J, Marca R La, Tr G, Ehlert U. Discriminating stress from cognitive load using a Wearable EDA Device. Technology 2010;14:410–7.

[44] Karthikeyan P, Murugappan M, Yaacob S. Multiple Physiological Signal-Based Human Stress Identification Using Non-Linear Classifiers. Elektron Ir Elektrotechnika 2013;19:80–5. doi:10.5755/J01.EEE.19.7.2232.

[45] Ahmadi N, Sasangohar F, Nisar T, Danesh V, Larsen E, Sultana I, Bosetti R. Quantifying Occupational Stress in Intensive Care Unit Nurses: An Applied Naturalistic Study of Correlations Among Stress, Heart Rate, Electrodermal Activity, and Skin Temperature. Undefined 2021. doi:10.1177/00187208211040889.

[46] Chen H, Cohen P, Chen S. How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. Commun Stat Simul Comput 2010;39:860–4. doi:10.1080/03610911003650383.

[47] Vazan R, Filcikova D, Mravec B. Effect of the Stroop test performed in supine position on the heart rate variability in both genders. Auton Neurosci Basic Clin 2017;208:156–60. doi:10.1016/J.AUTNEU.2017.10.009.

[48] Lucini D, Norbiato G, Clerici M, Pagani M. Hemodynamic and Autonomic Adjustments to Real Life Stress Conditions in Humans. Hypertension 2002;39:184–8. doi:10.1161/HY0102.100784.

[49] Maurizio Garbarino, Matteo Lai, Dan Bender, Rosalind W Picard, and Simone Tognetti. 2014. Empatica E3–A Wearable Wireless Multi-sensor Device for Real-time Computerized Biofeedback and Data Acquisition. In Proceedings of the International Conference on Wireless Mobile Communication and Healthcare (MobiHealth 2014).

[50] van Lier HG, Pieterse ME, Garde A, Postel MG, de Haan HA, Vollenbroek-Hutten MMR, Schraagen JM, Noordzij ML. A standardized validity assessment protocol for physiological signals from wearable technology: Methodological underpinnings and an application to the E4 biosensor. Behav Res Methods 2020;52:607–29. doi:10.3758/S13428-019-01263-9/TABLES/7.

[51] On the Filtering of Photoplethysmography Signals. Evaluation of Mental Stress and Heart Rate Variability Derived from Wrist-based Photoplethysmography Chongyan Chen1, a, Chunhung Li2, b, Chih-Wei Tsai2, c and Xinghua Deng

[52] Li S, Liu L, Wu J, Tang B, Li D. Comparison and Noise Suppression of the Transmitted and Reflected Photoplethysmography Signals. Biomed Res Int 2018;2018. doi:10.1155/2018/4523593.

[53] Ellis G. Filters in Control Systems. Control Syst. Des. Guid., Butterworth-Heinemann; 2012, p. 165–83. doi:10.1016/B978-0-12-385920-4.00009-6.

[54] Wanhammar L. Digital Filters. DSP Integr. Circuits, Academic Press; 1999, p. 115–86. doi:10.1016/B978-012734530-7/50004-0

[55] EMPSTER J. Recording and Analysis of Intracellular Electrophysiological Signals. Lab. Comput., Academic Press; 2001, p. 172–225. doi:10.1016/B978-012209551-1/50040-4.

[56] Pachi A, Ji T. Frequency and velocity of people walking. Struct Eng 2005;83:36–40.

[57] Azami H, Mohammadi K, Bozorgtabar B. An Improved Signal Segmentation Using Moving Average and Savitzky-Golay Filter. J Signal Inf Process 2012;03:39–44. doi:10.4236/JSIP.2012.31006.

[58] Schafer RW. ON THE FREQUENCY-DOMAIN PROPERTIES OF SAVITZKY-GOLAY FILTERS Ronald W . Schafer Hewlett-Packard Laboratories 1501 Page Mill Road Palo Alto , CA 94304. Digit Signal Process Work IEEE Signal Process Educ Work 2011:54–9.

Appendices

A.1 Ecological Momentary Assessment (EMA)

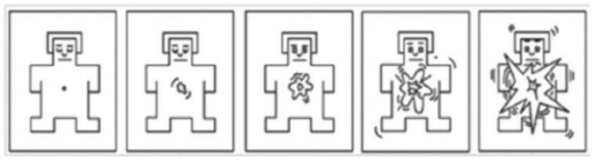
The EMA consists of nine questions which are about emotions linked to stress, the stress level, food and drinks (caffeine and alcohol) and physical activity, see below:

1. Below you can see different emoticons from unhappy to happy. Please indicate which emoticon fits you best **at this moment** by giving the corresponding number.

1 2 3 4 5

- 1
- 2
- 3
- 4
- 5

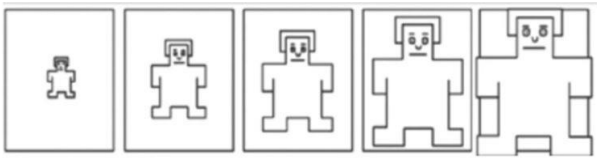
2. Below you can see different emoticons from relaxed to highly-energetic. Please indicate which emoticon fits you best **at this moment** by giving the corresponding number.



1 2 3 4 5

- 1
- 2
- 3
- 4
- 5

3. Below you can see different emoticons from being controlled to fully in-control. Please indicate which emoticon fits you best **at this moment** by giving the corresponding number.



1 2 3 4 5

- 1
- 2
- 3
- 4
- 5

4. What is your current stress level?
    - Not at all stressed
    - Slightly stressed
    - Moderately stressed
    - Very stressed
    - Extremely stressed

5. What is your average stress level of the past hour?
    - Not at all stressed
    - Slightly stressed
    - Moderately stressed
    - Very stressed
    - Extremely stressed

6. Did you ate anything during the past hour? If yes, please write down the time at which you ate. If no, please write down "No".

7. Did you drink any alcohol or had any caffeine (black/green tea and/or coffee) intake during the past hour? If yes, please write down the time at which you the intake. If no, please fill in "No".

8. What is your current physical activity? Choose from the list below.
    - Lying
    - Sitting
    - Standing
    - Slow walking (<3km/hour)
    - Moderate to fast walking (>3km/hour)
    - Running
    - Cycling
    - Driving a car
    - Sports
    - Other activity

9. Which physical activities did you perform during the last hour? Choose from the list below, you can select multiple answers.
    - Lying
    - Sitting
    - Standing
    - Slow walking (<3km/hour)
    - Moderate to fast walking (>3km/hour)
    - Running
    - Cycling
    - Driving a car
    - Sports
    - Other activity

A.2 Median vs. Butterworth filter

Introduction

The photoplethysmography (PPG) signal may contain noises and artifacts in the signal due to motion, baseline drifts or random noise due to external influences such as electromagnetic interference [51]. Therefore, as first pre-processing step a filter is implemented. In the literature, a band-pass Butterworth filter or a median filter is used to filter the PPG signal of the wearable E4 [29,32]. However, there is no gold standard for the type of filter that should be used in this setting. Therefore, both filters will be analyzed as the filters have different frequency domains.

Theory

The median filter is a non-linear filter which replaces the sample with the median of the samples in the chosen window. The median filter can be used to reduce noise depending on the window size. If the window is large, then the frequency of the output signal is lower, thus median filter can remove high-frequency noise [52].

The Butterworth band-pass filter is a filter in which the frequency response is created as flat as possible. The magnitude is maximally flat in the beginning and decreases in the passband and stopband (Figure 7) [53,54]. However, the filter does not have a linear phase response, resulting in a phase delay in the output signal [55]. In Figure7 the shift in phase can be seen from 180 to -180 degrees. This can be prevented by shifting the signal back and forth with *filtfilt* function in Matlab.

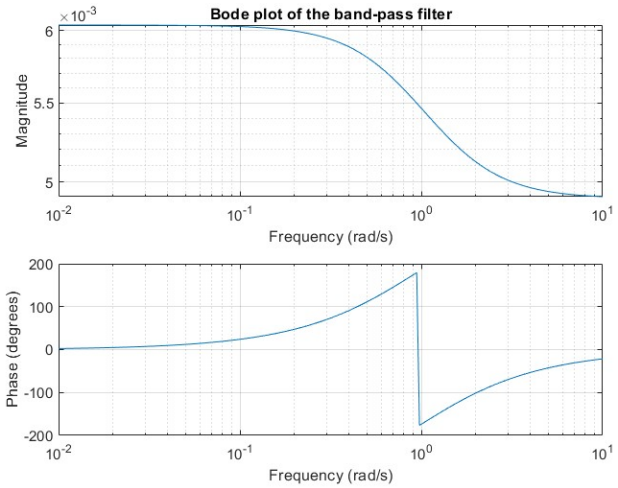


Figure 7: Bode plot of the band-pass filter

Method

To investigate which filter is most appropriate, three types of noises are added to a reference signal. The reference signal is a part of a night, since then it is assumed that the signal is most “noise-free”, because of minimal movement artifacts and relaxation of the subject. The following three noises are added to the signals: Random white noise, baseline noise and movement noise. Random white Gaussian noise was created with the function *agwn* in Matlab with a signal to noise ratio of 10 (y = awgn(x,10,'measured')), baseline nose was created by adding linear slope (rate of slope of 0.1) to the reference signal and the movement noise was created by adding a sine function (frequency = 1.8 Hz based on walking frequency [56], amplitude = 40) to the reference signal. The noise characteristics were arbitrary chosen. The same participants are used as described in the article above.

Results

In Figure 8 and 9 only the results of one participant is shown, but results of other participants were comparable. The band-pass filter reduces the amplitude of the peaks and does not include the dicrotic notch in the output signal (Figure 8). The median filter is almost identical to the reference signal. However, the dicrotic notch is not of interest in this study.

Results showed that the band-pass filter has a reduced amplitude compared to signal corrupted with noise (Figure 9). The amplitude of the median filter is closer to the signal corrupted with noise. Both filters eliminate the baseline noise. Furthermore, it can be seen that the white noise filtered by the median filter adopts the shape of random white noise signal, but it does reduce the noisy peaks. The band-pass filter eliminates the noise in the entire signal. Finally, it can be observed that the median filter is close to the shape of the corrupted signal with movement noise. The band-pass filter eliminates the shape of the movement noise.

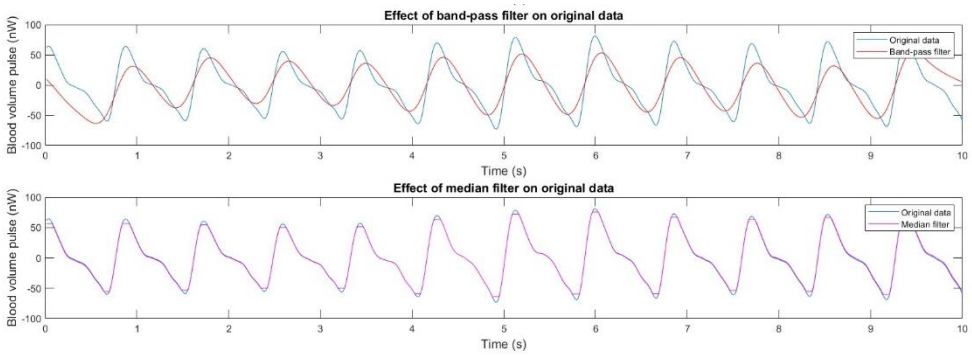


Figure 8: Effect of the band-pass and median filter on original data

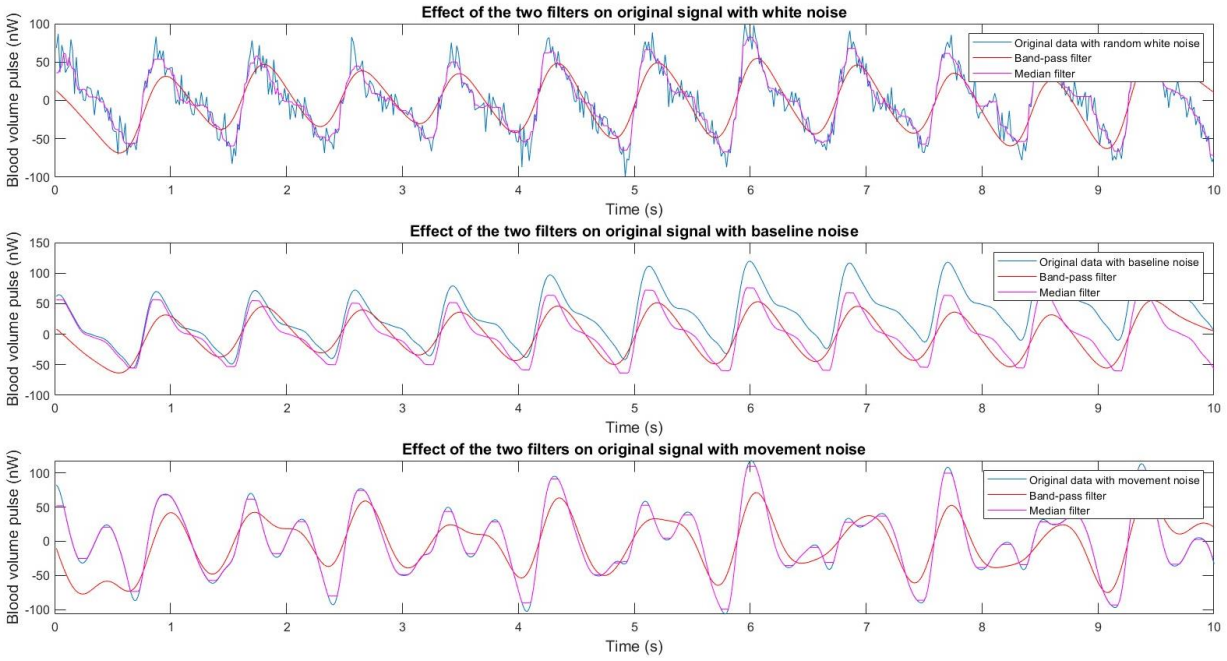


Figure 9: Effect of band-pass and median filter on white noise, baseline noise and movement noise

As it is difficult to judge based on visualization with the corrupted signal, Bland-Altman plots were created of the filtered signals which were corrupted with noise compared to the reference signal. The median filter shows smaller differences to the original signal compared to the band-pass filter, especially by the white noise and baseline noise (Figure 10). Only by the third participant more differences between median filter and original signal are seen, comparable to the band-pass filter.

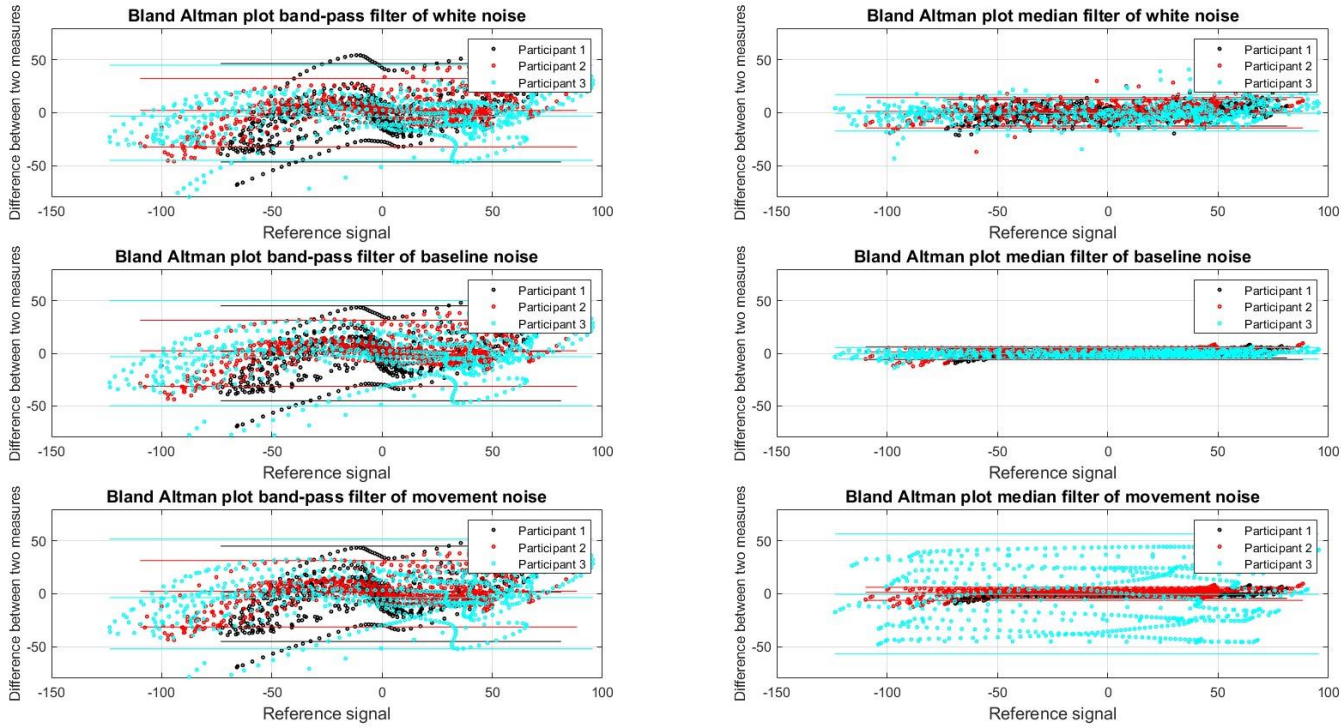


Figure 10: Bland Altman plots of band-pass filter and median filter of the three noises

Finally, since the features extracted from the PPG signal is based on peak detection, the last step is to compare the results in peak detection. The median filter is most close to the reference signal compared to the band-pass filter in both the heart rate (HR) and heart rate variability (HRV) (Table 3).

Table 3: Results from peak detection of the reference signal and median and band-pass filter with different noises

	HR (bpm)	**HRV (s)**
Original signal	70.6355	0.8494
Median filtered white noise	70.5175	0.8509
Band-pass filtered white noise	70.4587	0.8516
Median filtered baseline noise	70.4587	0.8516
Band-pass filtered baseline noise	70.4587	0.8516
Median filtered movement noise	71.2430	0.8422
Band-pass filtered movement noise	71.2430	0.8422

Conclusion

Based on the analysis, it was seen that the median filter is the most suitable filter to use in this case. The median filter is most close in HR and HRV compared to the reference signal, the shape is most close to the reference signal without adding noise to the signal. Furthermore, when looking at the Bland-Altman plots it was seen that the median filter was the closest to the reference signal with the random, baseline and movement noise.

A.3 Savitzky-Golay filter

The Savizky-Golay (SG) filter was used in this study to filter the skin conductance and skin temperature signals. This filter was most often used, especially with E4 devices [29,32]. The SG filter was based on fitting a polynomial with the local least-squares method, that results in smoothing the input data. The SG filter is capable of reducing noise while preserving the shape and amplitude of the signal, which why the SG filter is often used. It can be seen as a weighted moving average filter, in which the heightening is provided by the polynomial [57]. Frequency characteristics of the SG filter are that it has a very flat passband, with moderate attenuation in the stopband [58]. The result of filtering the skin conductance and skin temperature with SG-filter is shown in Figure 11. It can be seen that the SG-filter smooths the skin conductance signal by reducing sharp peaks in the signal and in the skin temperature signal the edges of the transitions between temperatures are smoothed.

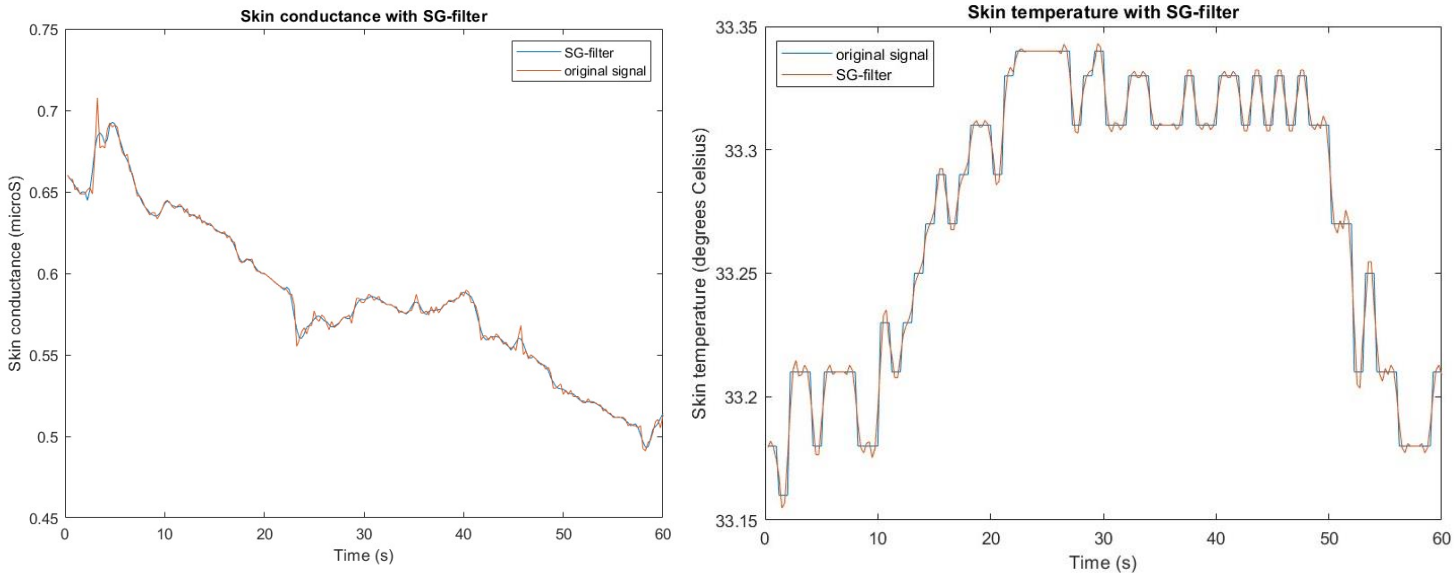


Figure 11: Skin conductance and skin temperature with SG-filter

A.4 Quality indicators

To detect artifacts in the physiological signals, quality indicators were used as shown in Figure 2 in the article.

Quality Indicator For PPG: Three quality indicators rules were applied: (1) Heart rate (HR) had to be between 40 and 180 beats per minute, (2) the maximal time between two PPG pulses may not exceed three seconds and (3) the ratio of the maximal time between two PPG pulses and the minimal time between two PPG pulses had to be less than 2.2. In Figure 12 examples of five second segments that were excluded due to meeting one of the three criteria are shown.

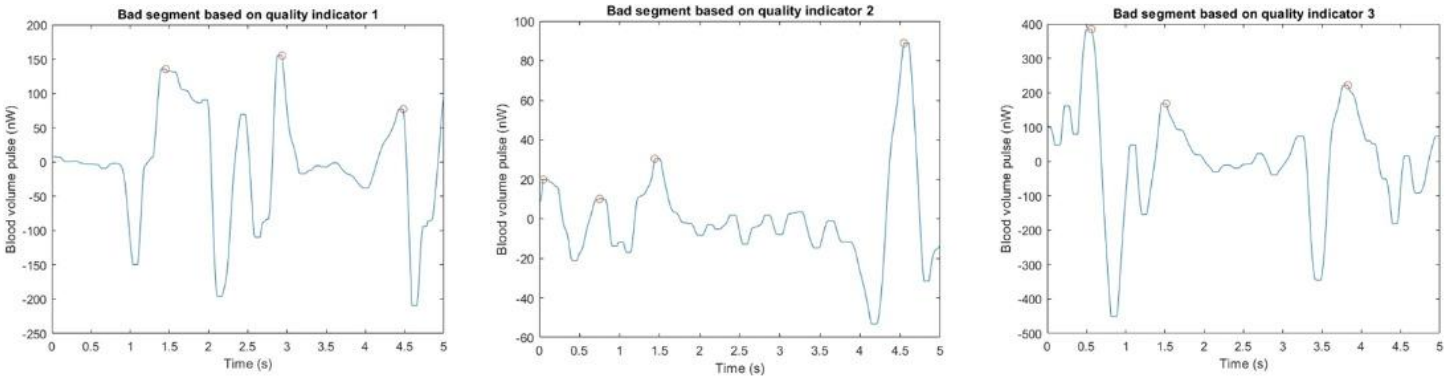


Figure 12: Examples of bad segments based on quality indicator of PPG signal

Next, a correlation template matching step was performed in which the threshold was set each day based on the first quartile of the distribution. If the mean of the correlation of the five second segment did not exceed the threshold, then the segment was classified as bad and not included in further analysis. The worst 25% of the correlated peaks was discarded each day. In Figure 13 bad and good segments based on the correlation template matching analysis are shown.

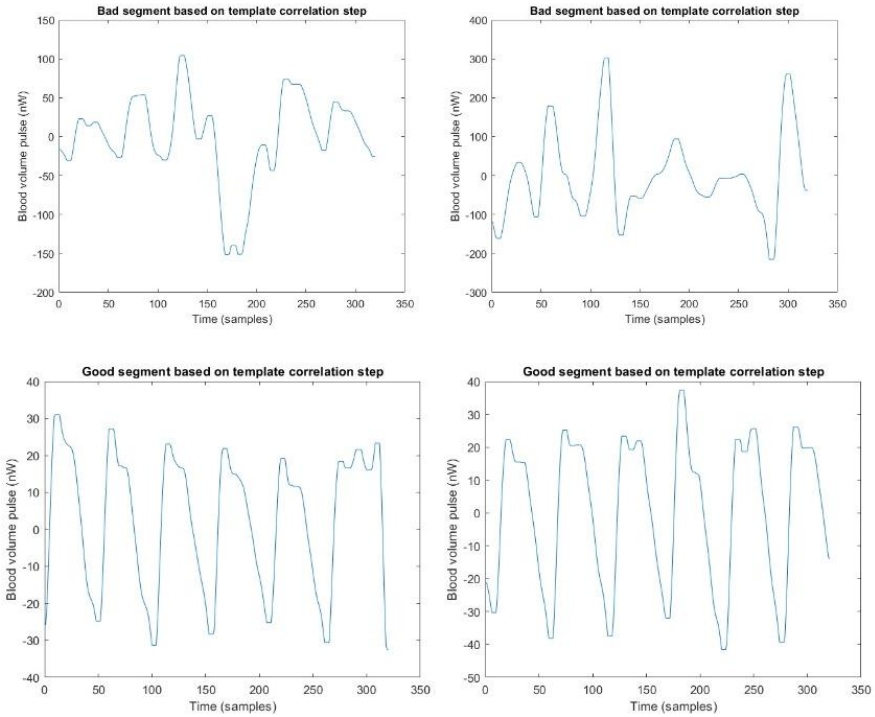
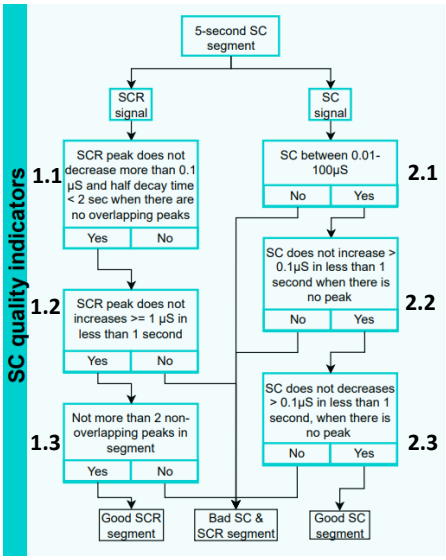


Figure 13: Examples of good and bad segments based on correlation template matching step in PPG processing

Quality Indicator For SC: There were quality indicators based on the whole signal (1), phasic and tonic, and quality indicators only based on skin conductance responses (2), the phasic part of the signal. These quality indicators can be seen in Figure 14 and also in Figure 2 of the article. Examples of five second segments that are classified as bad based on these quality indicators are shown in Figure 15.



Figuur 14: Skin conductance quality indicators

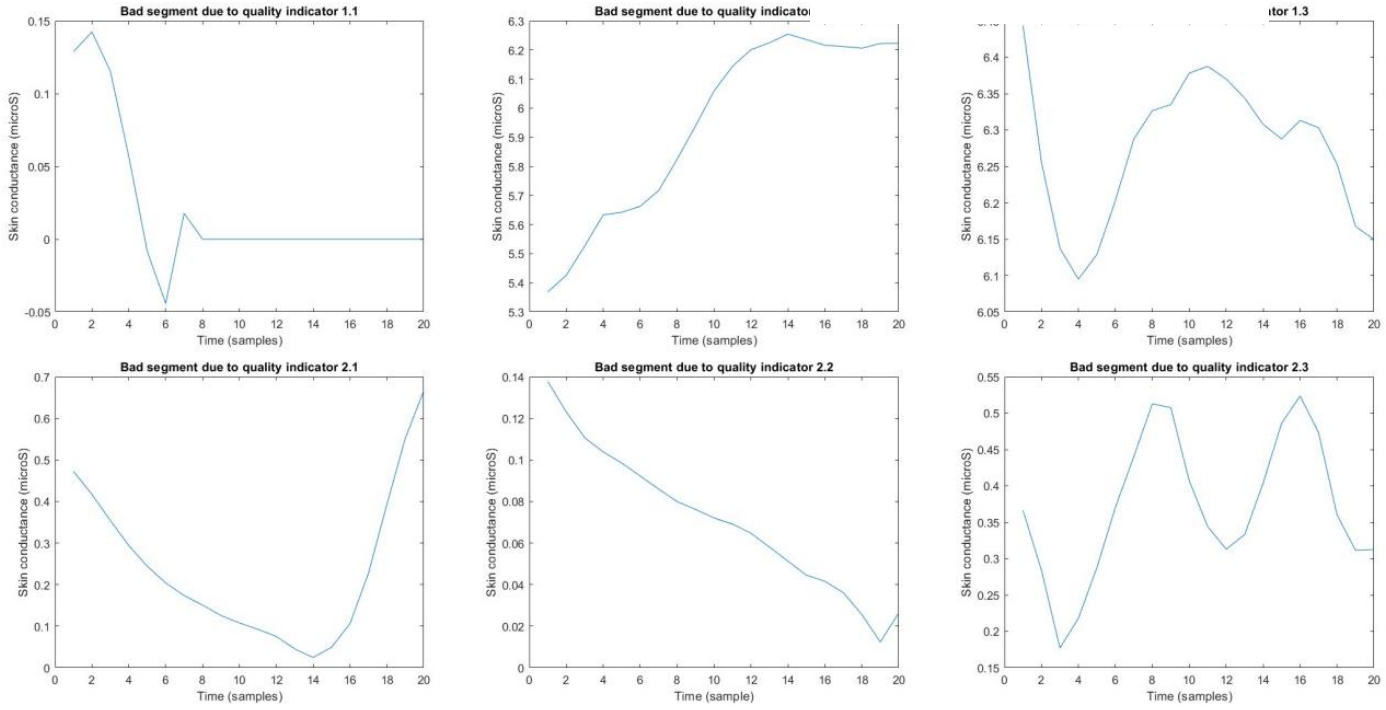


Figure 15: Examples of bad SC segments due to quality indicators

Quality Indicator For ST: The quality indicator of ST only consists of checking if the temperature is between 20 and 40 degrees Celsius in the five second segment (Figure 2). In Figure 16 an example is shown of a five second segment which did not met the quality indicator and is classified as a bad segment.

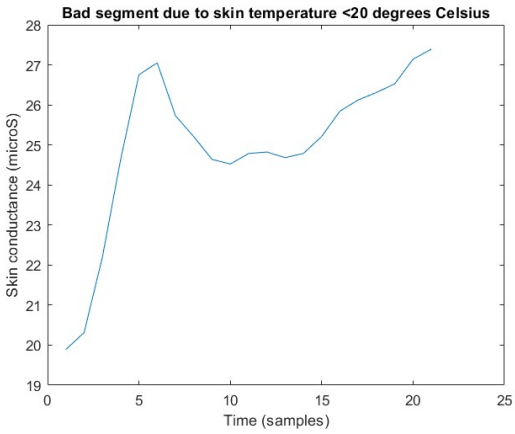


Figure 16: Example of bad segment due to quality indicator of ST

A.5 Results of linear model with absolute difference with 95% CI

In the three sub-sections (A.5.1-A.5.3) the results of the linear models for laboratory (Table 4), daily-life with current stress level (Table 5) and daily-life with the past one hour stress level (Table 6) are shown.

A.5.1 Laboratory results of the linear models

Table 4: Laboratory results of the linear models for each participant between the stress levels, the absolute difference with the 95% CI are presented

	S1 - S2 participant 1 (absolute difference, 95% CI)	S1 - S3-S4 participant 1 (absolute difference, 95% CI)	S2 – S3-S4 participant 1 (absolute difference, 95% CI)	S1 - S2 participant 2 (absolute difference, 95% CI)	S1 - S2 participant 3 (absolute difference, 95% CI)	S1 – S3-S4 participant 3 (absolute difference, 95% CI)	S2 – S3-S4 participant 3 (absolute difference, 95% CI)
mHR	-2.7127 (-1.27· 10^1, 7.35)	3.78 (-6.61, 1.41· 10^1)	6.49 (-0.38, 1.34· 10^1)	-5.28 (-1.38· 10^1, 3.21)	14.75 (4.67, 24.82)	6.22 (-3.85, 16.29)	-8.53 (-2.13· 10^1, 4.21)
SDNN	-0.02 (-0.20, 0.15)	-0.08 (-0.26, 0.10)	-0.06 (-0.18, 0.06)	0.07 (-0.11, 0.24)	0.11 (-0.05, 0.27)	0.11 (-0.05, 0.27)	-7.8· 10^{-17} (-0.20, 0.20)
pNN50	-0.24 (-0.67, 0.19)	-0.49 (-0.94, -0.04)	-0.25 (-0.55, 0.04)				
HF	-1.03· 10^5 (-1.03 · 10^4, 1.02· 10^4)	-1.02· 10^5 (-1.03 · 10^4, 1.02 · 10^4)	1.69 · 10^1 (-3.5 · 10^1, 6.9 · 10^1)	14.8 (-5.27· 10^1, 9.25· 10^1)	1.14· 10^2 (1.54 · 10^3, 1.77 · 10^3)	3.24· 10^2 (-1.33 · 10^3, 1.98 · 10^3)	2.09· 10^2 (-1.89 · 10^3, 2.31 · 10^3)
LF/HF	0.33 (-0.43, 1.08)	0.21 (-3.5· 10^{-3}, 6.9· 10^{-3})	-0.12 (-0.64, 0.39)	-0.13 (-0.51, 0.25)	-0.13 (-0.73, 0.47)	-0.09 (-0.69, 0.51)	0.04 (-0.72, 0.80)
mSCL	0.04 (-0.85, 0.93)	0.28 (-0.64, 1.20)	0.24 (-0.37, 0.85)	1.09 (0.31, – 1.86)	0.09 (-0.10, 0.28)	0.13 (-0.06, 0.32)	0.04 (-0.20, 0.29)
minSCL	-0.10 (-0.21, 0.01)	-0.08 (-0.19, 0.04)	0.02 (-0.05, 0.10)				
Absdiff2					1.1098· 10^{-4} (-6.3· 10^{-3}, 6.5· 10^{-3})	3.30· 10^{-3} (-3.1· 10^{-3}, 9.7· 10^{-3})	3.1845· 10^{-3} (-5.0· 10^{-3}, 0.01)
SCRR					-1.54· 10^{-5} (-0.26· 10^3, 0.22· 10^3)	3.01· 10^{-4} (-0.06· 10^3, 0.54 · 10^3)	3.12· 10^{-4} (0.01· 10^3, 0.62 · 10^3)
SCR dur					1.75· 10^3 (-1.07 · 10^4, 1.42 · 10^4)	4.95· 10^3 (-0.75· 10^4, 1.74· 10^4)	3.208· 10^3 (-1.25 · 10^4, 1.89 · 10^4)

SCR amp	-0.09 (-2.72, 2.54)	0.16 (-2.56, 2.87)	0.25 (-1.55, 2.05)	0.84 (-1.26, 2.94)	-0.04 (-0.35, 0.26)	0.41 (0.11, 0.71)	0.45 (0.07, 0.84)
SCR pow	-0.02 (-0.02, $-1.62\cdot 10^{-2}$)	-0.02 ($-1.8\cdot 10^{-2}$, $-1.57\cdot 10^{-2}$)	$5.99\cdot 10^{-5}$ ($-4.0\cdot 10^{-4}$,$1.60\cdot 10^{-3}$)	$1.01\cdot 10^{-3}$ ($-1.3\cdot 10^{-2}$, $3.3\cdot 10^{-2}$)	$7.60\cdot 10^{-5}$ ($-0.35\cdot 10^{3}$, $0.28\cdot 10^{3}$)	$2.48\cdot 10^{-4}$ ($-0.07\cdot 10^{3}$ $0.56\cdot 10^{3}$)	$2.86\cdot 10^{-4}$ ($-0.11\cdot 10^{-3}$ $0.68\cdot 10^{-3}$)
mST	2.01 (-1.37, 5.39)	2.02 (-1.47, 5.50)	$6.4122\cdot 10^{-3}$ (-2.30, 2.31)	-1.47 (-3.43, 0.49)	0.57 (-1.26, 2.41)	0.05 (-1.79, 1.88)	-0.53 (-2.84, 1.79)
SD ST	$-6.4918\cdot 10^{-3}$ (-0.05, 0.04)	-0.02 (-0.06 0.03)	-0.01 (-0.04, 0.02)	$-8.17\cdot 10^{-3}$ (-0.03, 0.02)	0.06 (-0.04, 0.1693)	$8.74\cdot 10^{-3}$ (-0.10 0.11)	-0.06 (-0.19, 0.08)
ST slope	$-1.7705\cdot 10^{3}$ ($-5.0\cdot 10^{-3}$,$1.5\cdot 10^{-3}$)	$-2.28\cdot 10^{-3}$ ($-5.6\cdot 10^{-3}$,$1.1\cdot 10^{-3}$)	$-5.14\cdot 10^{-4}$ ($-2.7\cdot 10^{-3}$,$1.7\cdot 10^{-3}$)				

A.5.2 Daily-life results of the current stress level of the linear models

Table 5: Daily-life results of the current stress level of the linear models for each participant between the stress levels, the absolute difference with the 95% CI are presented

	S1 – S2 participant 1 (absolute difference, 95% CI)	S1 – S3 participant 1 (absolute difference, 95% CI)	S2 – S3-S4 participant 1 (absolute difference, 95% CI)	S1 – S2 participant 2 (absolute difference, 95% CI)	S1 – S3 participant 2 (absolute difference, 95% CI)	S2 – S3-S4 participant 2 (absolute difference, 95% CI)	S1 – S2 participant 3 (absolute difference, 95% CI)	S1 – S3 participant 3 (absolute difference, 95% CI)	S2 – S3-S4 participant 3 (absolute difference, 95% CI)
mHR	0.07 (-13.62, 13.76)	-3.73 (-14.05, 6.58)	-3.80 (-18.88, 11.27)	-3.14 (-10.29, 4.00)	2.10 (-14.41, 18.60)	5.24 (-10.27, 20.75)	-1.14 (-11.01, 8.73)	-5.14 (-17.77, 7.48)	-4.00 (-17.24, 9.24)
SDNN	0.031 (-0.20, 0.26)	0.14 (-0.04, 0.31)	0.11 (-0.15, 0.360)	0.003079 (-0.18, 0.18)	-0.10615 (-0.18, 0.18)	-0.11 (-0.50, 0.28)	0.03 (-0.13, 0.20)	-0.0074197 (-0.22, 0.20)	-0.04 (-0.26, 0.18)
pNN50	0.022 (-0.68, 0.73)	0.317 (-0.21, 0.85)	0.30 (-0.48, 1.07)						
HF	1625.5 ($-1.51*10^{3}$,	1951.3 ($-0.41*10^{3}$,	325.78 ($-3.13*10^{3}$,	553.38 ($-1.07*10^{3}$,	-260.75 ($-4.00*10^{3}$,	-814.13 ($-4.33*10^{3}$,	-2056.4 ($-6.04*10^{3}$,	-2519.4 ($-7.60*10^{3}$,	-463.03 (-5.79, 4.86)

	$4.76*10^3$)	$4.31*10^3$)	$3.78*10^3$)	$2.17*10^3$)	$3.48*10^3$)	$2.70*10^3$)	$1.92*10^3$)	$2.56*10^3$)	
LF/HF	0.31 (0.07, 0.55)	-0.03 (-0.21, 0.16)	-0.34 (-0.60, -0.07)	0.04 (-0.04, 0.13)	-0.022 (-0.23, 0.18)	-0.07 (-0.26, 0.13)	0.07 (-0.15, 0.28)	-0.03 (-0.30, 0.25)	-0.10 (-0.38, 0.19)
mSCL	0.91 (-4.06, 5.89)	0.65 (-3.10, 4.39)	-0.268 (-5.75, 5.21)	-0.05 (-0.31, 0.20)	0.05 (-0.55, 0.64)	0.10 (-0.46, 0.66)	-0.01 (-0.12, 0.10)	0.09 (-0.05, 0.23)	0.10 (-0.04, 0.24)
minSCL	-0.05 (-0.29, 0.18)	0.04 (-0.14, 0.22)	0.09 (-0.17, 0.35)						
Absdiff2							$-1.43*10^{-3}$ (-0.02, 0.02)	$-8.66*10^{-4}$ (-0.03, 0.03)	$5.63*10^{-4}$ (-0.03, 0.03)
SCRR							$1.49*10^{-3}$ ($-4.8*10^{-3}$, $7.8*10^{-3}$)	$4.77*10^{-3}$ ($-3.3*10^{-3}$, $1.29*10^{-2}$)	$3.27*10^{-3}$ ($-5.2*10^{-3}$, $1.18*10^{-2}$)
SCR dur							$5.08*10^{-4}$ ($-3.8*10^{-3}$,$4.8*10^{-3}$)	$-2.71*10^{-4}$ (-0.01, 0.01)	$-7.79*10^{-4}$ (-0.01, 0.01)
SCR amp	0.35 (-1.36, 2.07)	-0.26 (-1.55, 1.04)	-0.61 (-2.50, 1.28)	0.35 (-0.03, 0.70)	-0.30 (-1.15, 0.54)	-0.64 (-1.43, 0.16)	-0.25 (-0.75, 0.26)	0.15 (-0.50, 0.80)	0.39 (-0.29, 1.08)
SCR pow	-0.03 (-0.29, 0.22)	-0.07 (-0.26, 0.12)	-0.04 (-0.32, 0.24)	$3.45*10^{-3}$ ($-1.7*10^{-3}$ 0.01)	$-8.49*10^{-4}$ (-0.01, 0.01)	$-4.29*10^{-3}$ ($-1*10^{-4}$, $1*10^{-4}$)	$5.08*10^{-4}$ ($3.8*10^{-3}$, $4.8*10^{-3}$)	$-2.71*10^{-4}$ ($-5.7*10^{-3}$ $5.2*10^{-3}$)	$-7.79*10^{-4}$ ($-6.5*10^{-3}$, $5*10^{-3}$)
mST	-1.97 (-4.37, 0.43)	-0.27 (-2.08, 1.54)	1.70 (-0.94, 4.35)	0.15 (-2.88, 3.19)	2.81 (-4.19, 9.82)	2.66 (-3.92, 9.24)	-2.43 (-6.85, 1.99)	0.64 (-5.01, 6.30)	3.07 (-2.86, 9.00)
SD ST	-0.04 (-0.12 0.04)	-0.01 (-0.07, 0.05)	0.03 (-0.06, 0.12)	-0.01 (-0.06, 0.04)	$-2.1374*10^{-3}$ (-0.11, 0.11)	0.01 (-0.09, 0.11)	-0.05 (-0.13, 0.03)	$-6.44*10^{-4}$ (-0.10, 0.10)	0.05 (-0.06, 0.16)
ST slope	$3.1823*10^{-3}$ ($-3.2*10^{-3}$, 0.01)	0.01 ($1.2*10^{-3}$, 0.01)	$2.7963*10^{-3}$ ($-4.2*10^{-3}$ 0.01)						

A.5.3 Daily-life results of the past one hour stress level of the linear models

Table 6: Daily-life results with the past one hour stress level of the linear models, the absolute differences with 95% CI are presented

	S1 – S2 participant 1 (absolute difference, 95% CI)	S1 – S3 participant 1 (absolute difference, 95% CI)	S2 – S3-S4 participant 1 (absolute difference, 95% CI)	S1 – S2 participant 2 (absolute difference, 95% CI)	S1 – S3 participant 2 (absolute difference, 95% CI)	S2 – S3-S4 participant 2 (absolute difference, 95% CI)	S1 – S2 participant 3 (absolute difference, 95% CI)	S1 – S3 participant 3 (absolute difference, 95% CI)	S2 – S3-S4 participant 3 (absolute difference, 95% CI)
mHR	0.26 (-5.47, 6.00)	-2.14 (-8.22, 3.95)	-2.40 (-9.24, 4.44)	-1.96 (-7.65, 3.73)	0.90 ($-1.36\cdot 10^1$, $1.54\cdot 10^1$)	2.86 ($-1.12\cdot 10^1$, $1.69\cdot 10^1$)	-5.49 ($-1.38\cdot 10^1$ 2.82)	-3.85 ($-1.21\cdot 10^1$ 4.39)	1.63 (-8.72, $1.20\cdot 10^1$)
SDNN	-0.04 (-0.16 0.09)	0.09 (-0.04 0.23)	0.13 (-0.02 0.28)	-0.03 (-0.14, 0.09)	-0.11 (-0.40, 0.19)	-0.08 (-0.37, 0.21)	-0.11 (-0.23, 0.02)	$2.24\cdot 10^{-3}$ (-0.13, 0.13)	0.11 (-0.05 0.27)
pNN50	-0.03 (-0.30, 0.23)	0.18 (-0.10, 0.47)	0.21 (-0.10, 0.53)						
HF	$-4.44\cdot 10^2$ ($-1.46\cdot 10^3$, $0.57\cdot 10^3$)	$3.00\cdot 10^2$ ($-0.78\cdot 10^3$, $1.38\cdot 10^3$)	$7.45\cdot 10^2$ ($-0.47\cdot 10^3$, $1.96\cdot 10^3$)	$-3.41\cdot 10^2$ ($-1.04\cdot 10^3$, $0.36\cdot 10^3$)	-6.02 ($-2.37\cdot 10^3$, $1.17\cdot 10^3$)	$-2.61\cdot 10^2$ ($-1.98\cdot 10^3$, $1.45\cdot 10^3$)	$-9.12\cdot 10^2$ ($-2.63\cdot 10^3$, $0.80\cdot 10^3$)	$-3.34\cdot 10^2$ ($-2.04\cdot 10^3$, $1.37\cdot 10^3$)	$5.77\cdot 10^2$ ($-1.56\cdot 10^3$, $2.72\cdot 10^3$)
LF/HF	$6.74\cdot 10^{-3}$ (-0.04 0.05)	$-1.67\cdot 10^{-2}$ (-0.06, 0.03)	-0.02 (-0.07 0.03)	0.03 (-0.02 0.08)	-0.03 (-0.16, 0.10)	$-6.52\cdot 10^{-2}$ (-0.19, 0.06)	0.033 (-0.09, 0.15)	$-6.63\cdot 10^{-3}$ (-0.12, 0.11)	-0.04 (-0.19, 0.11)
mSCL	-0.51 (-2.00 0.98)	-1.07 (-2.65, 0.51)	-0.56 (-2.33, 1.21)	-0.04 (-0.33, 0.25)	-0.11 (-0.84, 0.63)	-0.06 (-0.78, 0.65)	-0.06 (-0.28, 0.17)	0.05 (-0.18, 0.27)	0.06 (-0.17, 0.28)
minSCL	-0.02 (-0.07, 0.03)	-0.03 (-0.08, 0.02)	$-8.08\cdot 10^{-3}$ (-0.06, 0.05)						
Absdiff2							$-8.43\cdot 10^2$ ($-1.97\cdot 10^3$, $0.28\cdot 10^3$)	64.19 ($-1.05\cdot 10^3$, $1.18\cdot 10^3$)	$9.07\cdot 10^2$ (-0.49, 2.31)
SCRR							$-8.68\cdot 10^{-3}$ ($-0.02\cdot 10^{-3}$, $5.8\cdot 10^{-3}$)	$-1.84\cdot 10^{-3}$ ($-1.62\cdot 10^{-2}$, $1.25\cdot 10^{-2}$)	$6.83\cdot 10^{-3}$ (-0.01, 0.02)
SCR dur							-25.89 (-61.32, 9.54)	2.96 (-32.22, 38.14)	28.85 (-15.29, 72.99)
SCR amp	8.12 (-76.01 92.24)	-40.59 ($-1.30\cdot 10^2$, 48.67)	-48.71 ($-1.49\cdot 10^2$, 51.55)	-10.09 (-30.54 10.36)	-20.1 (-72.15 31.96)	-10.01 (-60.42, 40.40)	-25.89 (-61.32, 9.54)	2.96 (-32.22, 38.14)	28.85 (-15.29, 72.99)
SCR pow	-0.01 (-0.04 0.02)	-0.02 (-0.05, $6.8\cdot 10^{-2}$)	-0.01 (-0.04, 0.02)	$-1.183\cdot 10^{-3}$ ($-3.9\cdot 10^{-2}$,	$-1.53\cdot 10^{-3}$ ($-8.50\cdot 10^{-3}$, $5.40\cdot 10^{-3}$)	$-3.44\cdot 10^{-3}$ ($-7.1\cdot 10^{-3}$, $6.4\cdot 10^{-3}$)	$-5.41\cdot 10^{-3}$ ($1.34\cdot 10^{-2}$, $2.60\cdot 10^{-3}$)	$-2.94\cdot 10^{-3}$ (-0.01, $2.6\cdot 10^{-3}$)	$5.4052\cdot 10^{-3}$ ($-2.6\cdot 10^{-3}$, $1.34\cdot 10^{-2}$)

				1.5· 10^{-2})					
mST	0.58 (-1.04 2.19)	-0.13 (-1.83, 1.59)	-0.70 (-2.62, 1.22)	0.96 (-1.71, 3.62)	2.37 (-4.42, 9.15)	1.41 (-5.16, 7.98)	-0.61 (-3.01, 1.78)	-0.37 (-2.75, 2.01)	0.61 (-1.78, 3.01)
SD ST	5.252· 10^{-3} (-1.75· 10^{-2}, 2.80· 10^{-2})	2.89· 10^{-2} (4.6· 10^{-3}, 0.05)	0.02 (-3.6· 10^{-3}, 0.05)	-0.01 (-3.12· 10^{-2}, 6.1· 10^{-3})	-9.52· 10^{-4} (-5.7· 10^{-2}, 0.04)	3.01 · 10^{-3} (-0.04, 0.05)	-4.85· 10^{-3} (-0.04, 0.03)	-1.2· 10^{-2} (-0.05, 0.03)	-7.20· 10^{-3} (-5.64· 10^{-2} -4.20· 10^{-2})
ST slope	-5.52· 10^{-4} (-1.7· 10^{-3}, 6.0· 10^{-4})	-9.00· 10^{-4} (-2.1· 10^{-3} 3.0· 10^{-4})	-3.47· 10^{-4} (-1.7· 10^{-3}, 1.0· 10^{-3})						

Chapter 2: Background

2.1 Introduction

In the recent years, the prevalence of stress in the worldwide population has been increasing. In 2013 51% European working people reported that stress often occurs in work situations [1]. Stress can lead to positive short-term effects such as having more motivation or excitement. However, long-term stress can lead to chronic diseases such as diabetes [2]. In order to prevent stress, it has to be captured first[3]. Stress can be estimated by different methods, arising from the mechanisms behind the stress response. The gold standard for measuring stress is by the use of questionnaires, these are subjective assessments of stress and only provide a snapshot of someone's stress level [1,3,4]. A more objective method to capture stress is by measuring hormones such as cortisol, but this requires invasive sampling. With the rise of wearable sensors, the focus of stress assessment has shifted towards capturing stress using physiological signals. Since wearable sensors can measure physiological signals continuously, the door is opened to capture stress in daily-life [3].

2.2 Definition of stress

It is important to define stress, but there is no universal definition of stress. The first definition of stress was created by Hans Selye as "the nonspecific response of the body to any demand" [5]. Since the first definition, many different definitions have been created for stress depending on the context of the research. These definitions can be summarized as three types: Stress generated by surroundings, stress as physiological response to a stressor (similar to Seyle's definition) and stress as interaction between surroundings and the individual [6,7]. These definitions have all suffered criticism. The first two definitions were criticized for not acknowledging that the stress response can differ for stressors and stress response is person specific [7]. As a result, the definition of stress in research has shifted to looking at the interaction between the individual and the environment, often referred to as psychological stress. The definition of stress in this study is taken from Liu et al. [8]: "Stress is a state of threatened homeostasis provoked by a psychological, environmental, or physiological stressor." This definition is chosen, since it takes all three definitions and criticism into account.

A distinction can be made between acute and chronic stress. Acute stress arises from an incident and is only enduring for a short period of time, such as finishing work before a deadline. If the incidents are repeatedly occurring, the stress response to the incident will become chronic at a certain point [9]. Chronic stress arises due to a prolonged imbalance and induces negative mental and physical symptoms, influencing blood pressure and cholesterol levels. Chronic stress especially induces negative symptoms if the individual has poor coping skills [7]. The response to stress consists of three slow to fast axes, which regulates the acute to chronic stress response.

2.3 Physiological mechanisms of stress response

There are three mechanisms that regulate the stress response from acute to chronic, which are: neuronal, neuroendocrine and endocrine axes. Neural axes provides an acute response, neuroendocrine acute to chronic response and endocrine axes regulates a chronic response [10]. These three axes will be described below.

The autonomic nervous systems (ANS) is responsible for the regulation of homeostasis and the internal environment of the body. Therefore, it regulates different physiological activities in the body, such as the heart rate, muscles and gland secretion [4]. The ANS consists of the

parasympathetic nervous system (PSNS) and the sympathetic nervous system (SNS). The PSNS cares for relaxation and stabilization of the body, whereas the SNS makes the body ready for action [4,10]. When an individual is experiencing stress, the SNS is activated by the paraventricular nucleus of the hypothalamus. These are the first systems that are activated by a stress response, because of the fast neural pathways [10]. Next to the SNS, two more neural axes form the response to stress, namely the parasympathetic nervous system (PSNS) and neuromuscular nervous system. The PSNS inhibits or slows down functions in the targeted organs [10]. Besides, the neuromuscular system can be activated as a result of a stressor, when this is done excessively it can lead to neuromuscular dysfunctions [10]. The largest effects of the neurological axes are immediate and the effects do not have the potential to prolong to chronic effects, since the axon branches are not able to continue releasing neurotransmitters under high stimulation.

In order to maintain a more chronic reaction to stress [11], another physiological mechanism is activated, which is the neuroendocrine axis [10]. The neuroendocrine axis is activated via the SNS. With this mechanism, the body is ensuring to have enough energy for the "fight-or-flight" response to control the stress and to prevent losing homeostasis [11]. The "fight-or-flight" response leads to increased heart rate and sweat gland activity amongst other physiological effects [4]. The effect of the norepinephrine and epinephrine is an increased adrenergic activity in the body, which have similar effects as the direct innervation from the sympathetic nervous system [10]. The epinephrine's have a delay of thirty seconds before measurable effects can be detected, due to using systemic circulatory system as transport mechanism instead of neural pathways [10]. The response of the neuroendocrine axis has an ten times longer response duration compared to neural axes, it can range from intermediate to chronic stress responses [10]. The mechanism to prepare the body for the fight or flight response has also been referred to as the sympathetic adrenal medullary system (SAM).

At the same time that the paraventricular nucleus activates the SNS, the corticotropin-releasing factor (CRF) and arginine vasopressin (AVP) hormones are released in the adrenal cortical axes by the paraventricular nucleus [1,10,12]. These hormones encourage the pituitary gland to create the hormone adrenocorticotropic (ACTH). The function of ACTH is then to synthesize and release the glucocorticoids out of the adrenal cortex, also called the hypothalamic-pituitary-adrenal (HPA) axis [1,10]. Cortisol creates negative feedback loops by releasing ACTH, reducing the paraventricular nucleus activity and reducing the production of CRF in the hippocampus [1]. The endocrine axes are the last pathway to react to the stressor, because it is almost fully dependent on the circulatory system for transport and the need of higher intensity of the stimulus for activation and can result in chronic stress [10].

In summary, the neuronal axes are firstly activated and lead to an immediate response. After that, the neuroendocrine axis is activated leading to an intermediate response, which can potentially be prolonged to a chronic response. Lastly, the endocrine axes are activated and provide a chronic response to the stressor. An overview of the three pathways and their response can be seen in Figure 17 [10]. For acute stress the neuronal axes and the neuroendocrine axes are the most important pathways to form the response. For chronic stress, the endocrine axes is the key pathway that create the response, but the neuroendocrine axis can also play a role. Thus, the different mechanisms for stress response can overlap. The neuroendocrine mechanism and the endocrine axes are most common to overlap since both may respond to chronic stressors [10].

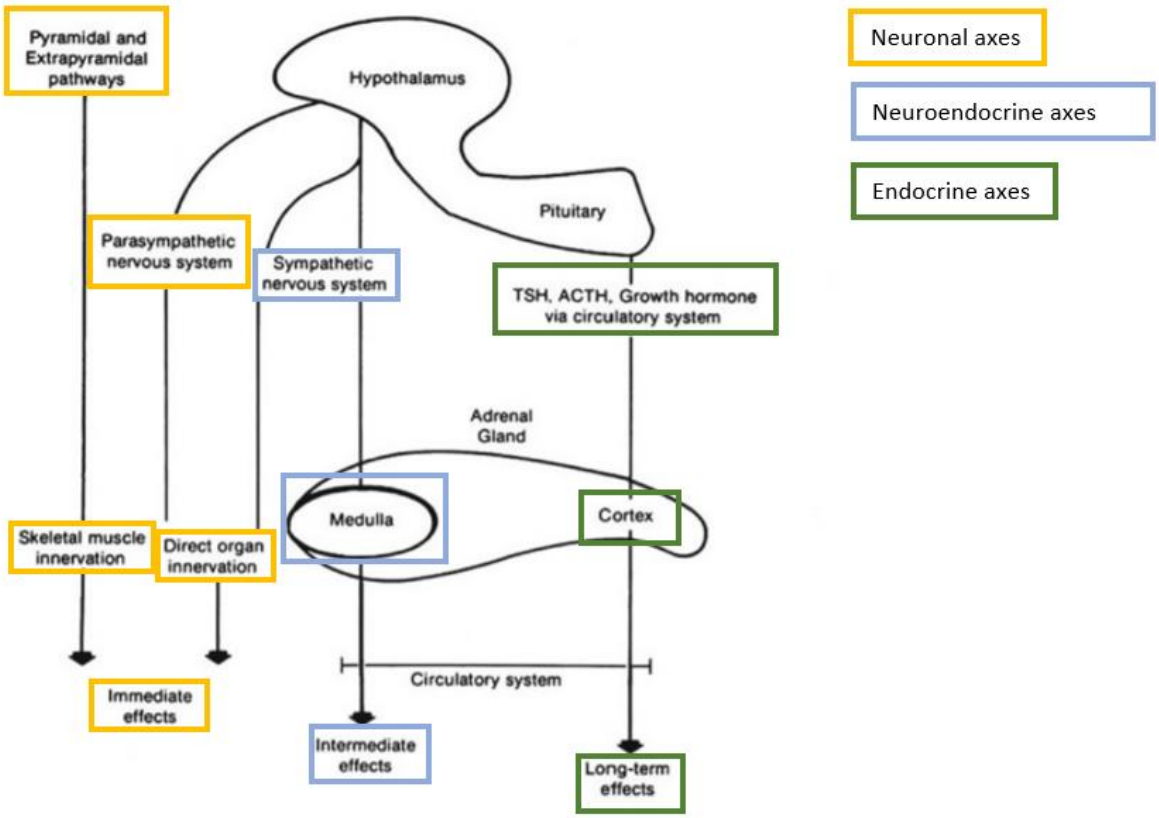


Figure 17: The three pathways of the stress response, taken from [8]. The neural axes is formed by the parasympathetic nervous system, sympathetic nervous system and pyramidal and extrapyramidal pathways, leading to an immediate response. The neuroendocrine axes pathway is formed by the sympathetic nervous system and medulla, leading to intermediate responses. Lastly, the pathway of the endocrine axes is formed by pituitary via the cortex leading to long-term responses.

2.4 Physiological signals related to stress

The stress response is created by three pathways: Neutral axes, neuroendocrine axis and endocrine axis. The neural axes response can be measured with physiological signals. However, these responses are only immediate and do not last long. The neuroendocrine axes produces a response for a longer time (immediate effects) and can also be measured by these physiological signals. Next to that, adrenal medullary catecholamines could be assessed by blood analyses, but this is invasive. The stress response of the endocrine axis is mostly captured by measuring cortisol, which is also invasive [4,10]. Therefore, physiological signals are used to capture stress which are primarily created by the neuroendocrine axes.

Stress responses can be measured through physiological signals, such as; blood volume pulse, skin conductance, blood pressure and skin temperature [1]. In this thesis, the focus is on these signals as these are easy and accurately measurable with a wearable device [1]. In the following sections each of these signals will be discussed in relation to stress. Additionally, the relevant characteristics of the physiological signals, also called features, will be covered and can be seen in Table 4. A choice between mean/median and standard deviation/interquartile range will be made based on the distribution of the feature.

Table 4: Features of the three physiological signals and feature abbreviation with references, which are studies in which the features were used for stress detection.

Feature abbreviation	Full name of feature	Feature description	Extracted from type signal	Reference
mHR	Mean/median heart rate	Measures the overall effect of the SNS (increase heart rate) and PNS (decrease heart rate) [1,13].	PPG	[3] [4] [14] [15]
HRV	Heart rate variability	The variation in the duration of the normal-to-normal interval [16], it captures neurocardiac function which is the interaction between the heart and the brain. This interaction is initiated by processes of ANS and heart-brain interconnection [13,17].	PPG	[4] [18] [19] [20] [14] [15]
SDNN	Standard deviation/interquartile range of normal-to-normal intervals	Measures total HRV that arises from all the factors (random and recurring sources) that contribute to HRV [13,16].	PPG	[7] [3] [14] [21]
pNN50	Proportion of the successive normal-to-normal intervals that differ more than 50 ms	Extracts periodic variability from baseline heart rate [16].	PPG	[22] [14] [21]
RMSSD	Root of the mean/median of the sum of squares of differences in normal-to-normal intervals	Measure of the variance in normal-to-normal intervals of heat rate and estimates PNS changes due to breathing in heart rate variability [13,17].	PPG	[3] [22] [14]
LF	Heart rate variability in the low frequency band (0.04-0.15 Hz)	Index for sympathetic modulations and activity of vagus nerve (major nerve fiber in PNS) to regulation of cardiac processes [4,23].	PPG	[20]
HF	Heart rate variability in the high frequency band (0.15-0.4 Hz)	Measure of parasympathetic activity, resembles variations in heart rate due to breathing also known as respiratory sinus arrhythmia [13].	PPG	[20]
LF/HF	Ratio of the low and high frequency band of heart rate variability	Measure of the sympathovagal balance, which means the influence of SNS and PNS on an autonomic state [4].	PPG	[7] [3] [20]
mSCL	Mean/median skin conductance level	Mean tonic component of skin conductance, which is changes in order of minutes, reflects general arousal level of an individual and depends on ambient temperature for example [24,25].	SC	[26] [27]
minSCL	Minimum skin conductance level	The lowest value of the tonic component of SC.	SC	[22]

SD SCL	Standard deviation/interquartile range of skin conductance level	Measure of variation in the tonic component of SC, reflecting change in general arousal level and ambient temperature for example [25].	SC	[22]
SCR diff2	Signal power of galvanic skin response in the second difference	Measure of the fast high-frequency changes in the responses by investigating energy of the phasic part of the SC signal of the second difference [28].	SC	[29] [27]
mSCR	Mean/median skin conductance response	Measure that reflects the stimulation of the SNS, which results in higher-frequency variability in SC (phasic component) [25].	SC	[4]
SCRR	Skin conductance response rate	Measure of the number of SCRs in a specific time frame divided over length of that time frame. It measures the frequency of the physiological response to stressors [4,24].	SC	[4] [22] [30]
SCR dur	Total duration of the skin conductance response	It is the sum of the duration of the SCRs. It can be defined as the time from the start of the increase in skin conductance (as caused by production of sweat) till there is an decrease in SC [4].	SC	[4]
SCR mag	Total magnitude of the skin conductance response	Measure of the strength of the firing of the sympathetic nerve in the skin [4,31].	SC	[4] [30]
SCR area	Sum of the area of skin conductance responses	Determined by the amplitude and duration of the SCRs [3].	SC	[3]
SCR pow	Signal power of the skin conductance response	Measure of the signal energy of the physiological response to certain stressor [3,28].	SC	[3]
mST	Mean/Median skin temperature	Measure of direct connection with the blood flow in blood vessels [32].	ST	[3] [33] [34]
SD ST	Standard deviation/interquartile range of the skin temperature	Change in temperature or blood flow in blood vessels [32].	ST	[3]
ST slope	Slope of the skin temperature	Slope of skin temperature is determined by a linear line fitted to the data. It indicates when there is rise or fall in temperature and blood flow in blood vessels to the skin [35].	ST	[3]

SNS: Sympathetic Nervous System, PNS: Parasympathetic Nervous System, ANS: Autonomic Nervous System, HRV: Heart rate variability, SCRs: Skin Conductance Responses, PPG: Photoplethysmography, SC: Skin Conductance and ST: Skin Temperature

2.4.1 Photoplethysmography

The photoplethysmography (PPG) measures the variation in blood flow volume. From the PPG, the heart rate (HR) and heart rate variability (HRV), the beat-to-beat interval, can be derived. The PSNS

controls HR and HRV in rest conditions, leading to a decreased HR and an increased HRV due to vagal activity. The vagus nerve manages involuntary body functions such as digestion. Whereas, in stressed situations, the SNS is dominant leading to increased HR and decreased HRV as result of increased activity of sympathetic path and reduction of vagal activity [36].

The features of the PPG can be divided into time and frequency domain features. The mean HR increases during stressed moments in laboratory setting [4] and daily-life [3]. HRV is also an important time domain feature for capturing stress in laboratory setting [14,15], it decreases under stressed conditions [19,20]. In daily-life stress detection models, HRV was also important features [18,21]. Next to that, root mean squared of the squared differences of normal-to-normal intervals (RMSSD) is related to stress in laboratory [14,22] and daily-life [3]. Another important parameter to capture stress is the standard deviation of the normal-to-normal intervals (SDNN) in laboratory [14,22] and daily-life [3,7,21]. Furthermore, the proportion of the successive normal-to-normal intervals that differ more than 50ms (pNN50) showed a significant difference between stressed and rest moments in laboratory [14] and daily-life [21].

The HRV can also be determined in frequency domain in which low frequency (LF) band is 0.04-15 Hz and high frequency (HF) band is 0.15-0.4 Hz. The HF band reflects the parasympathetic neural activity, LF corresponds to sympathetic and parasympathetic neural activity and the ratio LF/HF reflects sympathetic information [7]. LF is increased during stressed conditions in laboratory setting [20], whereas HF is decreased during stressed conditions in laboratory setting [20]. The ratio LF/HF is associated to stress in laboratory setting [20] and daily-life [7].

In summary, literature showed that time-based heart rate features (mean HR, HRV, SDNN, RMSSD and pNN50) and frequency based heart rate features (LF, HF and ratio LF/HF) were associated with stress responses in laboratory and daily-life settings.

2.4.2 Skin conductance

The second physiological signal is the skin conductance (SC), which captures the sweat production. The main function of the sweat production is thermoregulatory. However, sweating is also regulated by the SNS, which controls emotional sweating. The thermoregulatory sweat glands are only found in limited areas of the body, for example the arm pits, whereas the emotional sweat glands can be found throughout the body. The emotional sweat glands play the largest role of these three glands in sweat production [37]. The emotional sweat glands are innervated by the sympathetic nerves and therefore are involved in the emotional stress, next to the thermoregulation [30]. During a stress response, the SNS is activated resulting in the sweat glands filling up. As a result, the SC increases before the sweat is removed [31]. After the sweat is removed, the SC decreases [31] (Figure 18).

The skin conductance level (SCL) is the slowly varying part of the skin conductance or tonic component and the fast changing part is the skin conductance response (SCR) or phasic component [38]. SCRs are rapid changes in the SC signal as response to a neural stimulus from SNS. SCRs are specified as a peak (rise in SC) followed by a gradually decline due to removal of sweat (Figure 2 [39]). Different measures can be extracted from SCRs, for example the amplitude. SCL corresponds to slow changes in the SC signal and is often computed over a longer period of time. This means that SCL captures the general level of arousal of a subject [40], which increases when stress occurs due to the sweat glands filling up [1]. The change in SCL is normally between 1 and 6 µS [7,39]. Next, the features of both parts of the signal are discussed.

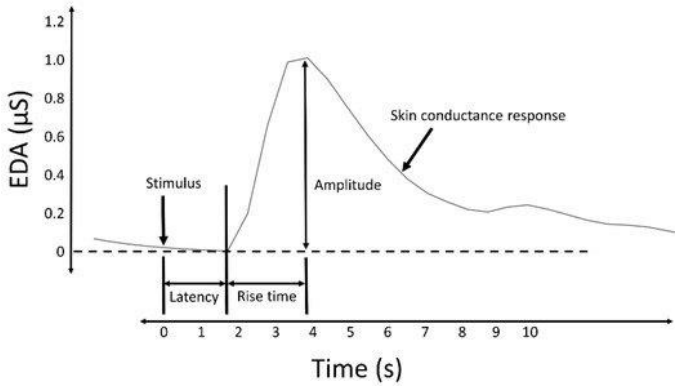


Figure 18: Skin conductance response [39]

The mean SCL is associated to stress in laboratory setting [26,27,29]. Also, minimum SCL is shown to be important in classifying stress in laboratory setting [22]. Furthermore, the standard deviation of SCL can be used in stress detection models in laboratory setting [22].

The SCR usually occurs within a couple of seconds after a neural stimulus. The amplitude of SCR is associated with a stress response in laboratory [4,30] and daily-life setting [41]. Also, the SCRs rate provides information about the stress response in laboratory setting [4,30]. Next to that, the total duration of SCRs increases during stressed conditions in laboratory setting [4]. Another important SCR feature for stress detection is the mean SCR, this was shown in laboratory setting [4,29]. Furthermore, the absolute second difference of SCR is important in classification models of stress detection in laboratory [27,29]. The SCR area is shown to be higher during stressed conditions in daily-life [3]. Lastly, the signal power of SCR is related to stress in daily-life, in which a higher signal power is associated with higher levels of stress [3].

In summary, three SCL features (mSCL, minSCL and SD SCL) and seven SCR features (SCR amp, SCRR, SCR dur, mSCR, SCR diff2, SCR area and SCR pow) are related to stress in laboratory and daily-life settings.

2.4.3 Skin temperature

The last physiological signal used in this thesis is the skin temperature (ST). When a subject is exposed to stress, the blood flow to the skin will be reduced, due to direct increased SNS activity. The SNS increase results in narrowing the blood vessels, which leads under the presence of norepinephrine. Thus, the skin temperature is decreased in stressful situations due to the reduced blood flow [17,7].

The mean ST is shown to be responsive to stress in laboratory [33,34] and daily-life setting [3]. Also, the standard deviation of ST is related to stress in daily-life setting [3]. Finally, the slope of ST is higher during unstressed conditions compared to stressed conditions in daily-life [3].

2.4.4 Confounding factors to stress response

The stress response is mediated by the ANS, but the ANS also regulates numerous other processes leading to mediating factors for the stress response. One of these factors is physical activity, which activates the SNS comparable to the stress response. As a result, physical activity may blur the effect of the stress response on the physiological signals. In addition, physical activity distorts the physiological signals with movement artefacts, especially in ambulatory measurements. Consequently, specific quality indicators for the physiological signals are needed to detect these movement artefacts [6]. Not only physical activity can lead to inaccuracies in the physiological signals, also placement of the sensor and respiratory movement influence the accuracy of the signals

[4]. Another confounding factor is alcohol, since that activates the HPA axis and thus the ANS too [41]. Additionally, caffeine activates the ANS as well, which may lead to masking the effect of the stress response on the physiological signals [42]. Next to that, the intake of food also stimulates ANS, since it regulates metabolism and cardiovascular system after food intake. This means that food intake is another confounding factor for the stress response [43]. Finally, ambient temperature is a confounding factor for the skin conductance signal, as one of the pathways that is included in the skin conductance response is regulating thermoregulatory sweating. Other pathways include high physical activity, which increases sweating due to the generated body heat [39].

2.5 State of the Art

Previous research has shown that stress can be captured by physiological signals from wearable sensors in laboratory settings [1,3,7,44]. Also, high classification accuracies were achieved for stressed moments in laboratory based on physiological responses [29]. This implies that physiological responses capture the stressed moments in a controlled setting. However, the induced stress in a controlled setting is not comparable to the stress that occurs in daily-life. Additionally, the stress captured in laboratory is only done for a short amount of time. With the rise of the wearable sensors, the opportunity to measure stress in the home situation arises. Therefore, the focus of studying stress has shifted from laboratory towards stress detection in ambulatory settings. Detection of stress is vital as stress is a major contributing factor to chronic disease [8]. So, detecting stress can aid the improvement of lifestyle and prevention of chronic diseases.

In ambulatory setting, the past research can be divided into semi-ambulant, which is observing stress in a specific situation, and monitoring stress in daily-life setting [3,45,46]. Healey and Picard for example showed that skin conductance and heart rate had the highest association with the stress level of the participants while driving [47]. The SWEET study by Smets et al., showed associations between physiological signals and stress based on self-reports in daily-life [3]. Although these results show that it is promising to use physiological signals in daily-life for stress detection, the classification accuracies in daily-life are lower compared to laboratory settings [46].

The stress detection in daily-life is challenging due to confounding factors, lack of a continuous stress reference and person-specific stress responses [1]. Physical activity is an important confounding factor, since it influences the physiological signals and the quality of the signals also after the activity has ended [1]. Another challenge, is that there is no continuous stress reference for daily-life stress measurements to compare the results of the physiological signals to. In laboratory setting, the moments when a stressor is applied are used as reference for stressed moments. Whereas, in daily-life there is no continuous reference available when a person is stressed, how often the person is stressed or how long the person is stressed. To overcome the challenge of having no stress reference, stress research relies, for now, on self-reported questionnaires [4,7]. Ecological Momentary Assessments (EMA) are often used, which are self-reports that are repeatedly asked throughout the day to reduce recall bias. The EMAs can be time-based, send at random, specific moments, or event-based. Besides the reference for stress, another challenge is that stress responses tend to be person-specific. This was shown with a large-scale study by Smets et al [3]. This suggests that personal models could yield higher classification accuracies than general models, which was confirmed by an ambulatory study of Can et al [48]. This suggests the need for personalized techniques in ambulatory setting. Finally, in daily-life it is difficult to capture the baseline of one's physiological signals, which is a rest state of one's physiological signals. This is especially difficult, since this might change throughout the ambulatory study. In laboratory, the baseline is captured by first letting the participant rest for ten minutes before starting with the protocol for example. However, in daily-life it is more difficult to capture a baseline measurement, since it can depend on

someone's sleep for example. A method to capture the baseline is to measure each night, as done in an ambulatory study by Smets et al [3].

Personalized techniques for ambulatory settings were addressed in research by Smets et al., in which clusters on Montreal Imaging Stress Test in daily-life were created [6]. The results suggest that the type of features that are associated with a stress test in daily-life and daily-life stressors can be similar, which implies that these features can be a promising tool to improve the classification accuracy [6]. In addition, the results showed that the stress test may be useful for personalized calibration, or in other words the stress test may be used to capture a baseline of the physiological signals [6]. Although the results indicate that a stress test could be used for normalization and clustering of features, the models with information from the stress test showed low classification accuracies. So, further research is necessary to determine if the stress responses to the stress test and daily-life stressors are similar.

Next to using a stress test for improving ambulatory models, also models based on laboratory data were used. Hovspian et al [18]., created a model that was trained on laboratory data and applied it to laboratory and field data, leading to a prediction accuracy of 90% for laboratory data and 72% for the field data. This implies that the models built in laboratory show lower accuracies in daily-life, suggesting that the stress responses are different in laboratory compared to daily-life. Thus, to improve the accuracy of stress classification more information about the relation between the stress responses during laboratory stressors, stress test in daily-life and daily-life stressors is needed.

2.6 Scope of the thesis

This thesis focuses on (1) establishing a framework for a multi-sensor approach, and (2) comparing the stress-related features captured in laboratory setting to stress-related features captured in daily-life setting. With the aim to determine how the accuracy of stress detection in daily-life can be improved. Next to laboratory and daily-life stressors, this research also includes a stress test in daily-life to bridge the gap between a fully controlled and a fully uncontrolled setting. With this comparison, suggestions can be made for further research on how to improve the stress detection in daily-life.

References

[1] Smets E, De Raedt W, Van Hoof C. Into the Wild: The Challenges of Physiological Stress Detection in Laboratory and Ambulatory Settings. IEEE J Biomed Heal Informatics 2019;23:463–73. doi:10.1109/JBHI.2018.2883751.

[2] Plarre K, Raij A, Hossain SM, Ali AA, Nakajima M, Al'Absi M, Ertin E, Kamarck T, Kumar S, Scott M, Siewiorek D, Smailagic A, Wittmers LE. Continuous inference of psychological stress from sensory measurements collected in the natural environment. Proc 10th ACM/IEEE Int Conf Inf Process Sens Networks, IPSN'11 2011:97–108.

[3] Smets E, Rios Velazquez E, Schiavone G, Chakroun I, D'Hondt E, De Raedt W, Cornelis J, Janssens O, Van Hoecke S, Claes S, Van Diest I, Van Hoof C. Large-scale wearable data reveal digital phenotypes for daily-life stress detection. Npj Digit Med 2018 11 2018;1:1–10. doi:10.1038/s41746-018-0074-9.

[4] Sun F-T, Kuo C, Cheng H-T, Buthpitiya S, Collins P, Griss M. LNICST 76 - Activity-Aware Mental Stress Detection Using Physiological Sensors. LNICST 2012;76:282–301.

[5] Saturday L, Selye H. Stress and the General Adaptation Syndrome. Br Med J 1950;1:1383–92. doi:10.1136/BMJ.1.4667.1383.

[6] Smets E. Towards large-scale physiological stress detection in an ambulant environment 2018.

[7] Wijsman J. Sensing stress: stress detection from physiological variables in controlled and uncontrolled conditions 2014. doi:10.3990/1.9789036537858.

[8] Liu YZ, Wang YX, Jiang CL. Inflammation: The common pathway of stress-related diseases. Front Hum Neurosci 2017;11:316. doi:10.3389/FNHUM.2017.00316/BIBTEX.

[9] Rohleder N. Stress and inflammation – The need to address the gap in the transition between acute and chronic stress effects. Psychoneuroendocrinology 2019;105:164–71. doi:10.1016/J.PSYNEUEN.2019.02.021.

[10] Everly GS, Lating JM. The Anatomy and Physiology of the Human Stress Response. A Clin Guid to Treat Hum Stress Response 2002:15–48. doi:10.1007/0-306-47800-5_2.

[11] Fink G. Stress, Definitions, Mechanisms, and Effects Outlined: Lessons from Anxiety. Stress Concepts, Cogn Emot Behav Handb Stress 2016:3–11. doi:10.1016/B978-0-12-800951-2.00001-7.

[12] Agorastos A, Chrousos GP. The neuroendocrinology of stress: the stress-related continuum of chronic disease development. Mol Psychiatry 2021 2021:1–12. doi:10.1038/s41380-021-01224-9.

[13] Shaffer F, McCraty R, Zerr CL. A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability. Front Psychol 2014;0:1040. doi:10.3389/FPSYG.2014.01040.

[14] Salahuddin L, Cho J, Jeong MG, Kim D. Ultra Short Term Analysis of Heart Rate Variability for Monitoring Mental Stress in Mobile Settings. Undefined 2007:4656–9. doi:10.1109/IEMBS.2007.4353378.

[15] Hickey BA, Chalmers T, Newton P, Lin CT, Sibbritt D, McLachlan CS, Clifton-Bligh R, Morley J, Lal S. Smart Devices and Wearable Technologies to Detect and Monitor Mental Health Conditions and Stress: A Systematic Review. Sensors (Basel) 2021;21. doi:10.3390/S21103461.

[16] Billman GE. Heart rate variability - A historical perspective. Front Physiol 2011;2 NOV:86. doi:10.3389/FPHYS.2011.00086/BIBTEX.

[17] Shaffer F, Ginsberg JP. An Overview of Heart Rate Variability Metrics and Norms. Front Public Heal 2017;5:258. doi:10.3389/FPUBH.2017.00258.

[18] Hovsepian K, Al'absi M, Ertin E, Kamarck T, Nakajima M, Kumar S. CStress: Towards a gold standard for continuous stress assessment in the mobile environment. UbiComp 2015 - Proc 2015 ACM Int Jt Conf Pervasive Ubiquitous Comput 2015:493–504. doi:10.1145/2750858.2807526.

[19] Taelman J, Vandeput S, Spaepen A, Van Huffel S. Influence of mental stress on heart rate and heart rate variability. IFMBE Proc 2008;22:1366–9. doi:10.1007/978-3-540-89208-3_324.

[20] Lucini D, Norbiato G, Clerici M, Pagani M. Hemodynamic and Autonomic Adjustments to Real Life Stress Conditions in Humans. Hypertension 2002;39:184–8. doi:10.1161/HY0102.100784.

[21] Tharion E, Neelakantan N, Parthasarathy S. Short-term heart rate variability measures in students during examinations. Natl Med J India 2009;22.

[22] de Vries GJJ, Pauws SC, Biehl M. Insightful stress detection from physiology modalities using Learning Vector Quantization. Neurocomputing 2015;151:873–82. doi:10.1016/J.NEUCOM.2014.10.008.

[23] Laborde S, Mosley E, Mertgen A. Vagal Tank theory: The three Rs of cardiac vagal control functioning - resting, reactivity, and recovery. Front Neurosci 2018;12:458. doi:10.3389/FNINS.2018.00458/BIBTEX.

[24] van Dooren M, de Vries JJGGJ, Janssen JH. Emotional sweating across the body: Comparing 16 different skin conductance measurement locations. Physiol Behav 2012;106:298–304. doi:10.1016/J.PHYSBEH.2012.01.020.

[25] Amin MR, Faghih RT. Tonic and Phasic Decomposition of Skin Conductance Data: A Generalized-Cross-Validation-Based Block Coordinate Descent Approach. Undefined 2019:745–9. doi:10.1109/EMBC.2019.8857074.

[26] Jacobs SC, Friedman R, Parker JD, Tofler GH, Jimenez AH, Muller JE, Benson H, Stone PH. Use of skin conductance changes during mental stress testing as an index of autonomic arousal in cardiovascular research. Am Heart J 1994;128:1170–7. doi:10.1016/0002-8703(94)90748-X.

[27] Wijsman J, Grundlehner B, Liu H, Penders J, Hermens H. Wearable Physiological Sensors Reflect Mental Stress State in Office-Like Situations. Undefined 2013:600–5. doi:10.1109/ACII.2013.105.

[28] Grundlehner B, Brown L, Penders J, Gyselinckx B. The design and analysis of a real-time, continuous arousal monitor. Proc - 2009 6th Int Work Wearable Implant Body Sens Networks, BSN 2009 2009:156–61. doi:10.1109/BSN.2009.21.

[29] Smets E, Casale P, Großekathöfer U, Lamichhane B, De Raedt W, Bogaerts K, Van Diest I, Van Hoof C. Comparison of Machine Learning Techniques for Psychophysiological Stress Detection. Commun Comput Inf Sci 2015;604:13–22. doi:10.1007/978-3-319-32270-4_2.

[30] Setz C, Arnrich B, Schumm J, Marca R La, Tr G, Ehlert U. Discriminating stress from cognitive load using a Wearable EDA Device. Technology 2010;14:410–7.

[31] Storm H. Changes in skin conductance as a tool to monitor nociceptive stimulation and pain. Curr Opin Anaesthesiol 2008;21:796–804. doi:10.1097/ACO.0B013E3283183FE4.

[32] Palanisamy K, Murugappan M, Yaacob S. Descriptive analysis of skin temperature variability of sympathetic nervous system activity in stress. J Phys Ther Sci 2012;24:1341–4. doi:10.1589/JPTS.24.1341.

[33] Karthikeyan P, Murugappan M, Yaacob S. Multiple Physiological Signal-Based Human Stress Identification Using Non-Linear Classifiers. Elektron Ir Elektrotechnika 2013;19:80–5. doi:10.5755/J01.EEE.19.7.2232.

[34] Ahmadi N, Sasangohar F, Nisar T, Danesh V, Larsen E, Sultana I, Bosetti R. Quantifying Occupational Stress in Intensive Care Unit Nurses: An Applied Naturalistic Study of Correlations Among Stress, Heart Rate, Electrodermal Activity, and Skin Temperature. Undefined 2021. doi:10.1177/00187208211040889.

[35] Ahmed MU, Begum S, Funk P, Xiong N, von Scheele B. A multi-module case-based biofeedback system for stress treatment. Artif Intell Med 2011;51:107–15. doi:10.1016/J.ARTMED.2010.09.003.

[36] Carlos L, Vanderlei M, Pastre CM, Hoshi RA, Dias De Carvalho T, Fernandes De Godoy M, Rua Bela MV. Basic notions of heart rate variability and its clinical applicability. Rev Bras Cir Cardiovasc 2009;24:205–17.

[37] Baker LB. Physiology of sweat gland function: The roles of sweating and sweat composition in human health. Temp Multidiscip Biomed J 2019;6:211. doi:10.1080/23328940.2019.1632145.

[38] Singh RR, Conjeti S, Banerjee R. A comparative evaluation of neural network classifiers for stress level analysis of automotive drivers using physiological signals. Biomed Signal Process Control 2013;8:740–54. doi:10.1016/J.BSPC.2013.06.014.

[39] Posada-Quintero HF, Chon KH. Innovations in Electrodermal Activity Data Collection and Signal Processing: A Systematic Review. Sensors (Basel) 2020;20. doi:10.3390/S20020479.

[40] Amin MR, Faghih RT. Identification of Sympathetic Nervous System Activation from Skin Conductance: A Sparse Decomposition Approach with Physiological Priors. IEEE Trans Biomed Eng 2021;68:1726–36. doi:10.1109/TBME.2020.3034632.

[41] Ralevski E, Petrakis I, Altemus M. Heart rate variability in alcohol use: A review. Pharmacol Biochem Behav 2019;176:83–92. doi:10.1016/j.pbb.2018.12.003.

[42] Koenig J, Jarczok MN, Kuhn W, Morsch K, Schäfer A, Hillecke TK, Thayer JF. Impact of Caffeine on Heart Rate Variability: A Systematic Review. Https://HomeLiebertpubCom/Caff 2013;3:22–37. doi:10.1089/JCR.2013.0009.

[43] Lu C-L, Rr WCO, Chen JDZ. Postprandial Changes of Sympathovagal Balance Measured by Heart Rate Variability. Dig Dis Sci 1999;44.

[44] Karthikeyan P, Murugappan M, Yaacob S. A review on stress inducement stimuli for assessing human stress using physiological signals. Proc - 2011 IEEE 7th Int Colloq Signal Process Its Appl CSPA 2011 2011:420–5. doi:10.1109/CSPA.2011.5759914.

[45] Healey JA, Picard RW, Healey JA. Detecting Stress During Real-World Driving Tasks Using Physiological Sensors n.d.

[46] Can YS, Gokay D, Kılıç DR, Ekiz D, Chalabianloo N, Ersoy C. How Laboratory Experiments Can Be Exploited for Monitoring Stress in the Wild: A Bridge Between Laboratory and Daily Life. Sensors 2020, Vol 20, Page 838 2020;20:838. doi:10.3390/S20030838.

[47] Healey JA, Picard RW. Detecting stress during real-world driving tasks using physiological sensors. IEEE Trans Intell Transp Syst 2005;6:156–66. doi:10.1109/TITS.2005.848368.

[48] Can YS, Chalabianloo N, Ekiz D, Fernandez-Alvarez J, Riva G, Ersoy C. Personal Stress-Level Clustering and Decision-Level Smoothing to Enhance the Performance of Ambulatory Stress Detection with Smartwatches. IEEE Access 2020;8:38146–63. doi:10.1109/ACCESS.2020.2975351.