

Detecting stress patterns based on physiological measurements in real-life scenarios using existing wearables

By

Thies Keulen

GRADUATION REPORT

Submitted to

Hanze University of Applied Science Groningen

in partial fulfillment of the requirements

for the degree of

Fulltime Master Sensor System Engineering

2019

ABSTRACT

Detecting stress patterns based
on physiological measurements
in real-life scenarios
using existing wearables

by

Thies Keulen

During the last decades there has been an uprising in stress related health problems. To counter this growing problem, a system is proposed to detect stress events at an individual level. These events can then be used as feedback in efforts to prevent chronic stress symptoms from developing.

To detect stress, test subjects are monitored using the Huawei Watch 2's PPG signal. This signal was used in combination with a peak detection algorithm to determine heart rate and HRV parameters. A group of (lecturing) staff is monitored during a controlled test sequence using STROOP and MAT(H) tests.

The heart rate and RMSSD/SDNN features show most separation power between rest and stress phases. Test subjects did not respond to proposed test sequence as expected. Therefore classification based on gathered datasets could not accurately be determined with sufficient statistical backing.

Afterwards the test subjects were monitored for at least five days. Data gathered from this field test does show possible stress events and stress trend lines. Further studies should look into improvements to test setup, test sequence and test subject selection.

DECLARATION

I hereby certify that this report constitutes my own product, that where the language of others is set forth, quotation marks so indicate, and that appropriate credit is given where I have used the language, ideas, expressions or writings of another.

I declare that the report describes original work that has not previously been presented for the award of any other degree of any institution.

Signed,

Thies Keulen

ACKNOWLEDGEMENTS

I would like to thank Berend Jan van der Zwaag and Johan Blok for their supervision and feedback throughout this project. Furthermore I would like to thank the Hanze University of Applied Sciences for the time and opportunity to work on this project.

For her guidance throughout the initial phase and the search for the project definition I would like to thank Marietta de Rooij. Additionally I would like to thank all (lecturing) staff and colleagues involved in my education for their efforts during the past years.

Last but not least I would like to thank everyone close to me for their support and the opportunity to go through this process.

CONTENTS

DEFINITIONS AND ABBREVIATIONS	6
LIST OF TABLES.....	7
LIST OF FIGURES.....	8
1. RATIONALE	9
2. SITUATIONAL & THEORETICAL ANALYSIS.....	11
2.1. Lab test protocol	11
2.2. Data labelling	12
2.3. Detecting stress.....	13
2.3.1. State of the art	13
2.3.2. Test length	16
2.3.3. Test population size	16
2.4. Ethics	17
2.5. Data handling	18
3. CONCEPTUAL MODEL.....	19
4. RESEARCH DESIGN	20
5. TEST SETUP.....	22
6. RESULTS.....	26
6.1. PPG signal.....	27
6.2. Rerun.....	29
6.3. Stress detection	30
6.4. Heart rate.....	34
6.5. Field test.....	35
7. DISCUSSION.....	36
8. CONCLUSION.....	37
9. RECOMMENDATIONS.....	39
10. REFERENCES	41
Appendix A : Test subjects information brief	45
Appendix B : Informed consent form.....	46

DEFINITIONS AND ABBREVIATIONS

PWV	Pulse wave velocity
PPG	Photoplethysmography
GSR	Galvanic skin response
HRV	Heart rate variability
GDPR	General Data Protection Regulation
ERB	Ethical review board
TSST	Trier Social Stress Test
CPT	Cold pressor test
MAT	Mental arithmetic task
VAS	Visual analogue scale
FFT	Fast Fourier Transform
Stroop	A test focused on the conflict between dominant and non-dominant reactions. For example reading a color word on a card which is written in a different color [28].
SSL/TLS	Secure Sockets Layer / Transport Layer Security (updated version of SSL)
Hanze UAS	Hanze University of Applied Sciences

LIST OF TABLES

Table 1 Peak detection algorithm comparison	24
Table 2 Peak detection comparison advanced test	24
Table 3 Percentage of useable data from first lab tests	26
Table 4 Rerun heart rate detection rate	29
Table 5 Wilcoxon results test subject #5	30
Table 6 F1-scores of the classifiers	32

LIST OF FIGURES

Figure 1 Number of absent days due to work stress per absent employee (left) and costs due to work stress per absent employee in euro (right) in the Netherlands[1]	9
Figure 2 ECG of a heart in normal sinus rhythm [41].....	14
Figure 3 Lab test timing diagram	20
Figure 4 Physical activity skipping.....	23
Figure 5 Slavic peak detection on advanced test signal	25
Figure 6 Window slicing in lab dataset	26
Figure 7 PPG signal degradation	27
Figure 8 Garbage collection events (left) and improved memory management (right)	28
Figure 9 Accepted slices test subject #5 rerun	29
Figure 10 Wilcoxon p-value results test subject #5 rerun (75% threshold).....	30
Figure 11 Correlation matrices test subject #5.....	31
Figure 12 Separation between rest and stress data	32
Figure 13 VAS score labelling by test subjects.....	33
Figure 14 Heart rate during lab test of all test subjects	34
Figure 15 Stress before start of stress phase.....	34
Figure 16 Heart rate during field test with significant event.....	35
Figure 17 Heart rate field test with linear increase trend	35
Figure 18 Effect of reducing sampling rate on the PPG signal.....	39

1. RATIONALE

Over the last decades there has been a shift from mainly physical labour to work more focused on knowledge and service. With this change, the physiological aspect of employment became more important [6]. The role of the employee changed from performing routine tasks to being a multi-skilled team player with the skill of mental adaptation [2][7]. This role change initiated the uprising in stress related health problems, see Figure 1.

Work stress can be defined as the harmful physical and emotional responses that may occur when there is an imbalance between job demands and the amount of control an employee has on meeting these demands [12]. Of all the health related absenteeism in 2017 in the Netherlands, 16% was due to work-stress [1]. This alone resulted in a cost of two billion euro to Dutch employers [1].

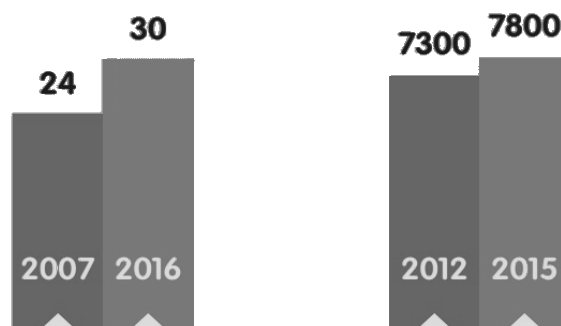


Figure 1 Number of absent days due to work stress per absent employee (left) and costs due to work stress per absent employee in euro (right) in the Netherlands[1]

The cause of stress related absenteeism is varied, with e.g. 38% due to high work demand and 45% due to experiencing low autonomy in executing these tasks [1]. These figures state the quantifiable amount of absent days, but they do not include the loss in productivity due to rising stress levels nor the costs involved in e.g. reintegration tracks [2].

To prevent an individual from developing chronic stress symptoms such as a burn-out, the aim is to develop a system that records all stress events. This system is used to indicate a rise in stress level at a stage in which the individual is often unable to yet indicate a change. With this information action can be taken to decrease the possibility of developing chronic stress and the health risks that come from it. The purpose of this research is to find a relationship between physiological parameters and stressful events. This relationship could then be used to detect stress events and prevent chronic stress symptoms from developing.

To detect stress in a real-world scenario, the most accurate readings can be made when the impact on the user is as little as possible [3]. This is because the awareness of being monitored can influence the users' physiological parameters (white coat effect). Recent studies with real-world measurements are with sensor systems that are not wearable in a normal working day environment [24][25]. This questions their suitability to everyday use.

This project will limit itself to the use of wearables that are readily available, affordable for the end user and non-limiting in the daily activities. Therefore this project is limited to using existing smart watches. Watches are commonly worn accessories and eliminate the white coat effect as much as possible from the stress event detection.

The following research question is deduced from the described problem, the desired situation and the scope:

“How can stress patterns be detected by measuring physiological variables using existing wearables in real-life situations to indicate a rise in the amount of stress events?”

To answer the main research question the following sub-questions are defined:

- What physiological variables are linked to stress?
- What method is most suitable for detecting stress events in real-world scenarios?
- What existing wearables meet the set requirements?
- What standardized test procedures allow for labelled datasets?
- How can the measured variables be labelled in relation to stress?
- How can the measured variables be combined to detect a pattern in stress events before a user can indicate it?
- How can the selected wearable be used to monitor a group of lecturers and other staff over a period of several weeks?

2. SITUATIONAL & THEORETICAL ANALYSIS

There has been a division between the engineering and physiological approach to measuring stress events. Both of these have been criticized, for their lack of incorporating environmental factors, non-specificity and lack of attention to individual differences [4].

Most research done into stress detection has been in controlled environments and with known stressors. An upcoming line of research is in measuring stress in real-world scenarios. The following characteristics separate the two approaches:

- There has been no conclusive evidence that stressors applied in a controlled environment give the same reactions as to stress in real-world scenarios [5], which is eventually where the system is used in.
- The validation of real-world stress events has proven to be challenging, due to user recall bias and overall lack of sensibility.
- The tests performed under lab conditions are suitable for labelling as being e.g. stress/not stressed with standardized and accepted test procedures.

With these limitations, a combination of these two approaches will be investigated.

2.1. Lab test protocol

For the labelling of gathered data (e.g. stressed / not stressed) a test protocol needs to be used that is verified to induce stress in a reliable way. Several of these test protocols have been developed and tested over the years [26][27].

Stroop test

The Stroop test [28] is a test focused on the conflict between dominant and non-dominant reactions. An example would be a test where the participant needs to read a color word on a card which is written in a different color.

TSST (Trier Social Stress Test)

The TSST test [29] is more focused on inducing psychological stress by being confronted by job interviews, public speaking and possible arithmetic's tasks all in front of a test panel. The downside of this test is that only one participant can be tested at a time and the relatively large amount of test panel participants needed [27].

CPT (Cold Pressor Test)

In a CPT test the participant is expected to experience stress due to induced pain. An example would be holding a hand into ice-cold water for as long as the participant can endure. Stress response is limited to the duration of the stressor [27].

Ergometer test

A test in which the participant is asked to cycle for a set period of time. The physical nature of the ergometer test has reliably induced stress responses [30]. One could argue the relation to non-physical stress response.

Mental arithmetic task (MAT)

Test consisting of only arithmetic tasks. These tasks are available for a limited time and immediate feedback is presented to the participant. In contrast to many other test methods, parallel testing is possible with MAT [27].

No-stress control test

To indicate the difference between stress and no-stress, a baseline needs to be established. A no-stress control test consists of relaxation or a milder version of the MAT test.

From these tests the TSST and STROOP tests show the most response in stress levels [24][26]. The lowest reaction is to be expected of the CPT test and the ergometer test also has a high response but is mainly focused on physical activity [26][27]. The MAT(H) test has mixed results, better performance can be expected when feedback is given by a human assessor as to a peer score [27][32].

To determine a multilevel stress pattern a combination of tests and rest periods will be used in the test protocol.

2.2. Data labelling

For a system to detect stress, it needs to be trained with labelled data. Looking at the laboratory tests the following labelling can be applied:

Yes/no stress label

In this method data in the rest phase of the protocol is being labeled as no stress and the period with applied stressor as stress. This method does not take into account individual differences in stress response time or overlap in events.

Multilevel Visual analogue scale (VAS) score

The VAS test assesses the current perceived stress level of the individual using a questionnaire [26]. This method relies on the participant to do the labelling and is therefore subjective to recall bias and inaccuracy [31]. On the other hand this method does allow for multilevel labelling to indicate a rise in stress level.

Unsupervised labelling

Determining clusters in features based on unsupervised learning methods. This would suit best in future application in real-world [32]. This method is also suitable for multilevel labelling, with the introduction of additional clusters. This learning method can be applied to both individual data and to the data of the entire group of participants. To determine clusters, the measured features would have to be scaled. This to be sure all features have equal importance in the clustering process. Also the system would have to compensate for the differences in sampling frequency.

For the verification of an unsupervised method, human labelling is still needed.

2.3. Detecting stress

Research focused on detecting stress can be categorized into three groups [4][8], all with their limitations:

- Questionnaires (time consuming, so far unreliable and recall biased)
- Invasive / intrusive (blood , saliva, urine and hormones , measurements are hindering the user)
- Non-invasive (heart rate, blood pressure, respiration and others).

From these three groups, a non-invasive approach seems to be the best for measuring stress as it does not rely on one's recollection and is not hindering during daily activities.

There has not been a physiological parameter found that can uniquely identify stress. The link from mental stress to physiological variables is subjective to numerous factors and can differ per individual. These factors act in the translation between the signals in the brain being translated into neuronal and hormonal components that in their turn trigger the physiological variables [4]. Several studies do confirm that the following physiological parameters have a clear correlation with mental stress [4]:

- Blood pressure
- Heart rate / HRV
- Skin conductance level
- Respiration rate
- ECG/ EMG.

2.3.1. State of the art

These five parameters will be investigated further to assess the suitability in relation to the requirements for this project. Per parameter the available techniques will be discussed and which other factors can also influence this parameter. Lastly the common factors influencing all five of these parameters will be discussed.

Blood pressure:

Normally blood pressure is measured using an inflatable cuff to temporarily cut off blood circulation. This technique is however unsuitable for continuous use and is hindering the user. Companies tend to stay away from adding blood pressure measurements to their wearable due to lacking accuracy and following regulatory approval [4]. Omron Healthcare has developed such a wearable but it is still undergoing clinical tests and pending FDA approval. The technique used is still an inflatable watch strap [5], therefore the hinder to the user is still to be seen. During the last decade research has been done into other techniques to detect blood pressure (e.g. using pulse wave velocity with photoplethysmography). These techniques look promising but are yet to be integrated into consumer available wearable devices [6][7][8]. Devices with these capabilities are in development and companies are racing for patents and FDA approval [9][10][11].

Other unique influencing factors besides stress:

State of the vascular system, kidney diseases, hormone disorders, sleep apnoea, and eating [23].

Heart rate / HRV:

Being one of the easier parameters to measure, heart rate measurement functionality has been available in wearables for years (e.g. Polar, Fitbit, and Huawei). Using photoplethysmography (PPG), the arrival of the artery pulse wave can be detected. The measurement does however suffer from user movement, as the measured difference is normally only around 2% [12]. This normal difference can easily be lost in the much larger impact of movement on the measured reflected light [12].

From the PPG signal, the heart rate variability (HRV) can also be determined. The HRV is defined as the fluctuation in the time between consecutive heart beats. Traditionally HRV is derived from an ECG signal [33].

Other influencing factors:

ambient temperature and body position [22].

An ECG signal contains peaks, segments and intervals. Any deviation in these features is of clinical significance. To determine the HRV, the time difference between R wave peaks in the ECG signal is measured (see Figure 2). The interval is called the R-to-R interval or RR. Next to the RR several other characteristics can be determined, in both time and frequency domain.

Previous studies frequently use the following characteristics to indicate stress[34][35][32], with NN being 'all normal RR intervals' [32]:

Time domain:

- Mean HR
- NN intervals
- NN50 (amount of consecutive NN intervals that differ by more than 50 ms).
- PNN50 (percentage of NN50 intervals to total amount of NN intervals).
- RMSSD
(Root mean square of successive differences between adjacent NN intervals). Value that represents the autonomic control of the heart [36]. See equation 1.
- SDNN
(Standard deviation of all NN intervals). Considered the "gold standard" for determining cardiac risk when recorded over 24 hours [37]. See equation 2.

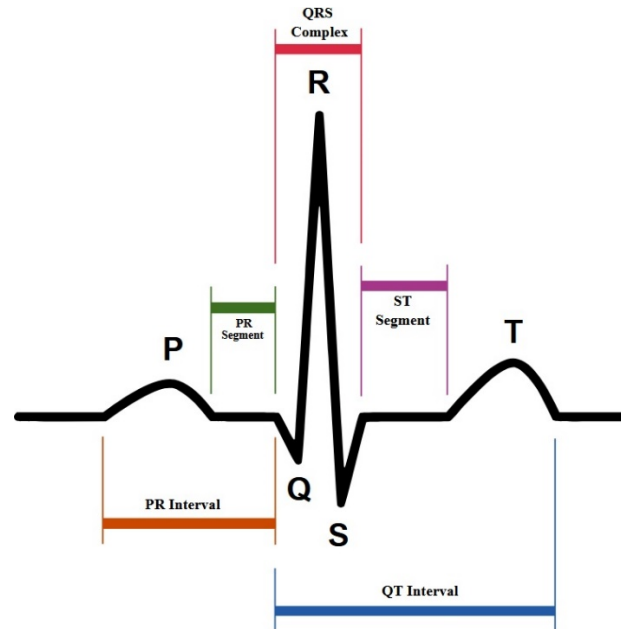


Figure 2 ECG of a heart in normal sinus rhythm [41]

$$RMSSD = \sqrt{\frac{\sum_{i=1}^N (RR_{i+1} - RR_i)^2}{N-1}} \quad (1)$$

$$SDNN = \sqrt{\frac{\sum_{i=1}^N (RR_i - RR_{mean})^2}{N-1}} \quad (2)$$

with N being the number of RR intervals

Frequency domain:

- LF HRV (Low frequency power (0.04–0.15 Hz))
- HF HRV (High frequency power (0.15–0.4 Hz))
- Ratio LF-HF HRV.

There is a trade-off in performance between characteristic calculations in time or frequency domain. With faster frequency domain calculations but with the drawback of having to do a Fast Fourier Transform (FFT) to get to the frequency domain and back. The limiting factors in this trade-off are the platform processing speed, storage size and transfer speed.

Although HRV is not a feature presented by current on the market wearables, the Huawei Watch 2 does allow for the extraction of raw PPG data. From this raw data, peak detection algorithms can be used to determine the before mentioned HRV characteristics [38]. Specific algorithm performance will be investigated for the used PPG data.

Skin conductance:

Galvanic skin response sensors (GSR) are used to detect changes in electrical properties of the skin. The skin conductivity is traditionally measured between two electrodes attached to two fingers of one hand [13][14][15]. The conductivity is increased by additional moisture produced by sweat glands. Consumer available wearables have slowly starting to introduce this functionality into smart watches/bracelets. These products are however still in development [16], specifically only for this parameter [17] or in a price range that makes it only available for research [18].

Other influencing factors:

ambient temperature [20].

Respiration rate:

Within the sport and fitness communities the measuring of respiration rate is well embedded. These measurements are usually done using a chest strap (e.g. Polar). Wearing a chest strap for an extended period (full days) will be hindering. Realizing this measurement in a wrist worn wearable has been a challenge. Consumer available wrist worn wearables with this functionality have yet to be developed. Apple has filled for a patent describing measuring respiration rate using a PPG array [19].

Other influencing factors:

- Blood oxygen level, blood pH, conditions that damage the rhythmicity center or level of consciousness [21] and speaking [24].

EMG / ECG:

Recent studies looked into the change in muscle activity due to stress. The use of electromyography (EMG) or electrocardiography (ECG) features appear to have promising results [24]. This technique requires the test subject to wear leads and/or electrodes to measure activity across muscles. Wearable devices featuring ECG require the user to press their finger on the smart watch to take a measurement (e.g. Apple smart watch series 4 and Cronovo smart watch).

Common influencing factors:

There are several factors that influence all of these parameters besides stress, being:

- Physical activity, fitness level and body mass index
- Age, gender, race and genetics
- Certain illnesses and medication
- Smoking, drugs and alcohol.

Most of these influencing parameters cannot be measured using wearables. With the exception of physical activity. That parameter can be measured using e.g. an accelerometer. Of the investigated devices, most include an accelerometer due to its relatively easy implementation and its use for user input and gesture/game control. Other influencing factors are eliminated during test subject selection where possible or accepted as part of the application to a large population.

2.3.2. Test length

In research found in literature the length of tests vary widely. In laboratory setting ranging from 2 to 10 minutes of stress, resting times of 2 to 10+ minutes and recovery monitoring of 2 to 75+ minutes [26][27][32]. Further study would be advised to find an optimum in these test protocol times. For now the test procedure will have to find the right balance between test length/variation and participant availability.

Looking at the real-life environment, a longer time span will be necessary to encounter significant stress events. Research found in literature again has widely varying time frames: 24 hours [34], 5 days [39] and 55 days [40]. Time frames seem mainly driven by project scope/funding or application rather than scientific proof. The target participants are staff/lecturers at a University of Applied Sciences. In this particular case the educational year is organized in quarters of ten weeks. To narrow down to a time span some assumptions are made. For instance one could say the same stress cycle occurs in every quarter (deadlines/grading vs lecturing and preparation stages). The exception would be the first and final quarter (first year start-up and graduation stress). Ideally the main stress events are covered in the test. On the other hand the test length needs to keep in mind the organisational/project limits and test subject impact.

2.3.3. Test population size

The validity of the developed algorithm depends on the size of the test population. Previous research does not describe a fixed number of minimum participants. Due to limitations in hardware availability two groups of 5 test subjects will run the proposed tests, one after the other. The project results determine if this population size allows for enough variation and detected stress events.

2.4. Ethics

With the application of measuring a users' physiological parameters the ethical implications need to be considered. The tests need to comply with the ethical guidelines in the Netherlands [58][59] and the recently introduced privacy law in the European Union (General Data Protection Regulation) [60]. An institution that can help with this is the "Ethical Review Board" from the Hanze University of Applied Sciences (Hanze UAS). This institution can provide two levels of services: advice or a formal letter of approval. An advice on the chosen test setup would in this project suffice (the actual advice would indicate further actions).

All participants need to sign an informed consent form, indicating they understand the tests, the application and their rights. In collaboration with the ethical review board an information brief and informed consent form were constructed (Appendix A and B). These forms were used during the tests.

When the proposed system comes to a finalized product, the following ethical issues need to be addressed. This system would lead to information about stress levels and events of the user, which could be miss-used. Most critically, the information could be used by the employer against the employee. An employer might be tempted to downplay sick leave when the cause is known to be stress. The knowledge of frequent stress events might allow the employer to change an employee's tasks in a way that does not benefit the employee. Moreover in the event that this information leaks to other employers it might be used in the selection process. An employee with known stress problems might be less tempting to take on board. In the Netherlands the inquiry into the state of health or the history of health is not allowed during the selection process. However this might not be the case in other countries.

Next to the possible problems that occur when this information is shared or leaked, the stress information could affect the user too. Having the knowledge of raised stress levels or events could lead to a larger reaction than needed by the user. For instance a user might consider leaving the company or slide into a depression. The knowledge of stress levels could also raise the stress levels themselves.

This project does not look into the actual feedback to the user nor the storage of this information when used as a finalized product. The project limits itself to seeking feasibility, future work will have to look into these concerns.

2.5. Data handling

In combination with the before mentioned ethical aspects of this project there are a number of requirements related to data handling. First and foremost the data needs to be kept private. Another key requirement is that the data needs to be stored for analysis later (e.g. unsupervised labelling/building a stress detection algorithm). The amount of data is key in the data handling decision. Looking into the data management side of the proposed system, there are a number of options.

The most secure option would be to store the data on the watch itself. However the amount of data storage on a wearable is limited (e.g. 2350 MiB on a Huawei Watch 2). The most commonly used approach is to send this data via Bluetooth to a smartphone with a large storage size. The introduction of a smartphone does however make the setup more complex for the user (separated apps/start up/charging) but also for development with making the app compatible with all user smartphone platforms (e.g. Android/Windows/iPhone). The introduction of a separate supplied smartphone would force the user to think about being measured and will introduce the before mentioned white coat effect.

Another option would be to directly send the measured data from the watch to a central storage server via Wi-Fi. This option would keep user impact to a minimum, with only having to wear the wearable and all data traffic and storage being taken care of by the system. This would keep the white coat effect to a minimum and test deployment over a larger test population easier. Sending the data to a central server would mean more work for the backend of the system.

Sending raw data to a server or smartphone like PPG data would be stressing the Bluetooth or Wi-Fi communication. This could force raw data to be analyzed into features on the wearable before being sent to storage. This requirement does limit the type of characteristics that can be measured due to the limited processing speed of wearables.

For increased security, data transmitted to a server should only be sent over a private network. An option would be to encrypt the data using e.g. SSL/TLS. Ultimately a trusted 3th party needs to supply a certificate for this communication.

3. CONCEPTUAL MODEL

The aim of this project is to detect stress events in a real-world environment using existing wearables. Based on the situational and theoretical analysis of previous chapters, the following model is derived.

A combination of controlled and real-life environments will be investigated in order to approach a verifiable and applicable solution. Research indicates that there are multiple physiological parameters linked to stress, with five being cited most. To accurately measure a person's physiological parameters, the user has to be hindered and aware as minimal as possible (minimize white coat effect). To accomplish this, non-invasive measurement methods will be used that can be worn during day-to-day activities. These five most cited physiological parameters are still difficult to measure using existing wearables. There are however several products in development. Heart rate / HRV is the only parameter that can be detected using existing wearables and is linked to stress.

Using the PPG sensor of the wearable the following heart rate characteristics will be recorded: mean HR, NN interval, NN50, PNN50, RMSSD, SDNN, LF HRV, HF HRV and the ratio LF-HF HRV. These features are not readily available and have to be derived from the PPG signal using a peak detection algorithm (with the exception of HR). The most suitable peak detection algorithm needs to be found. There are well known algorithms available for testing on the signal specific to the used device.

The mentioned parameters are not only influenced by stress but have a variety of influencing factors. Most of these cannot be measured using wearables, with the exception of physical activity. The addition of an accelerometer to detect motion would eliminate a part of the uninteresting events. Other influencing factors such as diseases and medication will be part of the test subject selection.

The selected wearable needs to be able to measure the before mentioned parameters, be readily available to the general public and non-hindering to the user. With these wearable requirements in mind, the Huawei Watch 2 is chosen as the existing wearable in this project. This wearable can as many other wearables measure both heart rate and physical activity but also has the option to access raw PPG data, which was not found in any other wearable.

Laboratory test protocols are varied in their use and their stress response. Studies on test effectiveness have mixed results in best method. However both TSST and Stroop induce significant stress, with the TSST being more complex in setup. The MAT(H) test can under the right conditions also apply significant stress. For identifying multilevel stress patterns a combination of laboratory tests should be used. The combination of Stroop and MAT(H) tests will be used in the laboratory test.

Test length is varying widely among research found in literature and is mainly driven by application/scope rather than scientific proof. Further research in this would be valuable but is not in the scope of this project. The laboratory test is limiting itself to 30 mins and the real-life test to 3 weeks. These tests will be held in the ending of quarter 2 and 3 of the academic year 2018/2019.

For the labelling of data a combination of available methods will be used. The unsupervised method to obtain bias free labels, yes/no labelling for verification of the unsupervised method and VAS method to investigate the benefit of multilevel labelling. For unsupervised learning the data will be pre-processed.

Taking the need of a smartphone out of the system would significantly decrease user test stress. However, sending data from a wearable to a server might require data to be pre-processed into features on the wearable. Sending data to a server also increases the possibility of sensitive data being leaked.

Subject tests need to comply with Dutch and EU regulations regarding ethics and privacy. To increase privacy, data will only be send over a secure connection on a private network in an anonymized way. The ethical review board of the Hanze UAS will be asked for ethical advice. Before testing begins, all participants need to sign an informed consent form.

4. RESEARCH DESIGN

Gather data

The required data will have to be obtained from the Huawei Watch 2 using an Android Wear app that needs to be developed. The HR is a standard feature that can be extracted from the smart watch but the other features need to be derived from the PPG signal. This PPG signal is available in the Android Wear app. On this signal, peak detection needs be performed. From these peaks the remaining features can be calculated. The features in the frequency domain would require additional computation time and research needs to be done if this is possible on the watch within the time constraints.

The data would then need to be transmitted to a storage server, while maintaining the test subjects' privacy. On this server a HTTP(S) server application needs to be developed to receive and store the data in a standard database. The connection from the wearables to the server needs to be secured to prevent the interception of sensitive information.

Lab test

Before the lab test the test procedure needs to be send to the ethical review board of the Hanze UAS for advice. Based on this advice the procedure might need adjustment. Based on the response and availability of the potential test subjects, a selection needs to be made based on the test criteria. The aim is to select participants over a large spectrum of age, gender and function within the Hanze UAS.

The lab test itself consist of two standardized test that need to be prepared:

1. Stroop test : https://www.pytoolkit.org/experiment-library/experiment_stroop.html
2. MAT(H) test : E.g. subtract 17 from 1223, start over on mistake. Encourage by use of a leader board.

These two standardized tests are combined with resting periods, see Figure 3.

The first group of five test subjects will start with the lab test and after data processing continue with the real-world test. The VAS scoring card and the yes/no labelling needs to be done during the lab test. After the lab test the unsupervised method needs to be run.



Figure 3 Lab test timing diagram

Real-world test

The real-world test would require more setup time to ensure the gathering of data over a longer period of time. During the test, maintenance and supervision is needed to further ensure this process.

Before the start of the second group the data of the first group needs to be analyzed and possible algorithms to detect stress events need to be developed.

Second test

The second group of tests acts as a verification of the first group and to increase statistical rigor.

Hypothesis

Based on the chosen parameters, the expectation is that stress events can be detected during lab and real-world tests. The heart rate and raw PPG signal will provide sufficient information to determine the distinction between stress and non-stress events. This while the addition of motion detection will sufficiently eliminate unwanted stress event labels. The set timeframe of the tests will cover a sufficient amount of stress events to be detected.

5. TEST SETUP

To measure the chosen parameters the Huawei Watch 2 is used. An Android Wear app was developed using Android Studio 3.3 [42] to facilitate these measurements. This application needs to sample the heart rate, PPG data and store it for further analysis. This application is deployed via Wi-Fi to the Huawei watches using the Android Debug Bridge protocol (ADB) [43].

Deployment

Unfortunately the watches do not natively support the WPA2 Enterprise protocol used to connect to Wi-Fi networks such as the “eduroam” network used by the Hanze UAS. In order to connect to the Wi-Fi network a separate application was installed (Wi-Fi manager for Wear OS) [44].

To deploy applications to the watches using the ADB protocol, the user needs to allow ADB debugging on the watch. This feature is accessible via the hidden “Developer options” on the watch [45].

Sensors

For the measurement of heart rate and PPG values, two sensors are accessed within the Android application. The heart rate sensor is an Android Wear native sensor type, the PPG sensor is a Huawei Watch 2 specific custom sensor type (65537). Documentation on this sensor type is unfortunately lacking, but by looking at the output (float array) only one data field shows change during use (3th entry). This change indicates the change in reflected light due to the fluctuation of blood flow in the wrist. With this data, pulse arrival time can be determined using a peak detection algorithm.

Data logging

During development of the Android Wear application the options for data logging were investigated. These options were, logging to:

- Research laptop using ADB
- A dedicated server using HTTPS
- The watch internal storage.

There is a trade-off between the logging options. The logging to a laptop was considered impractical due to laptop availability during the field tests. Logging to a dedicated server looked promising, because of its large storage and up time. However the connection to the eduroam network proved to be unstable, possibly due to the non-native implementation of the WPA2 Enterprise protocol. The mobile nature of the test subjects during field tests complicates this solution too. Temporary storage would still be needed during transition between network coverage areas or general connection loss. For continuation purposes of the tests, a solution was developed to store the measured data on the watch itself. This data can then be exported to researcher’s laptop after the test by use of ADB. Calculations showed a maximum storage use of 500 Mbyte per watch during field tests, with the Huawei Watch 2 having 2350 Mbyte available.

Movement

Due to the influence of physical activity to the heart rate, it was chosen to only measure and store the chosen parameters during periods of relative immobility. This first filter allows for a resulting dataset that is already cleared of unwanted data. To construct this initial filter the available features of the Wear OS were investigated.

The Android Wear OS allows for the specification of estimated accuracy per sensor. The Huawei Watch 2 updates this accuracy field in the Wear OS environment. With this information the only data that is collected is from periods where the heart rate estimation from the watch is at “High” accuracy. This can be seen in Figure 4, where the diagonal line connects two points before and after significant movement. This movement made the watch change its estimated heart rate accuracy to below “High”, which resulted in the program not recording the data during this period.

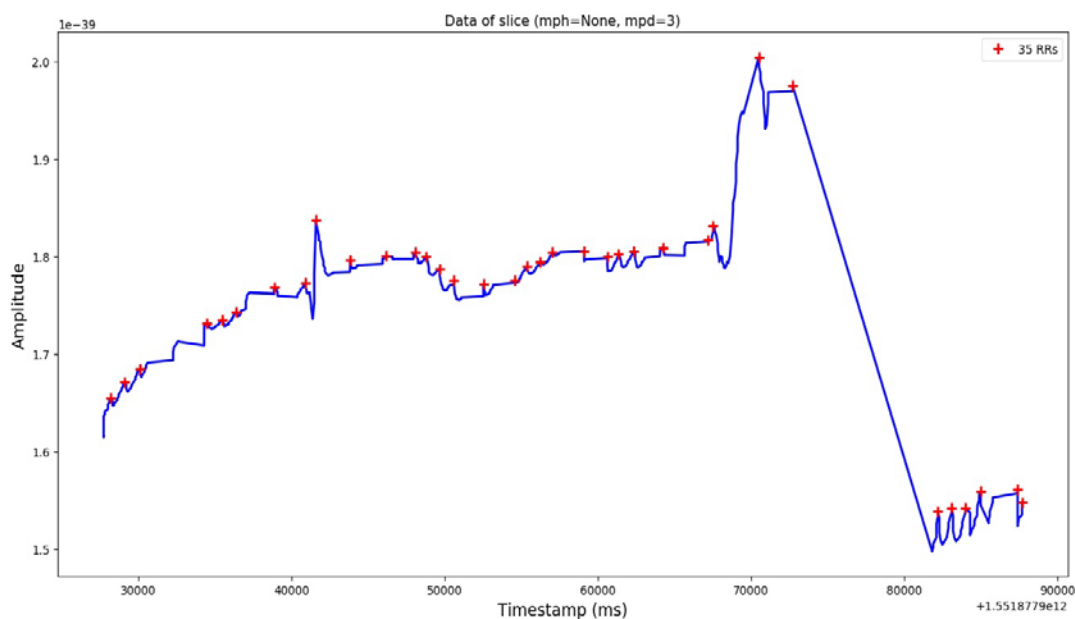


Figure 4 Physical activity skipping

Permissions

In order to make use of the above mentioned features in the Android Wear application a number of permissions have to be granted by the user. The following labels need to be set in order to grant these permissions:

- android.permission.INTERNET (used to debug using ADB)
- android.permission.BODY_SENSORS (allow the application to access heart rate and PPG sensor data)
- android.permission.WRITE_EXTERNAL_STORAGE
(used to write sensor data to storage space accessible by ADB).

Before all tests were executed, these settings were accepted on behalf of the test subjects.

Peak detection

Unfortunately, the Huawei Watch 2 does not supply peak detection data on the PPG signal. Therefore a comparison between existing peak detection algorithms was done. Most peak detection algorithms allow for minimum amplitude difference and/or minimum distance between peaks filtering. To detect peaks in the PPG data, only the minimum distance between peaks could be used as a filter. This limitation is due to each individual having their own amplitude differences. This minimum distance was calculated using a maximum heart rate of 200 and the sampling frequency. At the maximum heart rate the peaks are closest to one another and by use of the sampling frequency the amount of samples between this minimal peak distance was calculated.

To compare the performance of the peak detection algorithms during optimal conditions, a stationary situation test PPG signal was recorded. During the test 311 peaks were recorded. See Table 1 for the number of peaks detected by each algorithm. These algorithms were selected for their minimum distance filtering capability [46].

Algorithm	Peaks detected (311 actual peaks)
Scipy (find_peaks_cwt) [47]	2074
Scipy (argrelmax) [48]	311
Scipy (find_peaks) [49]	578
PeakUtils (peakutils.peak.indexes) [50]	551
Sixtenbe [51]	311
Slavic [52]	312 (also detected starting point)

Table 1 Peak detection algorithm comparison

This first test indicated a perfect result for three algorithms (Argrelmax, Sixtenbe and Slavic).

To determine the most suitable algorithm for this application a second test signal was recorded. This test signal was recorded by mimicking motion found during lecturing tasks (pointing/walking/writing on a board). Table 2 shows the performance differences of these three algorithms.

Algorithm	False positives	False negatives
Argrelmax	4	1
Sixtenbe	5	7
Slavic	2	5

Table 2 Peak detection comparison advanced test

Although the ArgreImax algorithm has the least amount of total false detections, the Slavic algorithm was selected for its least amount of false positives. The effect of false positives was considered to be most important, as this would introduce more false peak-to-peak intervals. The introduction of more false intervals would lead to more false data points being used in the algorithm to detect stress events. False negatives on the other hand would eliminate possibly good data points. In this case the correctness of the data outweighed the amount of data points. Another beneficial side of the Slavic algorithm is the relative easy coding (20 lines compared to 400+ lines or the use of the SciPy library). See Figure 5 for the peak detection of the Slavic algorithm on the advanced test signal, with both green and red markers indicating detected peaks.

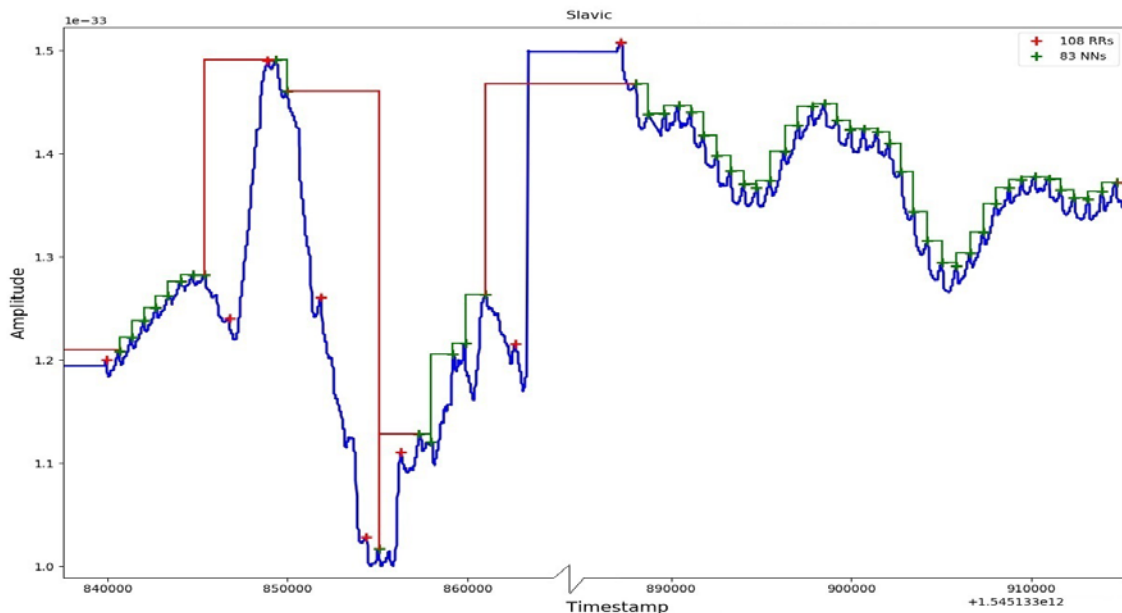


Figure 5 Slavic peak detection on advanced test signal

Normal peak intervals

Using the minimum and maximum time between peaks, the found RR peaks from the Slavic algorithm are filtered into so called 'normal' peaks (NN). NN peaks are a subset of the RR peaks, being those peaks that are detected within a set maximum time difference to a preceding peak. In Figure 5 it can be seen that the first peak detected in a series of peaks is labelled unusable (non-NN) due to its long time difference to a preceding peak.

From these NN peaks, NN intervals are determined. NN intervals are those intervals between NN peaks that are within a set time range. This time range is determined based on the maximum and minimum heart rate. For instance an interval that is larger than an interval that would occur at the minimum heart rate is rejected. In Figure 5, NN intervals are indicated in green and non-NN intervals are indicated in red.

6. RESULTS

After the lab tests were conducted the amount of useable signal was determined. This was done by first splitting the data by test phase (no-stress, stress, no-stress, etc.). These phases were then split up into windows of one minute, see Figure 6. The window size of one minute is consider the smallest time to determine HRV characteristics [53][54][55].

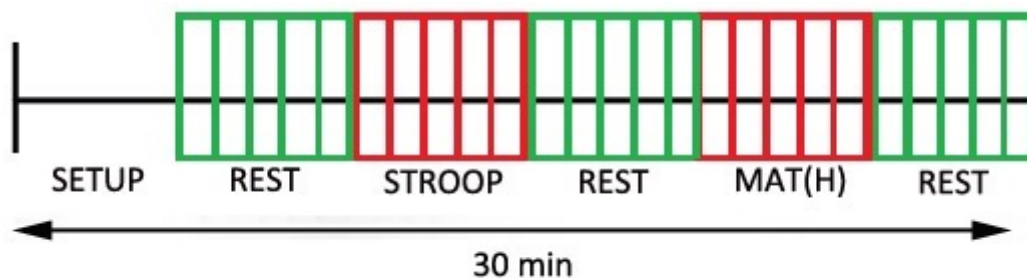


Figure 6 Window slicing in lab dataset

In order to determine if the data in a window is useable for feature extraction the number of detected NN intervals were compared to the heart rate. This window acceptance filter was introduced to make sure only the data was used when the PPG signal was of good enough quality. Without this filter, HRV features would be calculated for a window even if only a few peaks were detected during that window. These possible features would not be a good representation of the condition of the test subject. For a window to be considered 'ok', a specified percentage of NN intervals to heart rate need to be present. E.g. a window with a reported average heart rate of 80 needed 60 NN intervals to be considered 'ok' if the threshold was set to 75%. When a window did not meet this requirements the window was shifted by one peak distance and the cycle started again. Table 3 shows the amount of useable data after the first lab tests.

Test subject #	Minimum % of heart rate detected as NN intervals in window		
	25%	50%	75%
1	34.53%	12.40%	0%
2	11.54%	2.30%	0%
3	Not useable		
4	Not useable		
5	51.97%	21.71%	0%

Table 3 Percentage of useable data from first lab tests

These results are of limited use to determine correlation to stress. Therefore the data gathered during the lab tests was examined.

6.1. PPG signal

After the lab tests the PPG signal was analysed. During this process it became clear that there were large unusable parts within the signal. See Figure 7 for an example signal. This signal starts with a clear heart beat pattern and continues to be inconclusive, with a significant loss in sampling rate.

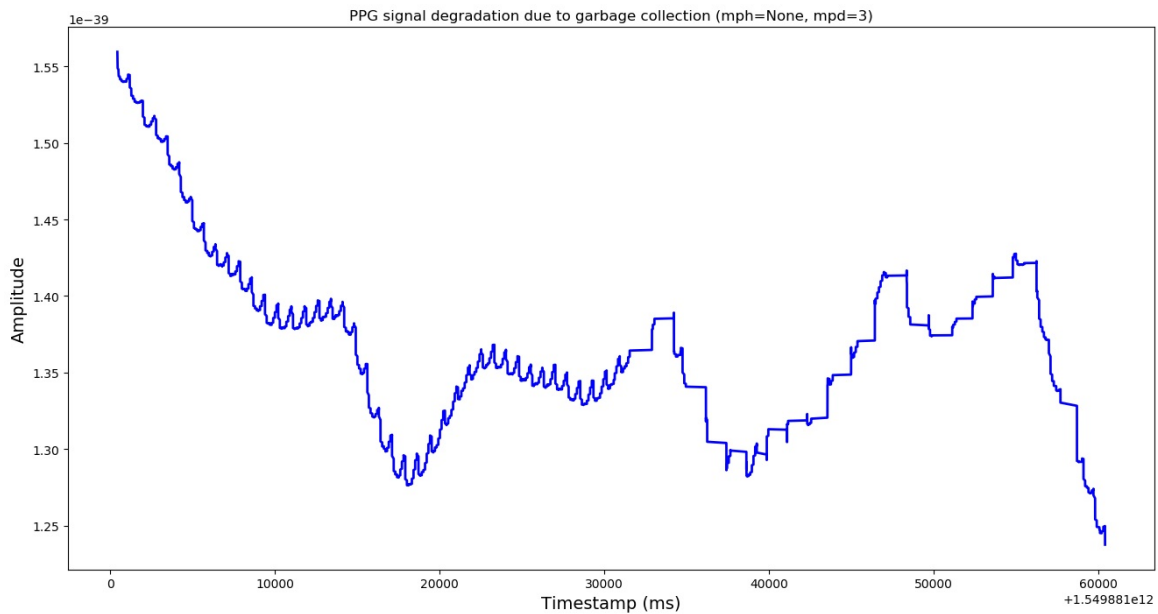


Figure 7 PPG signal degradation

This change in signal density resulted in a drop in useable NN intervals. Investigation of the problem resulted in two error sources, being:

- The garbage collection of the Android Wear app
- Application CPU priority level lowering when entering ambient mode after prolonged lack of interaction.

These problems combined resulted in frequent sampling frequency drop from 100 Hz to 1 Hz.

After the discovery of the sources of this problem, countermeasures were investigated for each source.

Garbage collection

In Android, the garbage collector is a background process that cleans up no longer used memory. In the Android Wear application the conversion from sensor data to log string created a significant amount of temporary Float variables. These short lived variables needed to be cleared from memory by the garbage collector. These clean up events are indicated in the Android Studio profiler as a bin symbol, see Figure 8. During these clean up events, the main code is halted and is unable to obtain new samples from the sensors. The impact of the garbage collection was mitigated by the use of the StringBuilder class and a change in the Java Virtual Machine settings (which also control the garbage collector). The following setting was used to increase available memory, change the garbage collection mode and target maximum delay.

```
org.gradle.jvmargs=-Xmx1024M -Xms1024M -XX:+UseG1GC -XX:MaxGCPauseMillis=200
```

Figure 8 (right) displays the drop in garbage collection events.

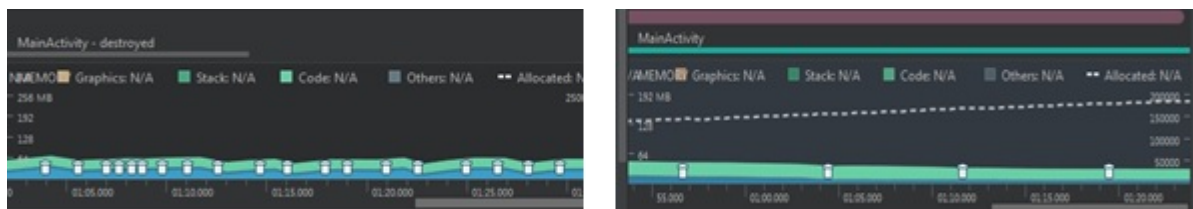


Figure 8 Garbage collection events (left) and improved memory management (right)

CPU priority

The main source of the frequency drop was the change in CPU priority level. In Android Wear an application will change to ambient mode, when the user does not interact with the device for several minutes. This is done to preserve the limited battery capacity. Applications in ambient mode are assigned a lower priority level, resulting in less time spend running on the CPU. In essence the application becomes a background process. This problem was tackled by forcing the application to stay out of ambient mode, using the following line of code.

```
getWindow().addFlags(WindowManager.LayoutParams.FLAG_KEEP_SCREEN_ON)
```

Signal loss

Next to the sampling frequency challenges, the PPG sensor used in this research has proven to be unstable in use. During several tests, the sensor output became corrupt with no clear cause, resulting in the output being 'Not a Number' (NaN). This does unfortunately mean that several tests were unusable. Possible cause of this problem could be the placement of the sensor or the way the Huawei Watch 2 translates the output of the AFE4405 chip [57] into the Android Wear sensor output fields.

6.2. Rerun

With the improvements made to the sampling application regarding garbage collection and CPU priority as mentioned in 6.1 the lab test was repeated. Unfortunately the rerun with test subjects 1-3 did again not result in gathered PPG data. Further investigation is needed to find the source or a workaround of the 'NaN' data problem. However the data gathered from test subject #5 proved very promising, see Table 4. This test of test subject #5 resulted in 18 correct windows in resting phase and 14 in stress phase (with 75% threshold), see Figure 9. Most of the cut data shown in this figure is due to incomplete windows at the point of switching between rest and stress phase.

	Minimum % of heart rate detected as NN intervals in window				
Test subject #	25%	50%	75%	95%	100%
1-3	NaN	NaN	NaN	NaN	NaN
5	84.51%	85.38%	79.44%	63.28%	18.41%

Table 4 Rerun heart rate detection rate

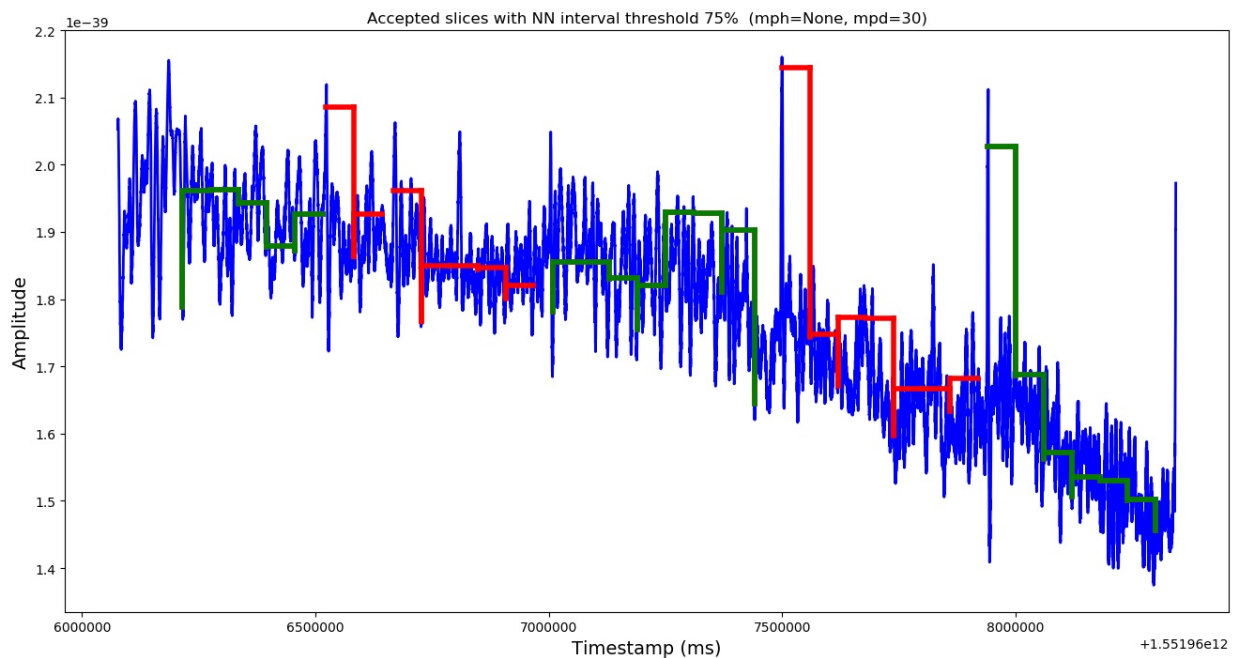


Figure 9 Accepted slices test subject #5 rerun

6.3. Stress detection

A possible relationship between gathered features and stress can be determined by looking at significant differences between rest and stress phases. If there are suitable features, the next step is to determine if a combination of features can be used to detect stress with significant confidence.

Wilcoxon

To determine if the features are suitable for stress detection, a Wilcoxon signed rank test was conducted. This statistical test determines if there are any significant differences between the data gathered during rest phase and stress phase. Figure 10 and Table 5 show that for test subject #5 the only three reliable features for stress detection are HR, RMSSD and SDNN (p-value < 0.05). In addition, it shows that the frequency domain features are outperformed by most of the time domain features.

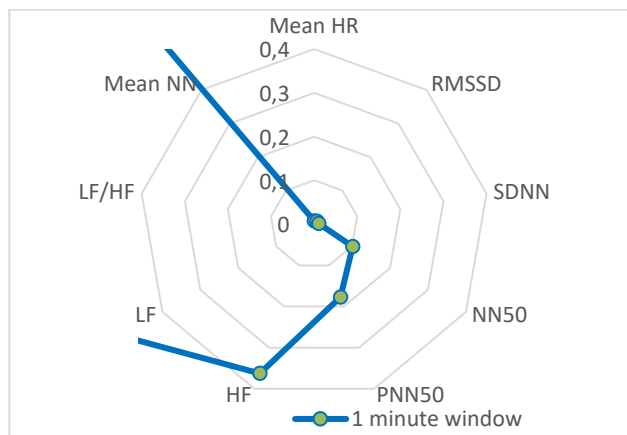


Figure 10 Wilcoxon p-value results test subject #5 rerun (75% threshold)

Feature	p-value
Mean HR	0.00763
RMSSD	0.009181
SDNN	0.011008
NN50	0.101708
PNN50	0.177007
HF	0.362686
LF	0.509797
LF/HF	0.593618
Mean NN	0.729891

Table 5 Wilcoxon results test subject #5

For the detection of stress, a combination of factors that change simultaneously would benefit the detection strength. To determine if any of the features have a significant pair wise correlation a custom correlation matrix was made using the Pandas library. Figure 11 displays this matrix of both rest and stress phases , where blue indicates a negative correlation and red a possitive correlation.

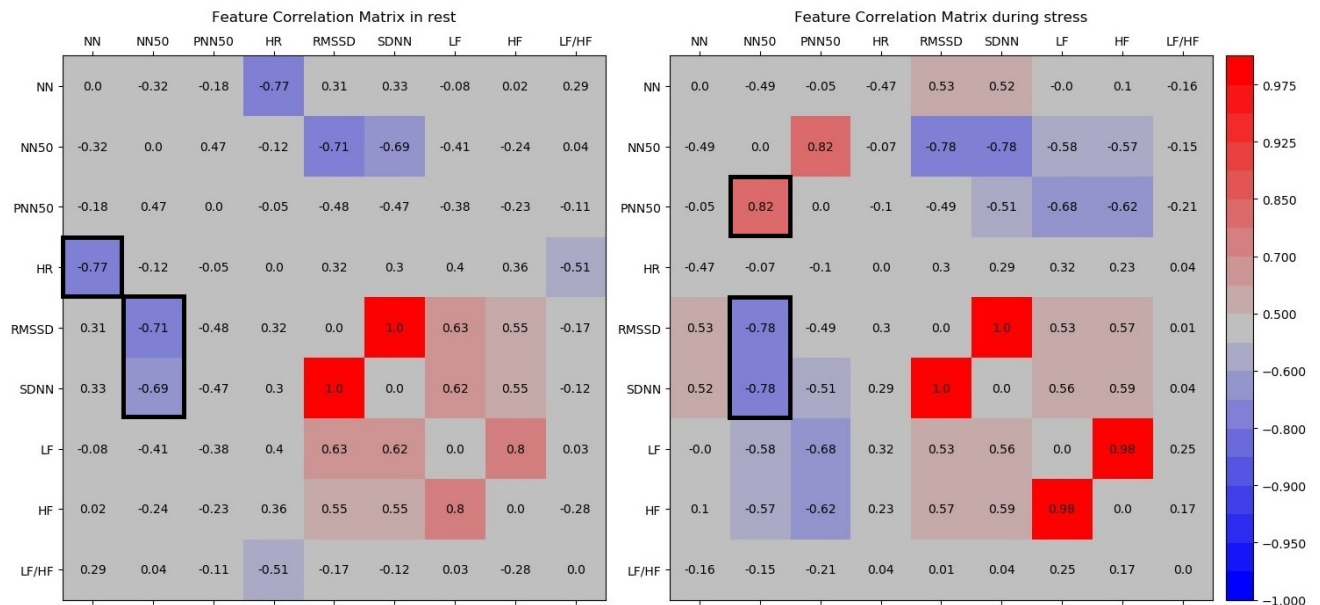


Figure 11 Correlation matrices test subject #5

Looking at the correlation there are several features that can act as stand-ins for one another. For instance RMSSD and SDNN are fully correlated, but also HF and LF. What can also be seen is that there is a change in correlation between NN50 and PNN50 during the rest and stress phase. Next to correlation between similar features, there are other correlations to be found (indicated with a black border in Figure 11):

- A weak correlation between RMSSD/SDNN and NN50
- In rest between HR and NN
- Between NN50 and PNN50 during stress phase.

Using these correlations one could focus on just these parameters in further study. In this study principle component analysis (PCA) was used to determine the most significant contribution to the separation power between stress and rest.

Self-labelling

In order to find a relationship between the features and stress, the self-labelling of the data points was investigated. This self-labelling can be done using a variety of machine learning techniques.

Figure 12 displays the result of a number of classifiers in combination with a number of dimensionality reduction methods (PCA kernels) [56]. In this figure the dots represent the rest or stress data points and the colored planes indicate the algorithms' clusters for the phases. Each subplot is a scatter plot of the two resulting principle components of the used PCA kernel. What can be seen is that the classifiers have a hard time determining the difference between the two datasets (rest/stress).

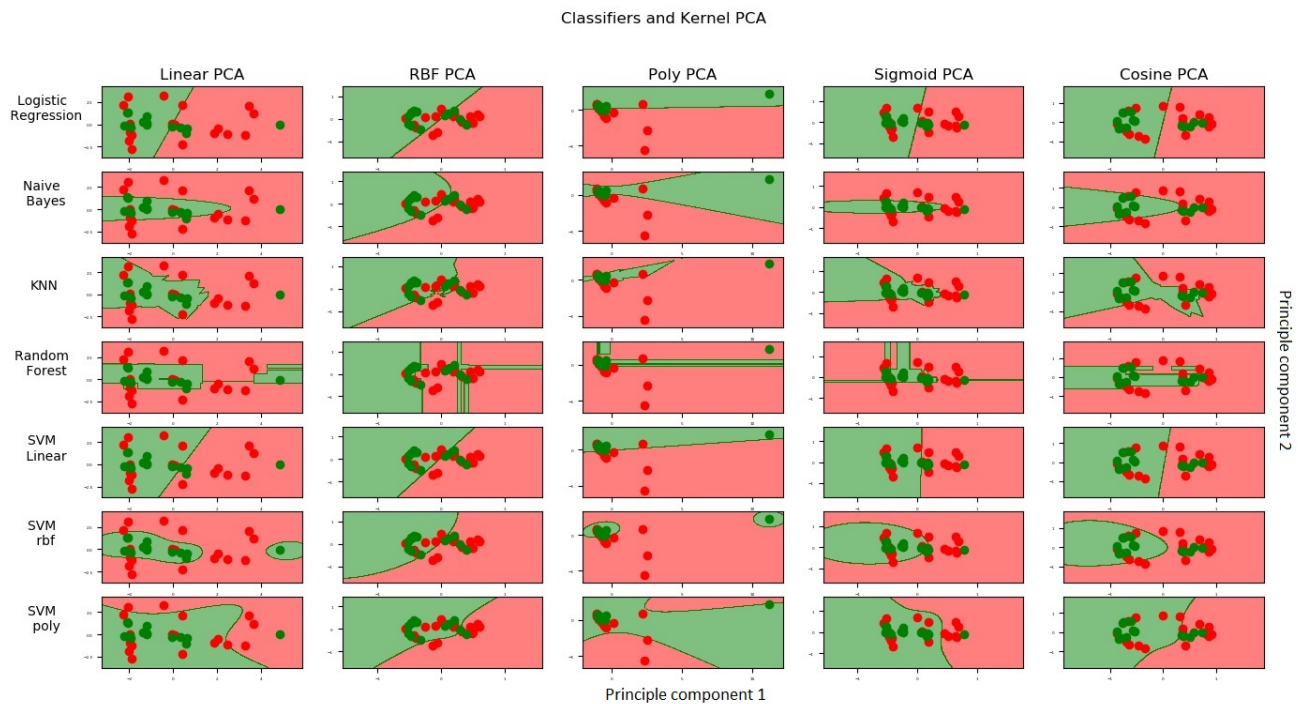


Figure 12 Separation between rest and stress data

This difficulty in separating the two datasets can also be seen in their respective F1-score when trained and tested (with training size 0.5), see Table 6. The F1-score represents the classifiers accuracy, taking into account both precision and recall. In this table also the DummyClassifier is considered for comparison, which randomly guesses the label.

Classifier	F1-score
LinearSVC	0.71
Naïve Bayes	0.82
LogisticRegression	0.75
KNeighborsClassifier	0.76
RandomForestClassifier	0
SVC Linear	0.7
SVC RBF	0.75
SVC Poly	0
DummyClassifier	0.63

Table 6 F1-scores of the classifiers

VAS score labelling

During the lab tests the test subjects were asked to indicate their stress level. This stress level was recorded as a score from one to ten at five points during the test. Figure 13 displays the levels from all participants. Interestingly the second stress test has a significantly higher stress impact. There are however clear differences per test subject. Some hardly react at all to the test, some recover fully from the tests and others carry their stress levels over to later phases.

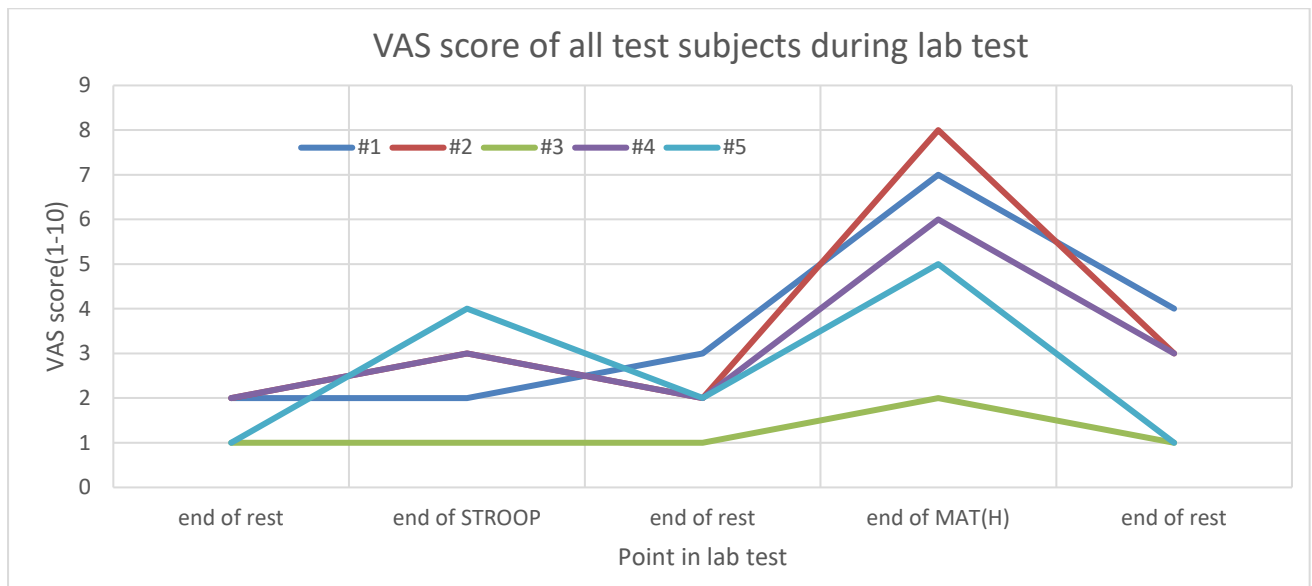


Figure 13 VAS score labelling by test subjects

In general it was considered hard to label the stress correctly. Several test subjects indicated they produced a value just for the sake of having a number and others said what they thought was expected from the test. The data did not correlate significantly to this labelling, therefore this approach was not continued further in the analysis.

6.4. Heart rate

The heart rate sensor did function correctly for all tests. With this gathered data, the effectiveness of the test can be determined. Looking at Figure 14, it can be seen that the heart rate of most test subjects hardly reacts to the test. This with the exception of test subject #4(which could be an anomaly). Test subject #2 does have a period of increased heart rate close to the MAT(H) test phase.

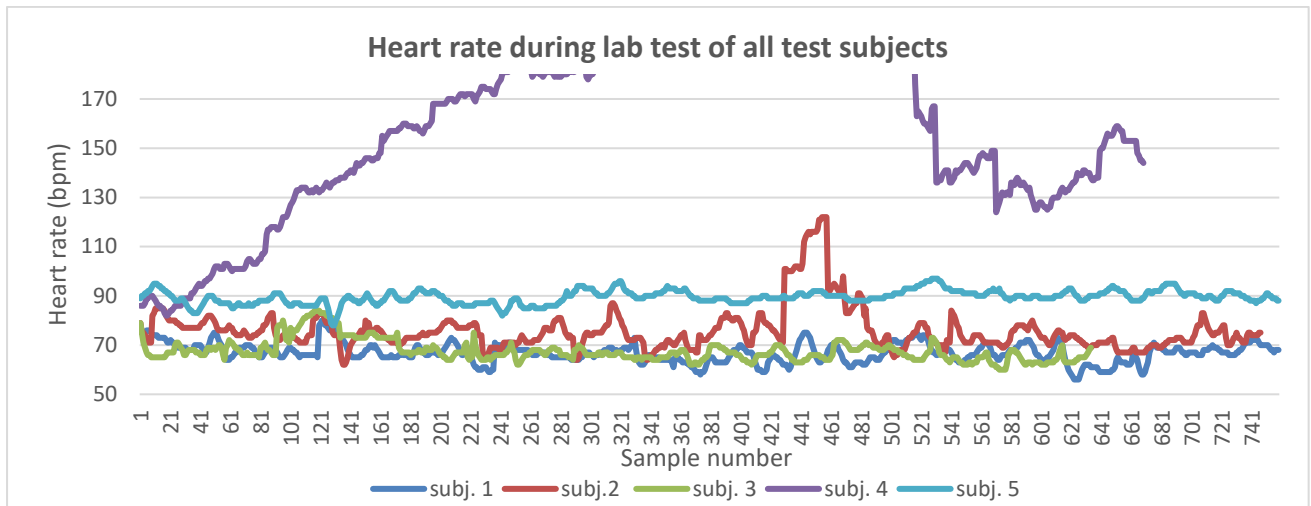


Figure 14 Heart rate during lab test of all test subjects

Zooming in on the data of test subject #2 an interesting phenomenon can be seen. Figure 15 displays this data, which is labeled according to the stress/rest phase. Interestingly the increased heart rate period is right before the second stress phase. This was also indicated by the test subject. The introduction or announcement of the second stress phase triggered a significant stress event. During the MAT(H) test the stress level of the test subject decreased again. This further limits the selective power between stress and rest phases.

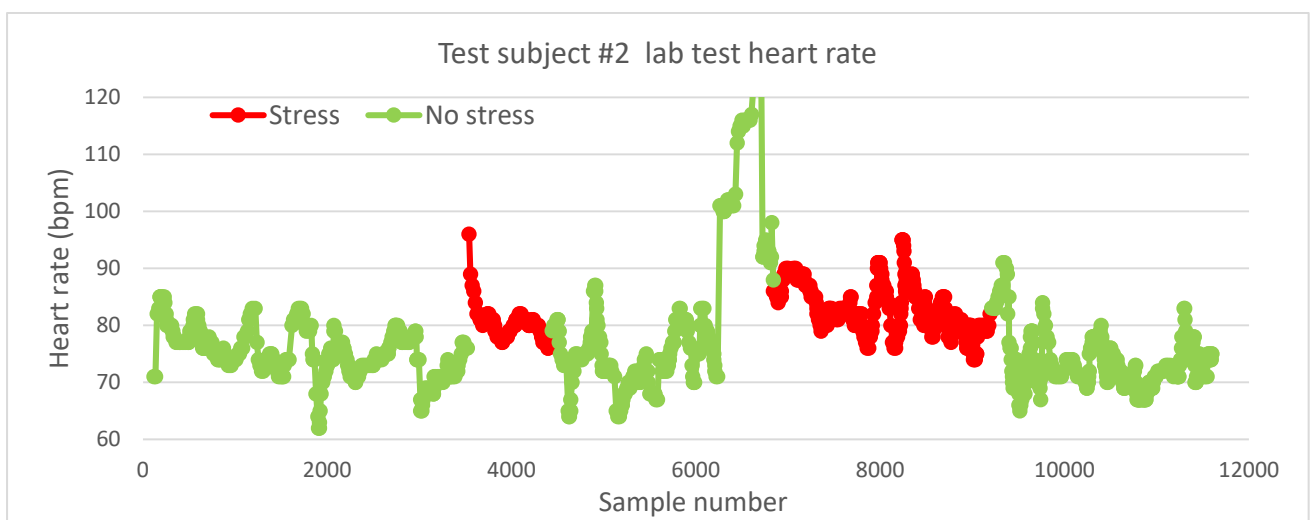


Figure 15 Stress before start of stress phase

After the indicated stress phases it can also be seen that the stress levels continue and take time to recover. The distinction between phases is not as clear as the yes/no labelling requires.

6.5. Field test

After the lab tests, the test subjects were asked to wear the watch for an extended period of time. During this period the test subjects started the app when they arrived at work and stopped the recording when they left work. During this period it often happened that the test subject was not working on certain days and as such no recordings were made during those days. The field test was stopped when at least five days of recording was done.

During the field test no labelling was done in order to keep the measurement as non-hindering as possible and reduce the white coat effect. Figure 16 shows the result of test subject #1. This figure shows over 24 hours of recording, in which several peaks in heart rate can be seen. The most significant increase can be seen on day 4, but after feedback from the test subject this turned out to be a period of significant physical activity. Therefore either the watch picked up a 'resting' heart rate in between activity or the heart rate lingers after activity.

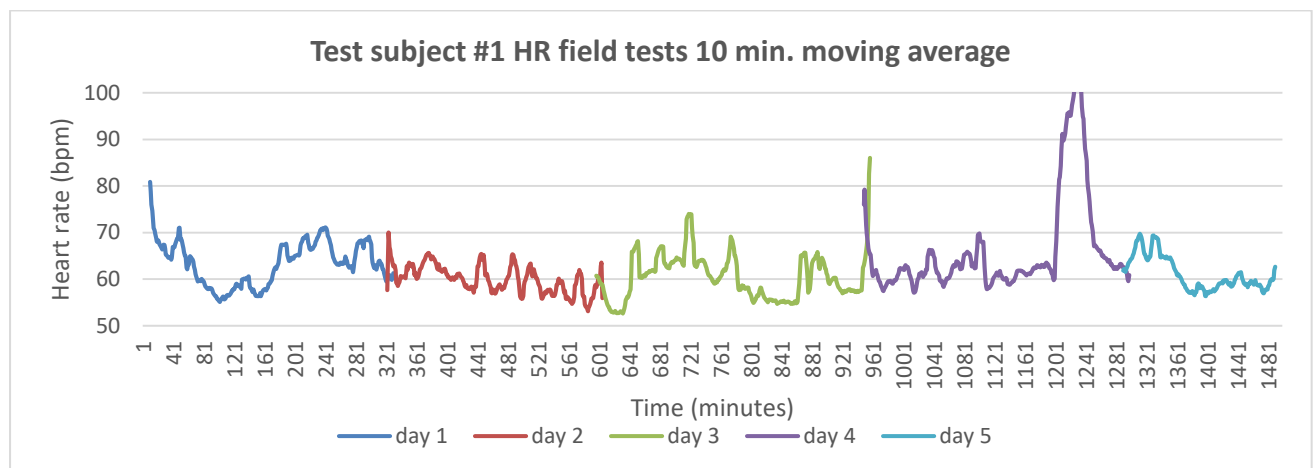


Figure 16 Heart rate during field test with significant event

The field test data of test subject #5 shows a linear increase over five measurement days, including some significant events, see Figure 17. These five days were recorded over an eight day period which included a non-monitoring weekend in the middle. During these days the average heart rate increased from 75 bpm to 85 bpm. This could be an indication of stress build-up. The increase also continues after the weekend, which could indicate an insufficient amount of rest.

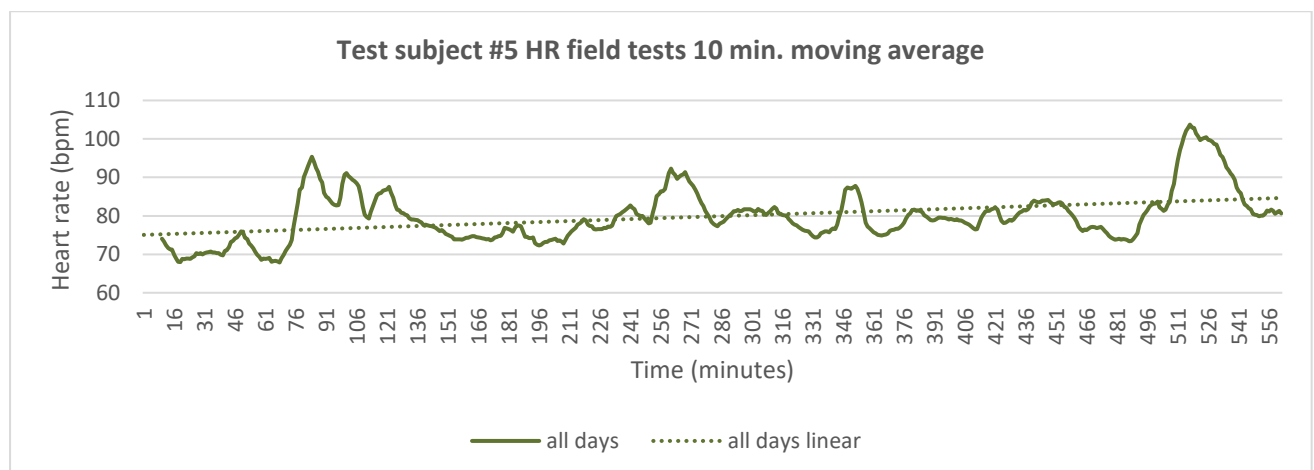


Figure 17 Heart rate field test with linear increase trend

7. DISCUSSION

Due to the limited use of the first tests, the planned control group tests could not be executed. Instead a rerun with the same test subjects was done. The resulting limited amount of test subjects does not allow for results with a strong statistical backing. Within the test subjects however there are already differences in response and stress detectability.

The rerun of the test with already used test subjects could have an influence on the response to the lab stress test. Test subjects could have gotten used to the stress exercise or at least better cope with it. The added stress of an unknown stress test is lost in this rerun. Further studies should use test subjects who have not had any similar test before in recent past.

In general the proposed stress test setup did not perform as expected from literature. Other studies found a clear distinction between stress and test phases in contrast to the results of this study. Some studies do also confirm the problems with test setups, e.g. due to test subjects enjoying the test instead of perceiving stress [34]. Test results are also influenced by the omission of test subjects or test subject grouping based on pre-test surveys [34][39]. In this study it is clear that individual test subjects react differently to the test. A test that allows for correct labelling of stress and rest phases in a diverse population of test subjects is yet to be found.

8. CONCLUSION

With the information gathered during the tests, the following conclusions can be drawn. The conclusion can then be used as input for further studies. The main research question that was drawn up was as follows:

“How can stress patterns be detected by measuring physiological variables using existing wearables in real-life situations to indicate a rise in the amount of stress events?”

To answer the main research question the following sub-questions were defined, each with their own answers:

- ***What physiological variables are linked to stress?***
Several studies have indicated that a number of physiological variables are linked to stress. However studies do also contradict each other or at least have varying statistical backing. In this research a set of variables was selected, being: mean HR, NN interval, NN50, PNN50, RMSSD, SDNN, LF HRV, HF HRV and the ratio LF-HF HRV. It is possible that other variables are also useful for the detection of stress. The HR and RMSSD/SDNN features show most separation power between rest and stress phases.
- ***What method is most suitable for detecting stress events in real-world scenarios?***
With the impact of the white coat effect, there is only the option of using a device that is already accepted by the general public for day to day use. Currently only a watch/bracelet is suitable for this application.
- ***What existing wearables meet the set requirements?***
There is only a limited set of wearables that can measure any variables linked to stress. Heart rate seems to be the standard parameter but there are a number of features in the works. A raw PPG signal is also hard to find due to the ever growing closed source approach to development in this field. For this project the Huawei Watch 2 was used. This wearable can measure the desired features, but proved to be unstable in use. A build in peak detection API would also benefit this application or even better a build in HRV feature output. Future study will also benefit from upcoming wearables that allow for the recording of other stress related variables e.g. blood pressure.
- ***What standardized test procedures allow for labelled datasets?***
There is a large amount of variation on the stress tests recommendations found in literature. There is also limited focus on the type of test subject being used. The test length is also debatable, especially with the limitation of several HRV features being an average over a minimum window of one minute.
- ***How can the measured variables be labelled in relation to stress?***
The labelling done by the test subjects has proven in literature to be unusable. Within this research it also became clear that test subjects can hardly justify a change or even make up a change just because it was expected. Labelling based on the phase in the test also has its limitations, but can be used with the right setup.

- ***How can the measured variables be combined to detect a pattern in stress events before a user can indicate it?***

It is clear that some features are more interesting than others when determining stress. There are several machine learning techniques that could indicate a significant event, but the accuracy of these algorithms needs to be determined using a better dataset. The test length and test subjects did not allow for a significant amount of stress events to determine a pattern.

- ***How can the selected wearable be used to monitor a group of lecturers and other staff over a period of several weeks?***

The logistics of a large test setup has proven to be challenging. Leaving the control of the measurement up to the test subjects has the benefit of minimal effort but also produces mixed quality datasets. None of the test subjects had any problem with wearing the device for an extended period of time.

Looking back at the hypothesis it is clear that most points are not valid at this point. The hypothesis consisted of the following components:

- ***Based on the chosen parameters, the expectation is that stress events can be detected during lab and real-world tests.***

The lab test procedure did not deliver stress events as expected. Therefore this link cannot be clearly defined. The real-world test does indicate possible events based on heart rate.

- ***The heart rate and raw PPG signal will provide sufficient information to determine the distinction between stress and non-stress events.***

Due to the limited amount of data points and the test structure, there can be no clear correlation found between stress and non-stress events. It may very well be possible that the chosen parameters are able to provide sufficient information, but revised study needs to confirm that possibility.

- ***This while the addition of motion detection will sufficiently eliminate unwanted stress event labels.***

During lab test the motion detection was of limited use. The test subjects were not engaged in any physical activity. The excessive use of arm movements was correctly kept out of the recorded data. For the field test this needs to be reconsidered. Certain physical activities, such as cycling do not trigger a halt in data recording, which introduces unwanted data.

- ***The set timeframe of the tests will cover a sufficient amount of stress events to be detected.***

The lab test needs to be extended to allow for more data points. The field test did already indicate a number of possible stress events, but due to the lack of correlation options of the lab data this cannot be verified. There is a trend visible in the data of one of the test subjects, but the gathered information will become more interesting when more recording days are added. The impact of non-recording days in between the series needs to be investigated.

A final conclusion on the detection of stress using wearables is that a significant amount of further study needs to be done. The link between the measured features and stress is too weak, especially for a diverse population. For now stress events are still best recorded by the people themselves. While further study is done, employees should be guided and facilitated to cope with possible stress build-up to limit stress related absenteeism.

9. RECOMMENDATIONS

With the results and conclusions of this study in mind there are points for improvement. For future studies the following points are recommended for further investigation.

To further reduce the garbage collection rate that disrupts the PPG sampling, it is possible to write more efficient code. The available Float to String converter function in common libraries use a temporary String variable to build the result. This can be improved by writing a custom Float to String function, that does not use a temporary variable but which uses a reused char array.

Looking at the results of the lab tests it is clear that the transition from stress to rest phases and vice versa are not instant. One way to overcome this issue is to redesign the lab test setup. Instead of several parts, a setup with only one relax and stress phase would be better. With this structure there is only one transition from relax to stress which reduces the amount of features being wrongly labelled. Another idea would be to remove data around the transition point, to reduce mislabelling (e.g. ± 5 minutes). Furthermore future study in the occurrence of stress appearing just before a stress phase could provide valuable information.

The test length needs to be increased. With the proposed setup only a maximum of 30 samples can be gathered. With the amount of transitions and window slicing there isn't enough data to accurately train any model. A setup of 30 minutes rest and 30 minutes stress would be better. Even with removing data around the transition point it would theoretically produce 50 data points.

If the sampling rate needs to be reduced in order to allow for other application tasks, the sampling rate can at least be reduced to 33 Hz for the PPG data. Figure 18 shows that the peaks are still clearly detectable at this sampling rate.

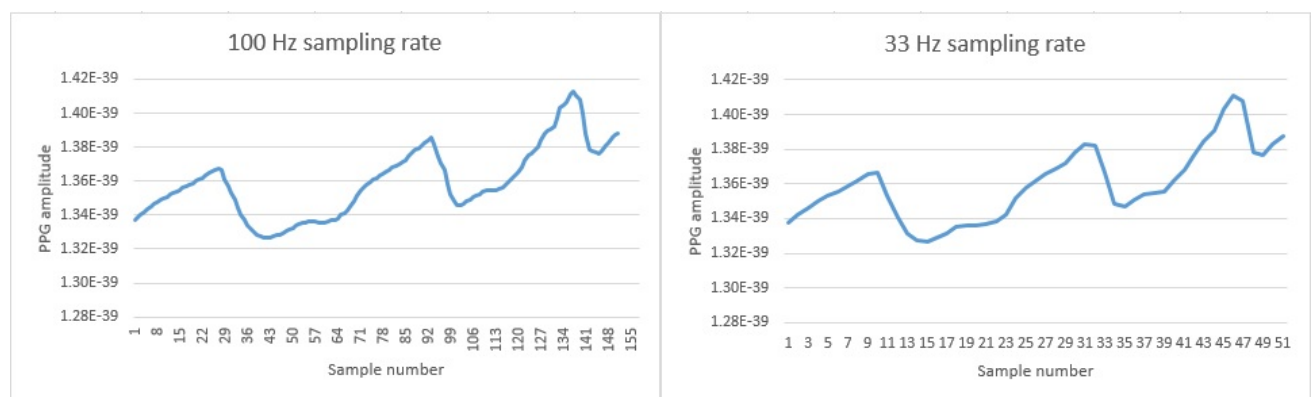


Figure 18 Effect of reducing sampling rate on the PPG signal

More study needs to be done on the effect of physical activity on stress detection. At this point there is too little information on the effect and the detection of these activities using wearables. For instance the detection of physical activity while keeping the wearable still is a concern.

For commercial use a more stable smart watch needs to be found and with sufficient documentation on the format of the output data.

This research used test subjects with at the very least a link to the technical domains. This could very well influence the response to proposed stress tests (Stroop and MAT(H)). Further research should look into the effect on several types of test subject groups.

Further study needs to look into the feedback and the application of the stress information. Correct application of the gathered data is of significant importance to the use and miss-use of the system.

10. REFERENCES

- [1] : TNO. (2017). Factsheet week van de werkstress 2017. Retrieved September 19, 2018, from <https://www.monitorarbeid.tno.nl/cijfers/factsheets/factsheet-werkstress>.
- [2] : Netherlands: EWCO comparative analytical report on Work-related Stress. (2011). Retrieved September 19, 2018, from <https://www.eurofound.europa.eu/publications/report/2011/netherlands-ewco-comparative-analytical-report-on-work-related-stress>.
- [3] : Lequeux, B., Uzan, C., & Rehman, M. B. (2017). Does resting heart rate measured by the physician reflect the patient's true resting heart rate? White-coat heart rate.
- [4] : Takahashi, D. (2018). Omron Healthcare blood pressure wearable. Retrieved September 19, 2018, from <https://venturebeat.com/2018/01/07/omron-healthcare-shows-off-wearable-blood-pressure-monitor-for-your-wrist/>.
- [5] : Stein, S. (2018). Omron Healthcare smartwatch CES. Retrieved September 19, 2018, from <https://www.cnet.com/news/omron-heartguide-blood-pressure-smartwatch-ces/>.
- [6] : Zhang, Q., Zhou, D., & Zeng, X. (2017). Highly wearable cuff-less blood pressure and heart rate monitoring with single-arm electrocardiogram and photoplethysmogram signals. *BioMedical Engineering OnLine*, 10.1186/s12938-017-0317-z.
- [7] : McCombie, D., Reisner, A., & Asada, H. (2006). Adaptive blood pressure estimation from wearable PPG sensors using peripheral artery pulse wave velocity measurements and multi-channel blind identification of local arterial dynamics. *Conf Proc IEEE Eng Med Biol Soc*. 2006;1:3521-4.
- [8] : Zhang, Y., Berthelot, M., & Lo, B. (2016). Wireless wearable photoplethysmography sensors for continuous blood pressure monitoring. *IEEE Wireless Health (WH)*.
- [9] : Socher, R. (n.d.). Innovative Bracelet Will Monitor Your Blood Pressure 24 Hours a Day. Retrieved September 19, 2018, from <https://www.wearable-technologies.com/2018/05/innovative-bracelet-will-monitor-your-blood-pressure-24-hours-a-day/>
- [10] : Allison, C. (2018, June 5). Asus' VivoWatch BP brings medical-grade blood pressure monitoring. Retrieved September 19, 2018, from <https://www.wearable.com/smartwatches/asus-vivowatch-bp-price-specs-release-date-6170>
- [11] : Langley, H. (2018, February 18). And finally: Samsung's blood pressure-tracking smartwatch and more. Retrieved September 19, 2018, from <https://www.wearable.com/smartwatches/samsung-gear-x-blood-pressure-monitor-992>
- [12] : Collins, T., Pires, I., Oniani, S., & Woolley, S. (2018, June 19). How reliable is your wearable heart-rate monitor? Retrieved September 19, 2018, from <https://theconversation.com/how-reliable-is-your-wearable-heart-rate-monitor-98095>
- [13] : Imotions. (n.d.). Galvanic Skin Response Measure emotional arousal & stress by measuring changes in the conductivity of the skin. Retrieved September 19, 2018, from <https://imotions.com/gsr/>
- [14] : Brown, B. (2017, July 24). Wearable Measures Skin Conductance to Assess Emotions. Retrieved September 19, 2018, from <https://healthtechinsider.com/2017/07/24/wearable-measures-skin-conductance-assess-emotions/>
- [15] : Shimmer. (n.d.). Shimmer3 GSR Unit. Retrieved September 19, 2018, from <http://www.shimmersensing.com/products/shimmer3-wireless-gsr-sensor>
- [16] : Simband. (n.d.). Simband Documentation. Retrieved September 19, 2018, from <https://www.simband.io/documentation/simband-documentation/>
- [17] : Broughall, N. (2018, February 7). Jawbone UP3 review. Retrieved September 19, 2018, from <https://www.techradar.com/reviews/wearables/jawbone-up3-review-1297589/review>
- [18] : Empatica. (n.d.). E4 Wristband documentation. Retrieved September 19, 2018, from <https://www.empatica.com/en-eu/research/e4/>

- [19] : Purcher, J. (2017, June 15). Apple Invents an Apple Watch Feature to measure Respiration Rate with Multi-Band Plethysmography. Retrieved September 19, 2018, from <http://www.patentlyapple.com/patently-apple/2017/06/apple-invents-an-apple-watch-feature-to-measure-respiration-rate-with-multi-band-plethysmography.html>
- [20] : Doberenz, S., Roth, W. T., Maslowski, N. I., Wollburg, E., & Kim, S. (2011). Methodological Considerations in Ambulatory Skin Conductance Monitoring. *International Journal of Psychophysiology : Official Journal of the International Organization of Psychophysiology*, 80(2), 87–95. <http://doi.org/10.1016/j.ijpsycho.2011.02.002>
- [21] : Dr. St. John, T.M. (2018, August 14). Factors That Affect Respiration Rate. Retrieved September 19, 2018, from <https://www.livestrong.com/article/108342-factors-affect-respiration-rate/>
- [22] : American Heart Association. (2015, July 31). All About Heart Rate (Pulse). Retrieved September 19, 2018, from www.heart.org/en/health-topics/high-blood-pressure/the-facts-about-high-blood-pressure/all-about-heart-rate-pulse
- [23] : Jaliman, MD, D. (2017, July 11). Causes of High Blood Pressure. Retrieved September 19, 2018, from <https://www.webmd.com/hypertension-high-blood-pressure/guide/blood-pressure-causes#1>
- [24] : Wijsman, J. L. P. (2014). Sensing stress: stress detection from physiological variables in controlled and uncontrolled conditions Enschede: Twente University Press (TUP) DOI: 10.3990/1.9789036537858, 10.3990/1.9789036537585
- [25] : Liu, Yun & Siqing, Du. (2017). Psychological Stress Level Detection Based on Electrodermal Activity. *Behavioural Brain Research*. 341. 10.1016/j.bbr.2017.12.021.
- [26] : Skoluda, Nadine & Strahler, Jana & Schlotz, Wolff & Niederberger, Larissa & Marques, Sofia & Fischer, Susanne & Thoma, Myriam & Spoerri, Corinne & Ehlert, Ulrike & M. Nater, Urs. (2015). Intra-individual psychological and physiological responses to acute laboratory stressors of different intensity. *Psychoneuroendocrinology*. 51. 10.1016/j.psyneuen.2014.10.002.
- [27] : Stress Effects on Mood, HPA Axis, and Autonomic Response: Comparison of Three Psychosocial Stress Paradigms. Giles GE, Mahoney CR, Bruny  TT, Taylor HA, Kanarek RB (2014) Stress Effects on Mood, HPA Axis, and Autonomic Response: Comparison of Three Psychosocial Stress Paradigms. *PLOS ONE* 9(12): e113618.
- [28] : Stroop, J.R., 1935. Studies of interference in serial verbal reactions. *J. Exp. Psychol.* 18, 643—662.
- [29] : Kirschbaum, C., Pirke, K.M., Hellhammer, D.H., 1993. The ‘TrierSocial Stress Test’ —a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology* 28(1—2), 76—81. Lazarus, R.S., Folkman,
- [30] : Allgrove, J.E., Gomes, E., Hough, J., Gleeson, M., 2008. Effects of exercise intensity on salivary antimicrobial proteins and markers of stress in active men. *J. Sports Sci.* 26 (6), 653—661.
- [31] : Kusserow, M., Amft, O., and Troster, G. (2013). Monitoring stress arousal in the wild. *IEEE Pervasive Computing Magazine*, 12(2):28–37.
- [32] : Huysmans, Dorien & Smets, Elena & De Raedt, Walter & Van Hoof, Chris & Bogaerts, Kathleen & Diest, Ilse & Helic, Denis. (2018). Unsupervised Learning for Mental Stress Detection - Exploration of Self-organizing Maps. 10.5220/0006541100260035.
- [33] : Rong-Chao Peng, Xiao-Lin Zhou, Wan-Hua Lin, and Yuan-Ting Zhang, “Extraction of Heart Rate Variability from Smartphone Photoplethysmograms,” *Computational and Mathematical Methods in Medicine*, vol. 2015, Article ID 516826, 11 pages, 2015. <https://doi.org/10.1155/2015/516826>.
- [34] : Mario Salai, Istv n Vass nyi, and Istv n K sa, “Stress Detection Using Low Cost Heart Rate Sensors,” *Journal of Healthcare Engineering*, vol. 2016, Article ID 5136705, 13 pages, 2016. <https://doi.org/10.1155/2016/5136705>.
- [35] : Mozos O. M., Sandulescu V., Andrews S., Ellis D., Bellotto N., Dobrescu R., et al. . (2017). Stress detection using wearable physiological and sociometric sensors. *Int. J. Neural Syst.* 27:1650041. 10.1142/S0129065716500416

- [36] : DeGiorgio, C. M., Miller, P., Meymandi, S., Chin, A., Epps, J., Gordon, S., ... Harper, R. M. (2010). RMSSD, a Measure of Heart Rate Variability, Is Associated With Risk Factors For SUDEP: The SUDEP-7 Inventory. *Epilepsy & Behavior* : E&B, 19(1), 78–81.
<http://doi.org/10.1016/j.yebeh.2010.06.011>
- [37] : Shaffer, F., & Ginsberg, J. P. (2017). An Overview of Heart Rate Variability Metrics and Norms. *Frontiers in Public Health*, 5, 258.
<http://doi.org/10.3389/fpubh.2017.00258>
- [38] : J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Transactions on Biomedical Engineering*, vol. 32, no. 3, pp. 230–236, 1985
- [39] : A. Sano and R. W. Picard, "Stress Recognition Using Wearable Sensors and Mobile Phones," *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, Geneva, 2013, pp. 671-676.doi: 10.1109/ACII.2013.117
- [40] : Martin Gjoreski , Hristijan Gjoreski , Mitja Luštrek , Matjaž Gams, Continuous stress detection using a wrist device: in laboratory and real life, *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, September 12-16, 2016, Heidelberg, Germany [doi>10.1145/2968219.2968306]
- [41] : Agateller (Anthony Atkielski), 13 January 2007 , accessed: 14-nov-2018 ,
<https://en.wikipedia.org/wiki/Electrocardiography#/media/File:SinusRhythmLabels.svg>
- [42] : Android Studio 3.3, Build #AI-182.5107.16.33.5199772, built on December 25, 2018 , JRE: 1.8.0_152-release-1248-b01 amd64 , JVM: OpenJDK 64-Bit Server VM by JetBrains s.r.o , Windows 7 6.1
- [43] : Android Developers. (n.d.). Android Debug Bridge (adb). Retrieved March 18, 2019, from
<https://developer.android.com/studio/command-line/adb>
- [44] : Appfour. (n.d.). Wi-Fi Manager for Wear OS (Version 1.0.190312) [Computer software]. Retrieved March 18, 2019, from
<https://play.google.com/store/apps/details?id=com.appfour.wearwifi>
- [45] : Android Developers. (n.d.). Configure on-device developer options. Retrieved March 18, 2019, from
<https://developer.android.com/studio/debug/dev-options.html>
- [46] : Tournade, Y. (2015, November 1). Peak Detection in the Python World. Retrieved from
<https://blog.ytotech.com/2015/11/01/findpeaks-in-python/>
Supporting github : <https://github.com/MonsieurV/py-findpeaks>
- [47] : SciPy.org. (n.d.). Scipy.signal.find_peaks_cwt. Retrieved December 21, 2018, from
https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.find_peaks_cwt.html
- [48] : SciPy.org. (n.d.). Scipy.signal.argrelextrema. Retrieved December 21, 2018, from
<https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.argrelextrema.html>
- [49] : SciPy.org. (n.d.). Scipy.signal.find_peaks. Retrieved December 21, 2018, from
https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.find_peaks.html
- [50] : Tournade, Y. (n.d.). Peakutils.peak.indexes. Retrieved December 21, 2018, from
https://github.com/MonsieurV/py-findpeaks/blob/master/tests/peakutils_indexes.py
- [51] : Bergman, S. (n.d.). Peakdetect Sixtenbe. Retrieved December 21, 2018, from
<https://gist.github.com/sixtenbe/1178136>
- [52] : Slavic, J. (n.d.). Findpeaks Janko Slavic. Retrieved December 21, 2018, from
https://github.com/jankoslavac/py-tools/blob/master/findpeaks/Findpeaks_example.ipynb
- [53] : Castaldo, R., Montesinos, L., Melillo, P., James, C., & Pecchia, L. (2019). Ultra-short term HRV features as surrogates of short term HRV: a case study on mental stress detection in real life. *BMC medical informatics and decision making*, 19(1), 12.
doi:10.1186/s12911-019-0742-y

- [54] : Esco, M. R., & Flatt, A. A. (2014). Ultra-short-term heart rate variability indexes at rest and post-exercise in athletes: evaluating the agreement with accepted recommendations. *Journal of sports science & medicine*, 13(3), 535–541.
- [55] : Salahuddin, L., Cho, J., Jeong, M., & Kim, D. (2007). *Ultra short term analysis of heart rate variability for monitoring mental stress in mobile settings*. Conf Proc IEEE Eng Med Biol Soc. :4656-9.
- [56] : B. (2017, November 15). Visualizing Classifier Boundaries Using Kernel PCA. Retrieved from <https://www.kaggle.com/jsultan/visualizing-classifier-boundaries-using-kernel-pca>
- [57] : Texas Instruments. (n.d.). AFE4405 : Ultra-Small Integrated AFE With FIFO for Wearable Optical Heart-Rate Monitoring and Biosensing. Retrieved May 28, 2019, from <http://www.ti.com/product/AFE4405#>
- [58] : KNAW; NFU; NWO; TO2-federatie; Vereniging Hogescholen; VSNU (2018): Nederlandse gedragscode wetenschappelijke integriteit. DANS. <https://doi.org/10.17026/dans-2cj-nvwu>
- [59] : National Ethics Council for Social and Behavioural Sciences, (2018). Code of ethics for research in the social and behavioural sciences involving human participants
- [60] : General Data Protection Regulation (GDPR). (n.d.). Retrieved June 5, 2019, from <https://gdpr-info.eu/>

Appendix A : Test subjects information brief

Information Sheet Master Thesis Thies Keulen

Detecting stress patterns based on physiological measurements in real-life scenarios using existing wearables

The purpose of this research is to find a relationship between physiological parameters and stressful events. This relationship could then be used to prevent chronic stress symptoms from developing.

To find this relationship, several other factors need to be excluded. Therefore applicants with medication related to stress or medication which have an effect on heart rate are excluded. In this study only employees of the Hanze University of Applied Sciences are allowed to participate.

The participant of this study will be asked to take part in two tests. A laboratory setting test, in which the participant is asked to wear a watch and will be exposed to stress inducing tasks for 30 minutes. After about two weeks the participant is asked to start the second test. This test consists of wearing the watch during working hours for approximately 10 days.

At any time the participant can withdraw from the study by informing the researcher (Thies Keulen) and handing in the provided smart watch.

The two tests will gather a number of (physiological) parameters of the participant. The parameters are heart rate, motion and light absorption data at the wrist. From this data several heart rate variability characteristics will be determined. This is done to determine a relation between the measured parameters and stress events.

The gathered data will only be linked to an identification number on the watch. This data will only be accessed by the researcher (Thies Keulen). For consistency the same watch needs to be used in both tests. This link between watch and participant will be destroyed after completion of the second test.

At any time the participant has the right to request a full erasure of this data or where possible rectification. The gathered research data will be used for the duration of the academic year 2018-2019. The anonymized data will be used for the thesis report and thesis defense presentation(s).

Participating in this research has the following risks:

- There is still a possibility of exposure of the gathered data.
- The discomfort of wearing a watch for an extended period of time (during working hours).
- The laboratory setting tests exposes the participant to stress inducing tasks. This will most likely have the desired effect of causing stress to the participant.

Study contact details for further information or complaints:

Thies Keulen
Hanze University of Applied Sciences
School of Communication, Media & IT
Zernikeplein 11 , D.2.34
050-595 5654 / 06 28210559
h.t.keulen@pl.hanze.nl

Appendix B : Informed consent form

Informed Consent Form for Master Thesis by Thies Keulen

Detecting stress patterns based on physiological measurements in real-life scenarios using existing wearables

Please tick the appropriate boxes

Taking part in the study	Yes	No
I have read and understood the study information dated 08/JAN/2019, or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction.	<input type="radio"/>	<input type="radio"/>
I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason.	<input type="radio"/>	<input type="radio"/>
I understand that taking part in the study involves wearing a smart watch during the following two tests:	<input type="radio"/>	<input type="radio"/>
- A test in a laboratory setting. In which I'm asked to participate in stress inducing exercises. This tests will take 30 minutes. During this test I'm asked to score my stress level on multiple occasions. I'm asked to fill in this scoring form myself.	<input type="radio"/>	<input type="radio"/>
- A test during working hours for a period up to 10 days. During this time I'm asked to perform my daily activities as normal as possible.	<input type="radio"/>	<input type="radio"/>
Risks associated with participating in the study		
I understand that taking part in the study involves the following risks:	Yes	No
- The possible exposure of my physiological data collected during the tests, consisting of heart rate, heart rate variability parameters, motion and light absorption data from the wrist.	<input type="radio"/>	<input type="radio"/>
- The discomfort from wearing a watch for an extended period of time (during working hours)	<input type="radio"/>	<input type="radio"/>
- Being exposed to stress inducing tests during the lab test (30 minutes). These tests are specifically made to induce stress. This is done to detect a difference in the before mentioned physiological parameters.	<input type="radio"/>	<input type="radio"/>

Use of the information in the study

	Yes	No
I understand that information I provide will be used for the thesis report and thesis defence presentation(s)	<input type="radio"/>	<input type="radio"/>
I understand that personal information collected about me that can identify me, such as [e.g. my name], will not be shared with others.	<input type="radio"/>	<input type="radio"/>
I agree to joint copyright of the information provided by me via tests evaluations and feedback to Thies Keulen	<input type="radio"/>	<input type="radio"/>

Signatures

_____	_____	_____
Name of participant	Signature	Date

I have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands to what they are freely consenting.

_____	_____	_____
Researcher name	Signature	Date

Study contact details for further information:

Thies Keulen
050-595 5654 / 06 28210559
h.t.keulen@pl.hanze.nl