



School of Medicin and Health (SMH)
Fredrik Bajers Vej 7D, 9220 Aalborg Øst
Phone: +45 99 40 37 48
blc@hst.aau.dk

AALBORG UNIVERSITET
STUDENTERRAPPORT

Title: Assessing mental workload using seismocardiography
Semester: 10th semester, Master Thesis in Sports Technology
Project period: 1st Feb – 6th June 2019
ECTS: 30
Supervisor: Anderson de Souza Castelo Oliveira
Project group: 10211

Jacob Guy Diemar

Mikkel Jul Hansen

The aim of the current study was to investigate the feasibility of assessing mental load using seismocardiography by means of heart rate variability (HRV) analysis and machine learning. Twelve participants completed a mental computer task on 3 difficulty levels, on two days separated by at least a week. Electrocardiography (ECG) and seismocardiography (SCG) recordings were concurrently obtained, and a performance score based on the mental task were computed. Participants furthermore subjectively rated their mental workload (MWL) using the NASA-TLX. Cardiac cycle intervals were independently extracted from both ECG and SCG recordings and the HRV was analyzed in both the time- and frequency domain. The HRV results, subjective ratings and performance scores were statistically tested using a Two-way ANOVA with repeated measures, between days and MWL levels. Intraclass correlation coefficients (ICC) were furthermore computed to assess the agreement between the ECG and SCG based HRV. Features from the cardiac cycle segmented SCG signals were extracted and used for classification of MWL levels using machine learning. Significant differences were found for both subjective ratings and performance scores between days and MWL levels. HRV measures showed significant difference in the Peak LF measure between MWL levels. ICC values between ECG and SCG based HRV varied between poor and excellent agreement. Classification of MWL using SCG signals was unsuccessful using the included features. It can be concluded that SCG seems to be feasible for running HRV analysis due to an automatic noise removal and cardiac cycle segmentation of SCG signals being successful. However, further work is required to potentially implement successful classification of MWL using SCG.

Characters: [103.765]
Pages: [62]
Appendix: [1]

Ved at underskrive dette dokument bekræfter hvert enkelt gruppemedlem, at alle har deltaget lige i projektarbejdet og at alle således hæfter kollektivt for rapportens indhold.

Assessing mental workload using seismocardiography

Jacob Guy Diemar & Mikkel Jul Hansen

Abstract

The aim of the current study was to investigate the feasibility of assessing mental load using seismocardiography by means of heart rate variability (HRV) analysis and machine learning. Twelve participants completed a mental computer task on 3 difficulty levels, on two days separated by at least a week. Electrocardiography (ECG) and seismocardiography (SCG) recordings were concurrently obtained, and a performance score based on the mental task were computed. Participants furthermore subjectively rated their mental workload (MWL) using the NASA-TLX. Cardiac cycle intervals were independently extracted from both ECG and SCG recordings and the HRV was analyzed in both the time- and frequency domain. The HRV results, subjective ratings and performance scores were statistically tested using a Two-way ANOVA with repeated measures, between days and MWL levels. Intraclass correlation coefficients (ICC) were furthermore computed to assess the agreement between the ECG and SCG based HRV. Features from the cardiac cycle segmented SCG signals were extracted and used for classification of MWL levels using machine learning. Significant differences were found for both subjective ratings and performance scores between days and MWL levels. HRV measures showed significant difference in the Peak LF measure between MWL levels. ICC values between ECG and SCG based HRV varied between poor and excellent agreement. Classification of MWL using SCG signals was unsuccessful using the included features. It can be concluded that SCG seems to be feasible for running HRV analysis due to an automatic noise removal and cardiac cycle segmentation of SCG signals being successful. However, further work is required to potentially implement successful classification of MWL using SCG.

Keywords: *seismocardiography, heart rate variability, machine learning.*

Introduction

Occupations in modern society have experienced a precedence of non-manual work over manual work (Stansfeld & Candy, 2006). This have led to an increase in the mental demands, which may lead to work related stress and the detrimental effects associated if the mental demands exceed the capabilities of the individual (Holmes, 2001). One of the occupations at risk is e-sport athletes, as this arising industry has led to increased professionalism and thereby prolonged exposure to high mental demands, as athletes needs to think strategically and make fast and smart decisions to achieve successful performance (Bányai et al., 2018). Prolonged exposure to these higher demands, also known as chronic stress, ultimately resulting in the inability to cope with one's work, both psychologically and emotionally, is known as one of the most frequent health issues these days (van Daalen et al., 2009). For individuals with chronic stress, multiple health issues arise like weakened immune system, volume changes in certain brain areas (Mariotti, 2015), increased risk towards hypertension, stroke or heart attack, lowered reproduction capability etc. (Slavich, 2016; APA, n.d.).

Several different measures of mental workload (MWL) have been used in the existing literature, which can be differentiated as being either physiological, subjective or task performance measures. The physiological measures relate to respiration, blood pressure, eye-tracking, brain activity, electrodermal and cardiac activity, whereas subjective measures relates to self-reporting of the experienced MWL (NASA-TLX, SWAT etc.). Performance measures include parameters such as completion time, reaction time and error rate. Cardiac activity has been the most common physiological measure of MWL, due to certain correlations between heart rate variability (HRV) and MWL and the relatively simple and unobtrusive employment. (Charles & Nixonl. 2019)

Multiple previous studies have implemented the use of electrocardiography (ECG) to obtain HRV (Taelman et. al., 2011; Blitz. et al., 1970; Charles & Nixon, 2019). Another emerging method for monitoring cardiac activity is seismocardiography (SCG), which utilizes highly sensitive accelerometers to capture heart-induced motion.

The first implementation of SCG dates back to 1956 (Mounsey, 1956), using big, bulky and insensitive accelerometers. However, the recent advances in accelerometer technology, resulting in lighter, smaller and more sensitive accelerometers, now allows the technology to be used to capture new information regarding cardiac activity (Taebi et al., 2019; Paukkunen, 2014). SCG signals contain several concurrent waveforms, representing distinctive systolic and diastolic components, with representations of specific events in the cardiac cycle like aortic valve opening, isovolumic contraction etc. (Paukkunen, 2014).

No previous studies have investigated possible correlations between SCG signal morphology and MWL. If examined, this could potentially lead to new screening methods for evaluating MWL, which could be used to evaluate mental performance over time or allocate team roles, for example in high demanding mental tasks like e-sport. Furthermore, accelerometers have shown to be useful for obtaining information in many applications, such as physical activity monitoring, posture correction, balance evaluation etc. making for a health monitoring system that could evaluate multiple parameters (Attal et al. 2015).

With existing knowledge on the use of ECG to evaluate mental load and the recent advances in SCG technology, the aim of the current study was to investigate the feasibility of assessing MWL using SCG.

Method

All participants (12 males) (Age = 26 ± 1 year, body mass = 81.8 ± 9.3 , BMI = 24.9 ± 1.8 kg/m²) had normal or corrected to normal vision and were right handed computer mouse users. Participants were all non-smokers, non-drug addicts and had no known mental or heart diseases. Participants were instructed to abstain from alcohol (24h), caffeine (12h), painkillers and sleep medicine (24h), prior to the experiment. All participants were furthermore instructed to sleep a minimum of 7 hours prior to the experiment (reported sleep = 7.4 ± 1.1 hours). All experimental trials were conducted between 9am - 3pm. All participants signed a declaration of consent.

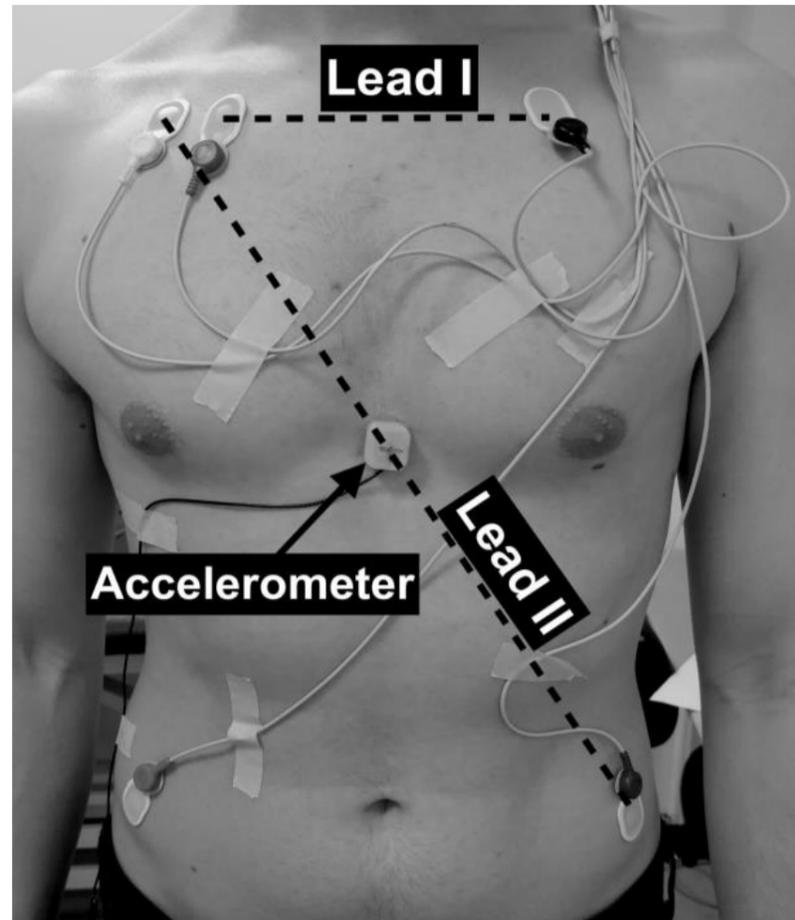


Figure 1: ECG electrodes and SCG accelerometer placement.

The hardware setup consisted of a combined ECG and SCG measuring system. A five electrodes setup for the ECG, as well as a single accelerometer for the SCG, was connected to the same amplifier/AD-converter (IWorx 214) for synchronized data-logging, sampling at 1000 Hz. The hardware was connected via USB to a computer, running a data-logging software (IWorx LabScribe V3.62).

For the recording of SCG signals, a small $\pm 2g$ low-noise ($5\mu g/\sqrt{Hz}$) capacitive sensing accelerometer (Silicon Designs model 1221) with a sensitivity of 2000 mV/g was used. The accelerometer was encased in a small (10x10x8mm) lightweight 3D-printed PLA box and was placed on the lowest part of sternum with double adhesive tape. Furthermore, the wire was secured to the participants chest forming a small loop, to limit wire movement noise, and made sure not to overlap with the ECG electrode wires (Figure 1).

A two-lead ECG, with Lead I and II configuration was used, utilizing five electrodes (Ambu[®] Neuroline 720, Ag/AgCl wet electrode) placed with one acting as ground (right leg), two as Lead II configuration (right arm (negative) and left leg (positive)) and two as lead I configuration (right arm (negative) and left arm (positive)) (Figure 1).

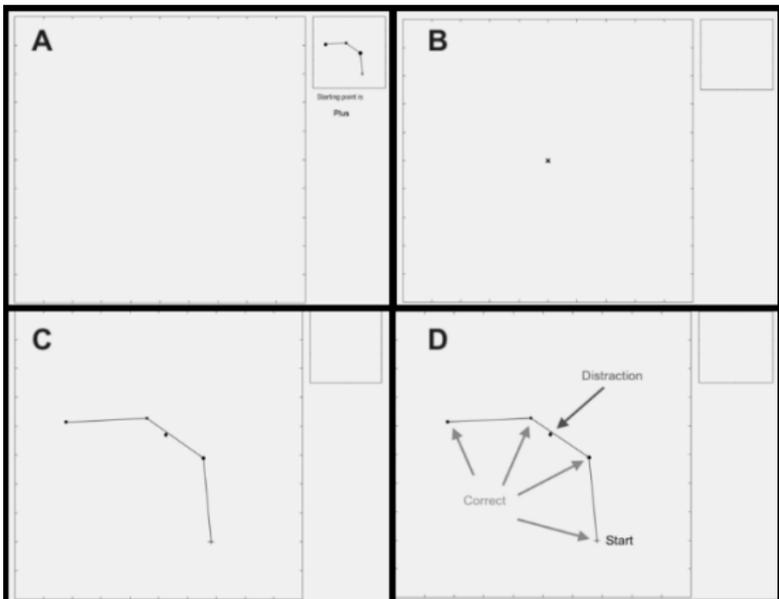


Figure 2: Mental task. **A:** memorization period, **B:** washout period, **C:** replication period, **D:** replication period with starting-, distraction- and correct pattern points indicated.

Mental task

A graphical user interface running a connect-the-dots game (WAME 1.0), developed at Aalborg University (Marandi et al., 2018A), was used to induce the MWL and consisted of cyclic computer operations where the participants had to recreate a specific pattern in a certain order. Each cycle involved a memorization period (MP), washout period (WP), and replication period (RP) (Figure 2). The task was displayed on a Dell E193FP 19-inch LCD monitor with 1280x1024 resolution and 75Hz refresh rate, and was placed approximately 55cm from the participants eyes with the center of the screen placed approximately 15° degrees below the horizontal line of sight (Marandi et al., 2018A) (Figure 3).

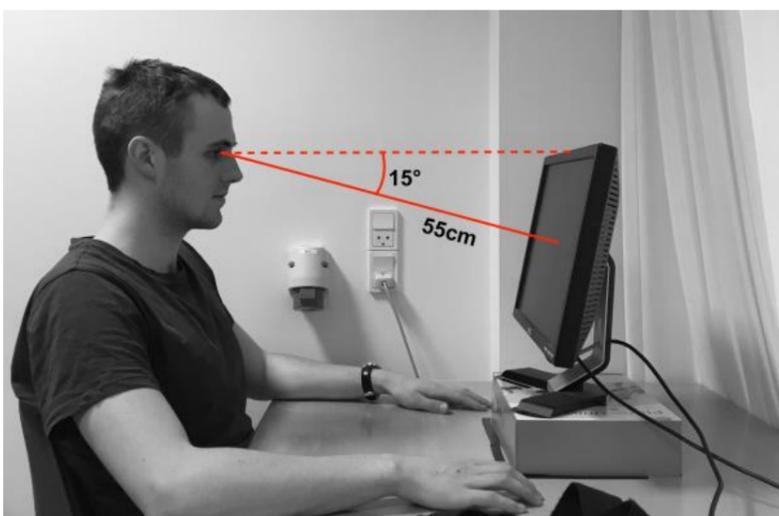


Figure 3: Mental task workstation with eyesight angle and distance illustrated.

The work panel contained a template panel with an appurtenant textual instruction and a replication

panel, which subtended respectively 5° and 20° of visual angle in both horizontal and vertical direction. The graphical user interface subtended approximately 27° of visual angle horizontally and 22° vertically. The area in which the participants were to complete the computer task was cordoned off to minimize the exposure of distracting elements in the surrounding environment.

Each mental task cycle was initiated by the MP, where a specific pattern composed by a series of points were showed on the template section (Figure 2, A). Each point could take various shapes i.e. plus, asterisk, circle, triangle, square, diamond, pentagram, and a short text indicated the starting point for the pattern. The MP was followed by the WP where the pattern in the template panel disappeared and a cross located in the center of the replication panel was used as indication of the WP (Figure 2, B). The mouse cursor was furthermore made invisible to avoid any prepositioning prior to the RP. The RP was initiated by a scaled version of the pattern points appearing on the replication panel, and the participant was now able to connect the points in the correct sequential order to replicate the same pattern shown during the MP (Figure 2, C). To indicate that the first point of the pattern was correctly clicked, the point was enlarged by a factor of two. Whenever the participant clicked on the points in the right order, a line was drawn to connect the newest correctly clicked point to the previous one, otherwise no line appeared. The number of pattern points (PP) to be connected and the geometrical complexity of the patterns were changed and thereby inducing three different levels of MWL referred to as low (PP = 4), medium (PP = 5) and high level (PP = 6). The geometrical complexity was changed such that the angles between any connecting lines were tightened with increased mental load (Marandi et al., 2018A). All MWL levels of the mental task also included a distraction point (DP), which were to be avoided in the replication of the pattern (Figure 2, D).

A library of randomly generated patterns was predefined and the order of patterns to be completed, were randomized for each participant. A time constraint of each period was predetermined and in accordance with Marandi et al., 2018A, being 2.06s, 2.34s, and 2.62s for MP and WP, and 4.11s, 5.06s, and 6.02s for RP in low, medium, and high levels of mental load respectively.

Experimental protocol

After participants had arrived, received information about the study and signed the declaration of consent, height and weight was measured, and sleep and substances ingested prior to the experiment (alcohol, caffeine, painkillers, sleep medicine etc.) was reported. The workstation at which the participants would be sitting, when completing the mental tasks, were adjusted to each participants height, to ensure a standardized visual distance and angle. Participants randomly drew an ID number matching a mental task level sequence. The sequence of the three different MWL tasks was counterbalanced across the participants

Next, the skin underneath each ECG electrode was prepared by removal of hair, light abrasion of the skin surface and cleansing using alcohol wipes, to lower the skin impedance. The accelerometer and electrodes were placed according to Figure 1.

The mental task was explained to participants and a training session consisting of at least 2 x 5 min (5 min low level and 5 min high level, repeated until familiar) was completed, to familiarize participants with the task, while equipped with the apparatus. After the familiarization period, participants weighted the different parameters of the NASA-TLX test, based on their experience with the task during the familiarization period. Next, participants rested for 10 min, before completing the first mental task level determined by the participant ID based sequence. After the mental task, participants were instructed to complete the NASA-TLX subjective rating. This was repeated until all 3 mental task levels (each consisting of 5 min) had been completed and rated, after which the trial was completed. The experimental protocol was completed twice on each participant on 2 different days, with at least 7 days in between.

Data analysis

An overall performance metric (OP) was computed for each task completed by each participant to address their dexterity. This OP metric was computed as the ratio of two other performance metrics, which quantifies how accurate and how fast the participant performed each task.

To account for the participant's clicking speed, the mean reaction time (MRT) was defined and computed in three different ways, depending on the degree of completion of the replication (Equation 1). If all the points in the pattern was correctly clicked,

the time intervals (TI) between the correct clicks (CC) and the first click with respect to the task onset time were averaged with respect to the number of pattern points (PP). If only some of the PP were correctly clicked, the remaining time of the replication period (RTRP) was added to the summation of time in between correct clicks and averaged with respect to the number of PP. If no PP were correctly clicked, MRT was equal to the length of the replication period.

$$MRT = \begin{cases} \frac{\sum_{i=1}^{CC} TI_i}{CC}, & \text{Completed pattern} \\ \frac{\sum_{i=1}^{CC} TI_i + RTRP}{CC + 1}, & \text{Partially completed pattern} \\ RP, & \text{No correct clicks} \end{cases} \quad (1)$$

The MRT was normalized with respect to the minimum of MRT across all participants (0.5120s). The parameter related to accuracy, selective attention (SelA), was defined as the ability to keep focused on a set of actions despite any distracting stimuli (Equation 2).

$$SelA = \frac{CC}{IC + PP + DC} \quad (2)$$

The SelA acquires the highest value when the number of CC is equal to the number of PP and where no incorrect clicks (IC) and clicks on the distraction point (DC) were performed. The OP was defined as the ratio between SelA and MRT, where a value of 1 account for the highest performance and 0 accounts for the lowest performance (Marandi et al., 2018A).

Subjective ratings

Overall weighted ratings were computed for each task completed by each participant. A weighting of the six different parameters (mental demand, physical demand, temporal demand, performance, effort, frustration) were computed based on the pairwise comparisons of the parameters after the familiarization period. A total weighted rating was computed for each MWL task.

Physiological measures

Both ECG and SCG were included as physiological measures of MWL, and the general steps involved in the data analysis has been illustrated in Figure 4. These steps include segmentation of the ECG signal, segmentation of the SCG signal, HRV analysis, agreement between ECG and SCG HRV measures, statistical testing and machine learning based on SCG signals.

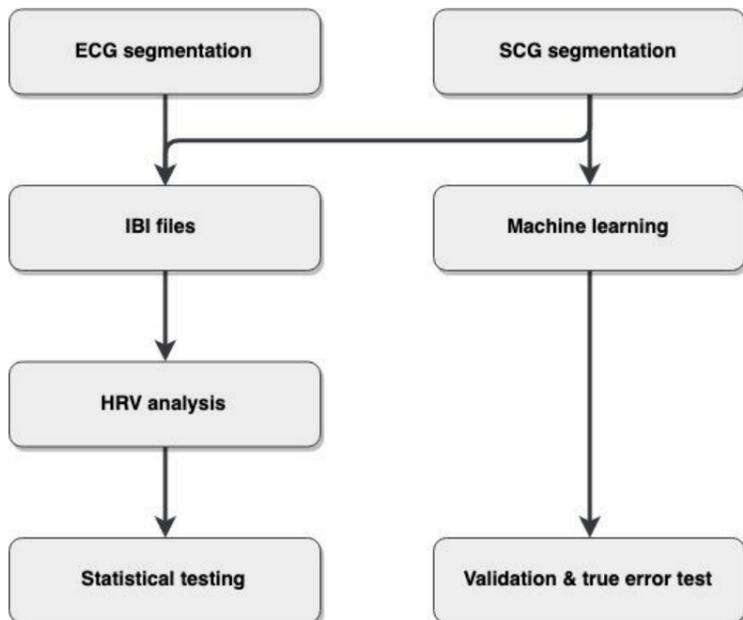


Figure 4: General steps involved in the data analysis of physiological measures

ECG cardiac cycle segmentation

For the analysis of ECG based HRV, a MatLab script was constructed to extract ECG based inter-beat intervals (IBI). Firstly, raw lead II data was extracted from the data matrix and structured based on the ID of each participant, test day and load level. Secondly, the MatLab toolbox BioSigKit (Sedghamiz, 2018) was implemented, using the Pan-Tompkins algorithm (Pan & Tompkins, 1985) for R peak detection. The ultimate end product is a text file consisting of IBI used for the HRV analysis.

SCG cardiac cycle segmentation

For the purpose of running HRV analysis based on SCG, another MatLab script was constructed, involving several steps as illustrated in Figure 5.

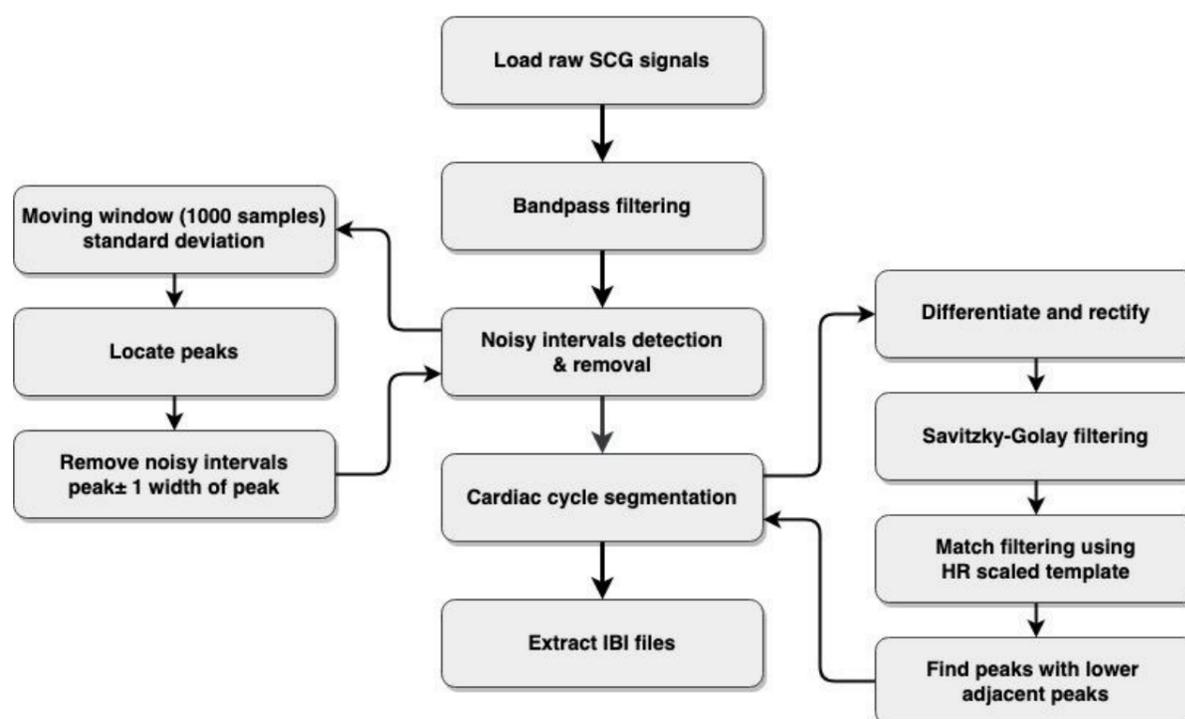


Figure 5: Processing steps involved in the SCG segmentation and creation of IBI files

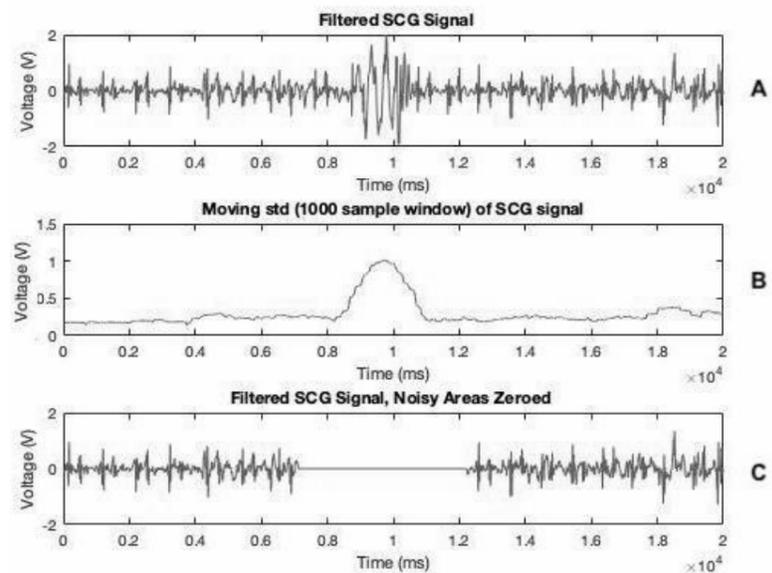


Figure 6: Steps involved in the automatic noise detection/cancellation of SCG signal. A) filtered SCG signal, B) moving std with a 1000 sample window, C) filtered SCG signal with zeroed noisy areas.

The goal was to automatically detect and remove noisy areas of the signal, and to segment individual cardiac cycles of the SCG signal without the use of the concurrent ECG measurements.

For the auto-detection and cancellation of noisy areas in the SCG signal (Figure 6, A) a moving standard deviation was calculated (window length=1000 samples) (Figure 6, B). The locations and widths of the noisy areas peaks were found using a minimum peak prominence of 0.3V, and the SCG was zeroed out around these peaks (± 1 width of the peak) (Figure 6, C) for later data analysis.

For the cardiac cycle segmentation, SCG data was filtered using a 4th order band pass Butterworth filter ($f_{c,low} = 0.5\text{Hz}$, $f_{c,high} = 80\text{Hz}$). The SCG signal was then differentiated, rectified and filtered using a 2nd order Savitzky-Golay filter with a window length of 101 samples for segmentation purposes (Figure 7, A,B,C).

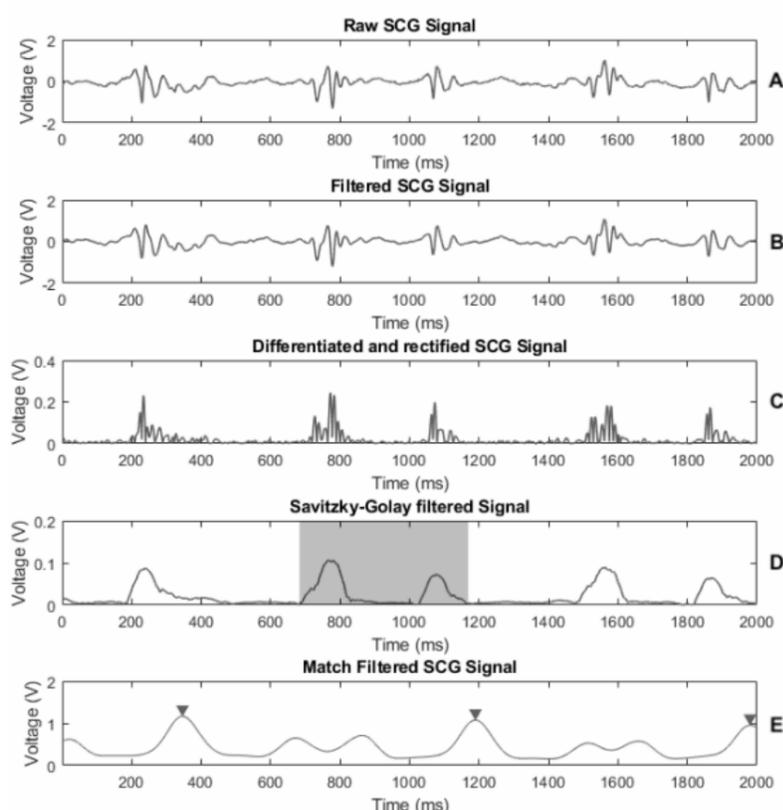


Figure 7: Steps involved in the automatic cardiac cycle segmentation of SCG signals. A) raw SCG signal, B) filtered SCG signal, C) differentiated and rectified SCG signal with template illustrated by grey area, D) Savitzky-Golay filtered signal, E) match filtered SCG based on template with peaks located.

Next, a template was defined from the Savitzky-Golay filtered SCG signal (Figure 7 D, indicated by greyed area). The differentiated, rectified and Savitzky-Golay filtered SCG signal was then match filtered using the time inverted and heart rate scaled template, resulting in Figure 7, E, where peaks with lower adjacent peaks were located, resulting in cardiac cycle segmented SCG signals. The auto segmented SCG based IBI were validated against ECG based IBI using a Bland-Altman plot, based on data with no noise removal from 2 random participants.

HRV analysis

The IBI files were preprocessed to remove ectopic beats, resulting in the computation of Normal-to-Normal intervals (NN-intervals) detected by a threshold of 3 standard deviations, as pure removal of IBI has been proved to perform superior to e.g.

linear and cubic spline interpolation (Lippman et al., 1994). The NN-intervals was furthermore detrended for low frequency trends, using the Wavelet Packet method. After preprocessing, the signal was analyzed in both the time- and frequency domain. The HRV time domain measures included were: average heart rate (MeanHR), standard deviation of NN intervals (SDNN), root mean square of successive NN interval differences (RMSSD), baseline width of the NN interval histogram (TINN), standard deviation of heart rate (sdHR), and the integral of the density of the NN interval histogram divided by its height (HRVTI). The HRV frequency domain measures was found using the Lomb Scargle method and included: the absolute power of the low-frequency (0.04–0.15 Hz) band (aLF), absolute power of the high-frequency (0.15–0.4 Hz) band (aHF), absolute power of all frequency bands (aTotal), percentage of the sum of aLF and aHF for the low frequency band (pLF), percentage of the sum of aLF and aHF for the high frequency band (pHF), the ratio of LF-to-HF power (LF/HF ratio), peak frequency of the low-frequency band (Peak LF), peak frequency of the high-frequency band (Peak HF).

Statistical testing of HRV analyses

The effect of MWL levels, and testing day, on HRV measures were statistically tested in SPSS, using a two-way ANOVA with repeated measures, followed by a pairwise comparison for load levels, using the Bonferroni correction, with a significance level of $\alpha = 0.05$. This includes the performance measures obtained by the mental task in MatLab, subjective ratings from the NASA-TLX, as well as time and frequency domain measures from both the ECG and SCG based HRV analysis. Mauchly's test of sphericity was implemented, and if violated, corrected for, using the Greenhouse Geisser correction.

The agreement between SCG and ECG based HRV analysis was assessed using the Intraclass Correlation Coefficient (ICC), with a mixed effects ICC with single measures, ICC(3,1). The ICC values were interpreted based on the following guideline adopted from Koo & Li, 2015; $ICC < 0.5$ are indicative of poor agreement, $0.5 < ICC < 0.75$ are indicative of moderate agreement, $0.75 < ICC < 0.9$ are indicative of good agreement and $ICC > 0.9$ are indicative of excellent agreement. ICC values were calculated for all HRV measures, used in the statistical tests.

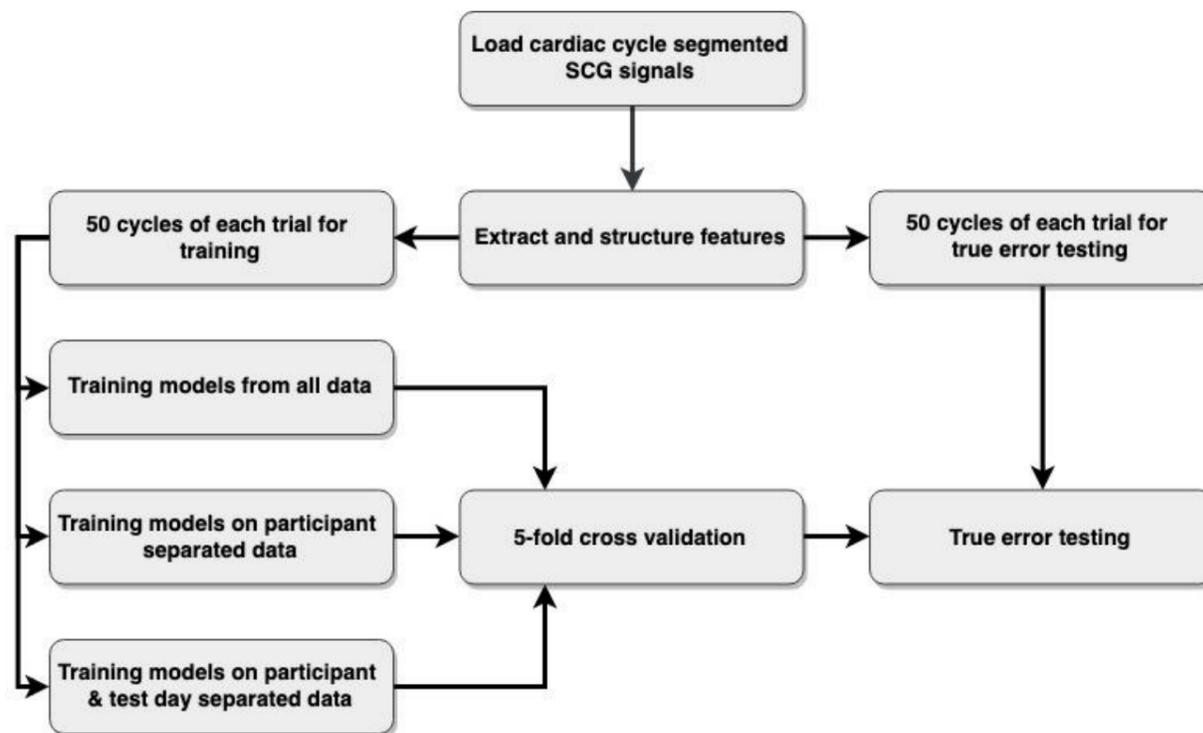


Figure 8: Schematic overview of the steps included to classify MWL levels using machine learning.

Classification of SCG signals

Machine learning was implemented to evaluate the ability to discriminate between MWL levels based on the segmented SCG signals as illustrated in Figure 8. Three different approaches to this classification process was selected, meaning that the classification algorithms were trained on the entire dataset for one approach, while the other two approaches were divided into participant and participant on the respective test day respectively. Different parts of the data were allocated for training and test purposes where 50 cardiac cycles were included from each MWL level, on each respective day, for each participant. This means that each approach to the classification process, all together,

participant separated, and participant and test day separated, contained different sample sizes being 3600, 300 and 150 samples respectively, for each trained model.

A total of 22 features were computed for each SCG segmented cardiac cycle including: mean, standard deviation, integral, median, variance, range, skewness, kurtosis, length, RMS, systolic max, location of systolic max, diastolic max, location of diastolic max, time from systolic max to diastolic max, first systolic min occurring before systolic max, location of systolic min, first diastolic min occurring before diastolic max, location of diastolic min, time from systolic min to diastolic min, time from systolic min to max, and time from diastolic min to max.

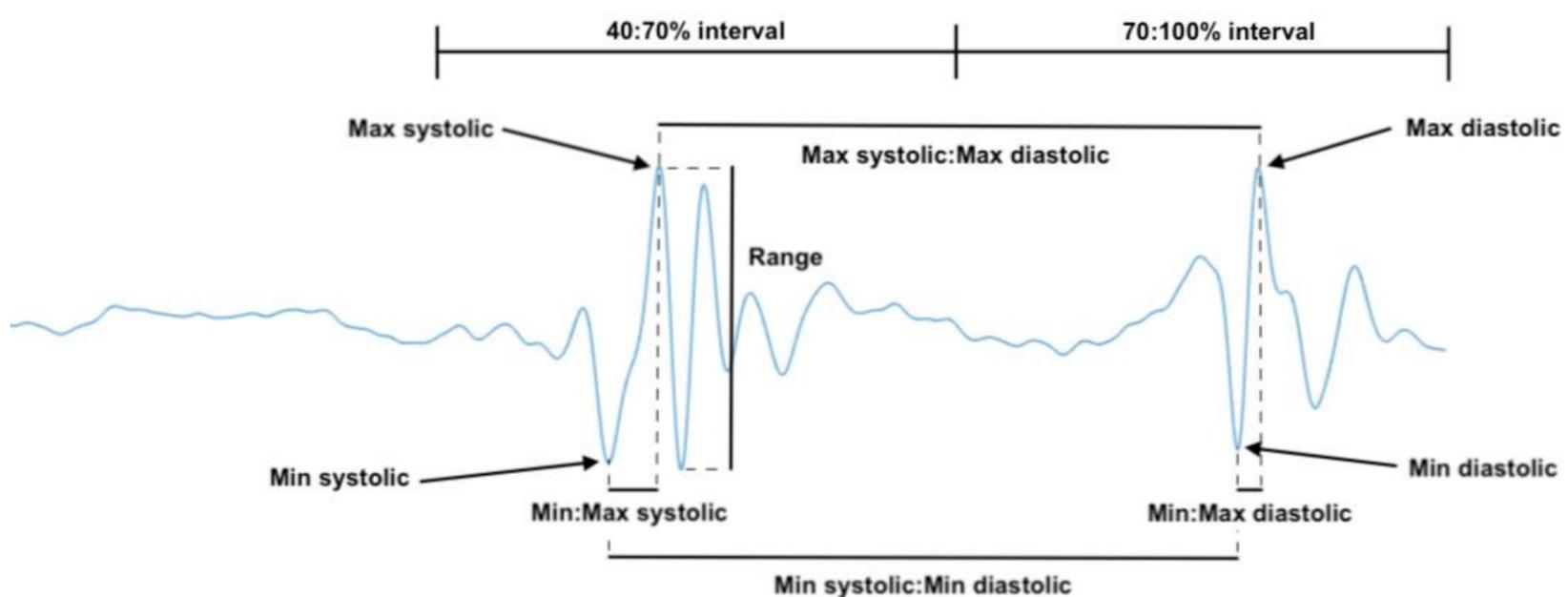


Figure 9: Illustration of certain features of the SCG signal for one cardiac cycle.

The systolic and diastolic features were found by searching within 40-70% and 70-100% of the signal respectively. Some selected features have been visualized in Figure 9. All features were rescaled to obtain values in the range 0-1. A principal component analysis was performed in order to reduce the dimensionality of the features while retaining any potential inter-class variation (Dougherty, 2013). The number of principal components included for further analysis was based on that the principal components comprised 95% of the total variance (Jackson & Donald, 1993). Decision Trees and K-Nearest Neighbor were chosen as the included classification algorithms and trained to predict the MWL level. The algorithms were validated using a 5-fold cross validation and the

superior variant of each classification algorithm were chosen for estimating the true error based on the testing data set.

Results

Table 1 provides an overview of the measures (ECG and SCG based HRV, performance measure, subjective rating) computed to assess MWL. The statistical analysis reveal that significant differences were found for ECG based Peak LF, SCG based aHF, aTotal and Peak LF, performance, and subjective rating. The appurtenant results of the two-way ANOVA with repeated measures and the post hoc test have been presented in Table 2.

		Day 1			Day 2		
		Low	Medium	High	Low	Medium	High
ECG Based HRV Analysis	MeanHR (bpm)	65.40±11.20	66.97±11.21	65.86±11.02	68.62±9.71	69.22±10.10	69.37±11.09
	HRVTI (ms)	10.85±2.17	11.85±1.89	11.36±1.68	11.55±1.27	11.16±1.96	12.11±1.60
	RMSSD (ms)	52.62±20.39	51.98±20.24	51.87±19.08	52.58±27.22	51.87±19.08	47.87±21.63
	TINN (ms)	250.6±104.9	273.9±97.0	263.8±80.8	271.0±155.8	272.0±125.3	267.7±93.7
	sdHR (bpm)	4.927±1.942	5.436±1.956	5.573±2.067	4.955±1.782	5.491±1.653	5.155±1.204
	SDNN (ms)	70.80±29.27	74.09±25.39	75.96±22.90	66.89±30.44	71.22±26.97	67.93±25.00
	aLF (ms ²)	0.008±0.002	0.008±0.002	0.008±0.003	0.009±0.004	0.008±0.004	0.007±0.004
	aHF (ms ²)	0.003±0.002	0.003±0.002	0.005±0.004	0.003±0.001	0.003±0.002	0.002±0.002
	aTotal (ms ²)	0.011±0.003	0.012±0.004	0.013±0.007	0.012±0.005	0.011±0.006	0.009±0.006
	pLF (%)	71.86±12.93	70.57±8.87	69.21±11.91	72.46±9.09	74.65±6.62	76.64±8.17
	pHF (%)	27.36±12.97	27.25±9.88	29.93±11.95	26.43±8.99	24.11±6.55	22.21±8.01
	LF/HF ratio	3.302±1.722	3.041±1.426	2.913±1.863	3.218±1.604	3.360±1.080	4.071±2.126
	Peak LF (Hz) *	0.120±0.000	0.100±0.000	0.089±0.003	0.120±0.000	0.100±0.000	0.090±0.000
	Peak HF (Hz)	0.259±0.068	0.216±0.053	0.245±0.069	0.240±0.064	0.241±0.073	0.228±0.071
SCG Based HRV Analysis	MeanHR (bpm)	65.19±11.27	66.72±11.22	65.46±10.95	68.54±9.79	68.97±10.11	69.26±11.16
	HRVTI (ms)	11.02±2.00	12.15±2.94	10.79±1.97	10.78±1.13	10.72±1.87	11.81±2.40
	RMSSD (ms)	53.94±21.42	52.24±19.05	53.88±19.25	53.79±27.72	53.88±19.25	48.83±20.34
	TINN (ms)	253.8±118.2	267.3±110.4	237.1±81.9	254.6±124.9	257.0±110.4	267.0±122.8
	sdHR (bpm)	4.827±1.941	5.200±1.972	5.282±1.723	4.946±1.949	5.418±1.669	5.009±1.319
	SDNN (ms)	69.60±28.85	71.29±24.50	73.27±21.38	66.39±31.52	70.71±26.85	65.86±24.73
	aLF (ms ²)	0.011±0.003	0.015±0.005	0.016±0.009	0.012±0.006	0.012±0.006	0.013±0.007
	aHF (ms ²) *	0.005±0.003	0.007±0.005	0.009±0.007	0.005±0.004	0.005±0.003	0.004±0.003
	aTotal (ms ²) *	0.016±0.005	0.022±0.009	0.025±0.015	0.018±0.009	0.016±0.009	0.017±0.010
	pLF (%)	70.62±14.43	68.23±9.94	65.95±12.86	70.68±9.59	73.49±7.93	74.59±7.24
	pHF (%)	28.83±14.50	29.30±11.77	33.14±12.33	28.66±9.79	25.52±7.48	24.28±7.14
	LF/HF ratio	3.192±1.812	2.928±1.757	2.456±1.610	2.868±1.359	3.153±1.043	3.347±1.090
	Peak LF (Hz) *	0.120±0.000	0.100±0.006	0.090±0.006	0.119±0.005	0.100±0.004	0.092±0.006
	Peak HF (Hz)	0.266±0.064	0.230±0.044	0.219±0.053	0.234±0.067	0.226±0.071	0.243±0.077
Performance score *	0.512±0.094	0.467±0.090	0.417±0.101	0.560±0.093	0.505±0.107	0.466±0.110	
Nasa TL-X rating *	53.000±17.228	63.212±11.278	69.879±9.347	46.879±14.102	53.091±18.815	64.394±12.704	

Table 1: ECG and SCG based HRV results, performance measures and subjective ratings in mean±1std. Statistically significant differences between one or more groups/days, are marked in bold and with a *.

	Two-way ANOVA, repeated measures				Pairwise comparison		
	Day	Load level	Day*Load level	Low - Med.	Med. - High	Low - High	
ECG Peak LF (Hz) *	$F_{1,10} = 1.0$ $p = 0.341$	$F_{1,10} = 3477$ $p = 0.000^*$	$F_{1,10} = 1.0$ $p = 0.341$	$p = 0.000^*$	$p = 0.000^*$	$p = 0.000^*$	
SCG aHF (ms ²) *	$F_{1,10} = 7.5$ $p = 0.021^*$	$F_{2,20} = 1.7$ $p = 0.211$	$F_{2,20} = 6.7$ $p = 0.006^*$	-	-	-	
SCG aTotal (ms ²) *	$F_{1,10} = 9.2$ $p = 0.013^*$	$F_{2,20} = 1.7$ $p = 0.217$	$F_{2,20} = 3.8$ $p = 0.039^*$	-	-	-	
SCG Peak LF (Hz) *	$F_{1,10} = 0.1$ $p = 0.756$	$F_{1,3,12,7} = 111.4$ $p = 0.000^*$	$F_{2,20} = 0.4$ $p = 0.698$	$p = 0.000^*$	$p = 0.000^*$	$p = 0.000^*$	
Performance score *	$F_{1,11} = 8.6$ $p = 0.014^*$	$F_{2,22} = 34.6$ $p = 0.000^*$	$F_{2,22} = 0.3$ $p = 0.713$	$p = 0.013^*$	$p = 0.001^*$	$p = 0.000^*$	
Nasa TL-X rating *	$F_{1,10} = 9.1$ $p = 0.013^*$	$F_{2,20} = 26$ $p = 0.000^*$	$F_{2,20} = 0.8$ $p = 0.449$	$p = 0.011^*$	$p = 0.007^*$	$p = 0.000^*$	

Table 2: F and p values of two-way ANOVA with repeated measures tests, with pairwise comparison (Bonferroni) for load levels, for parameters with significant differences. Bold font and * indicates significant difference.

These results reveal that ECG and SCG based Peak LF and performance score significantly decreased with increments of MWL, the subjective ratings significantly increased with increments of MWL. SCG based aHF and aTotal and subjective rating significantly decreased while performance score significantly increased from the first to the second test day. Interactions within significant different measures were found for SCG based aHF and aTotal. The computed ICC ($\pm 95\%$ CI) between ECG and SCG based HRV measures has been summarized in Table 3. All time domain features show good to excellent correlation, except for HRVTI which showed moderate to good agreement, while the frequency domain features ranged from poor to excellent agreement (Table 3). The validation of the cardiac cycle auto segmentation is presented as a Bland-Altman plot between ECG and SCG based IBI (Figure 10). The bias was found to be 0.006ms, upper- and lower limits ± 28 ms and no trend was discovered, representing a good agreement between methods when visually inspecting the location of datapoints in the plot.

SCG & ECG agreement		
Variable	ICC correlation	95% CI
MeanHR (bpm)	0,999	0.998-1
HRVTI (ms)	0,684	0.532-0.794
RMSSD (ms)	0,986	0.975-0.992
TINN (ms)	0,919	0.869-0.950
sdHR (bpm)	0,978	0.958-0.988
SDNN (ms)	0,987	0.975-0.993
aLF (ms ²)	0,226	-0.036-0.458
aHF (ms ²)	0,347	0.047-0.574
aTotal (ms ²)	0,210	-0.038-0.436
pLF (%)	0,901	0.807-0.945
pHF (%)	0,903	0.803-0.948
LF/HF ratio	0,814	0.695-0.887
Peak LF (Hz)	0,927	0.884-0.955
Peak HF (Hz)	0,670	0.511-0.784

Table 3: Two-way mixed single measures Intraclass Correlation Coefficient (ICC) measuring absolute agreement, between ECG- & SCG based HRV measures.

The cross-validation accuracies and true error of the two included algorithms classifying the three different levels of MWL has been presented in Table 4. This analysis generally showed poor classification accuracy, while a tendency towards slightly higher accuracies for the participant and day specific approaches occurred.

The most superior classification models associated to each different algorithm were coarse decision tree (4 splits) and a medium KNN (10 neighbors).

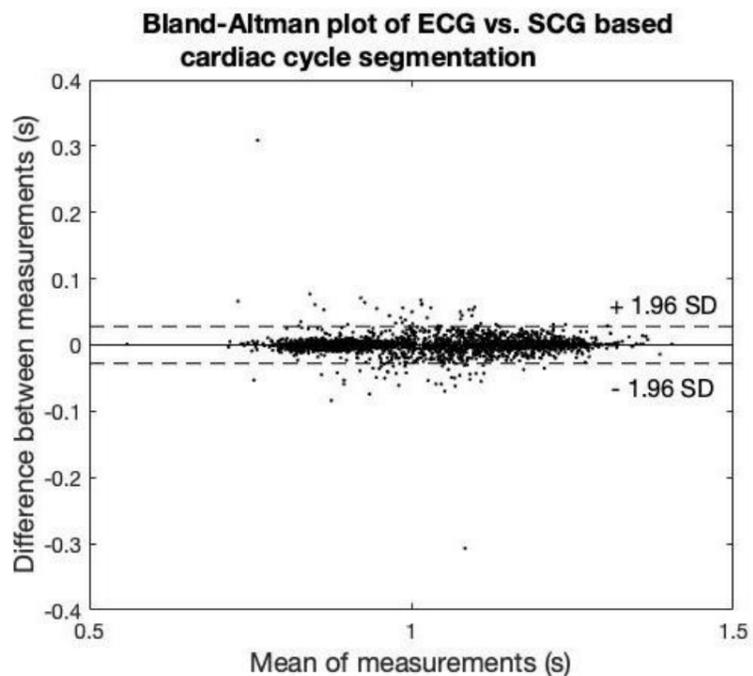


Figure 10: Bland-Altman plot illustrating agreement between ECG and SCG based IBI, $n=2750$.

Cross validation accuracy		
Data sample	Coarse decision tree	Medium KNN
Single day	48.1%	47.9%
Participant	45.4%	46.1%
All	34.5%	37.3%

True error		
Data sample	Coarse decision tree	Medium KNN
Single day	37.3%	38.8%
Participant	36.8%	37.1%
All	35.4%	34.9%

Table 4: Classification accuracies in percentage (%) from the initial 5-fold cross validation, and true error testing of trained models fed with new data

Discussion

Using SCG and the proposed auto noise detection and segmentation as an alternative and easier to use tool for HRV analysis, showed to be feasible, as well as providing almost similar results compared to the ECG based HRV according to Table 1, 2 and 3. This was expected as Tadi et al., 2015 reported almost identical HRV measures based on both ECG and SCG (Tadi et al., 2015). The SCG signal are however prone to noise, which was also observed in the present study to a varying degree. This ultimately

results in different input data for the HRV analysis, as only ectopic beats were removed from the ECG based signal, while larger periods of data were removed from the SCG based signal, due to noise. This might inflict the HRV results as the editing of NN intervals has been reported to change the results of the HRV analysis (Peltola, 2012). The HRV frequency domain measures are predominantly prone to removal of data and especially increases the power of both the LF and HF band (Salo et al., 2001). This accord well with the findings of the present study, as the ICC values of the HRV time domain features generally showed excellent agreement, except for HRVTI (moderate to good) and TINN (good to excellent), while the ICC values of the HRV frequency domain features varied between poor and excellent agreement (Table 3). The absolute power of the frequencies showed poor agreement, as suggested by Salo et al., 2001. Thereby caution should be exercised when assessing MWL from SCG based HRV as some frequency domain measures might be biased due to the removal of noisy areas.

Even though removal of IBI has been found to perform well for HRV analysis (Lippman et al., 1994), Salo et al., 2001 reported that interpolation proved to be superior to the removal of data, when dealing with larger periods of data that needs to be corrected for (Salo et al., 2001). This might indicate that interpolation of missing IBI should be considered for SCG based HRV analysis.

ECG and SCG based HRV analysis

Significant differences between MWL levels were found for both the ECG- and SCG based HRV analysis in the Peak LF (Table 2). This indicates that MWL can be distinguished by HRV analysis and the Peak LF parameter, based on either ECG or SCG signals. However, a review concerning HRV analysis of MWL using ECG, reports that other parameters within the HRV analysis have showed significant differences between MWL levels (Kim et al, 2018). The lack of these differences in the current study, could however be due to the mental task used in this study, and each participant's individual level of mental capability resulting in either an under- or overload. Also, the difference in difficulty between low, medium and high load level could have been too low to make a significant difference on other HRV parameters. This is supported by Charles & Nixon, 2019, as the different HRV measures and their link to MWL is dependent on the methodologies used to induce the MWL (Charles & Nixon, 2019).

Machine learning and SCG signals

The classification of MWL levels can be perceived to be unsuccessful, given the fact that pure guessing would have resulted in almost similar accuracies (Table 4). It can also be observed from Table 4 that the trained model based on all data performed worse, than the participant- and participant/day specific ones. This is most likely due to physiological differences between participants, which ultimately causes differences in the data, not caused by changing levels of MWL. For example, BMI affects the amplitude of aortic valve opening signal characteristic, as described by Sørensen et al., 2017, which corresponds to one of the present study's features: the first max value in the 40-70% interval of each cycle. Also, differences in arousal level between days, due to external factors, might be a contributing factor to the better accuracies found in the participant and day specific models compared to the participant specific models. This is supported by Lee et al., 2017 as different HR estimation accuracies from SCG, in different aroused situations, were reported (Lee et al., 2017). This indicates that the SCG signals morphology changes in regard to the level of arousal, which could complicate the segmentation of cardiac cycles and possibly also the distinction between different MWL levels.

When doing machine learning, variations between classes is needed in order to distinguish between these. However, when classifying SCG signals from different MWL levels, these variations between classes might drown in higher variations caused by coincidences or physiological differences (Sørensen et al., 2018). A previous study by Javaid et al., 2016 has shown that posture also has a big influence on the shape and frequency content of SCG signals, as *“posture can (1) distort the SCG signal, for example due to altering the body's mechanical vibration response, and (2) affect a person's cardiovascular physiology, for example due to changes in venous return.”*- Javaid et al., 2016, pp. 1. Although general posture during the experimental trials were fixed, leg extension could have influenced the venous return, and the torso angle and amount of contact with the chair, could have affected the body's vibration response, thereby resulting in differences in the SCG signal.

Previous studies examining classification of different MWL levels have reported relatively high classification accuracies in the range of 80-95%. These studies however included several different measure modalities such as ECG, electroencephalography, eye tracking and skin conductance, and achieved much lower classification accuracies when the models were not trained on data containing electroencephalographic measures (Hogervorst et al., 2014, Wilson & Russel, 2003). This indicates that the result of the classification of MWL levels in the present study (Table 4) might be in accordance with other cardiac based classification models. For future work of classifying MWL from cardiac cycle segmented SCG signals, it is proposed that more features are extracted for the machine learning, and that these potentially involves more fiducial points and frequency domain measures, to further evaluate the feasibility of this methodology. Furthermore, the features should be corrected for posture differences, as suggested by Javaid et al., 2016, as accelerometers can concurrently determine posture to a certain degree (Lugade et al., 2015).

Auto noise detection and cardiac cycle segmentation of SCG signals

For the purpose of moving towards a fully automatic MWL level classification algorithm, an automatic cardiac cycle segmentation algorithm, using matched filtering, was developed. This furthermore limited time, resources and expert knowledge needed for segmentation, as all SCG signals should not have to be manually annotated. This methodology was inspired by the work of Li et al., 2015, concerning automatic segmentation of SCG signals, but for the present study, the procedure was changed significantly to make the algorithm more robust to SCG signals where fluctuations, differences in amplitude and number of peaks, occurred. As described by Sørensen et al., 2018, SCG signals are not similar from beat to beat, and from participant to participant. Likewise, the SCG signals obtained in the present study showed high variance in regard to amplitude of peaks, number of peaks and number of fluctuations in the systolic and diastolic complex. By combining the methodology of matched filtering and the knowledge of differences in SCG signals, this method was able to auto segment SCG signals with higher variances in the signal morphology.

Furthermore, the algorithm for automatically detecting noisy areas in SCG signal was successfully implemented into the data analysis, as manual inspection showed successful removal of noisy intervals.

Since the template used for the matched filtering does not cover the same area of the signal as the ECG based IBI, a slight difference in IBIs were found as expected. However, a good agreement, between the proposed automatic cardiac cycle segmentation of SCG signals and the traditional Pan-Tompkins algorithm for segmenting ECG signals, was found in the Bland-Altman plot for concurrent measurements of ECG and SCG based NN-intervals (Figure 10).

WAME 1.0 mental computer task

As discussed earlier, the load level differentiation might have been the reason for the lack of multiple HRV measures being significant different between load levels, as well as a contributing factor to the poor accuracies of the classification models (Table 4). Even though the perceived MWL (based on the NASA-TLX) and performance measures were significant different between load levels (Table 2), these differences might not have been high enough to be physiologically distinguished. The parameters of each level could potentially be differentiated even more, for example by adjusting the completion time available or the length of memorization period, instead of only changing the complexity of the patterns and thereby the task. It is also noticeable that a learning effect have occurred, as both the performance measure and subjective rating show to be significantly different between the two test days (Table 1 & 2). This is however not conflicting with the purpose of the study, as only differentiation between the different MWL levels on both days was of interest.

Significant differences between load levels were however found in the present study, in both performance measures, subjective ratings and the Peak LF parameter from the HRV analysis (Table 2). Combined with the significant differences in heart rate measures found by Marandi et al., 2018B, using the same mental task, the WAME 1.0 mental task seems appropriate in terms of inducing different amounts of MWL. However, higher classification accuracies and multiple significant different HRV measures, could potentially be obtained by making minor adjustments to the settings of each load level in the mental computer task.

Future perspectives

Based on the findings of the present study, it is most likely that, with the implementation of the previously discussed alterations and modifications to this methodology, SCG measurements could be implemented in future methodologies for monitoring MWL. When compared to ECG systems, SCG is already easier to implement, since no electrode placement or skin preparation is needed. If the SCG could be obtained by smartphones, this would further ease the process. With the fast development of smartphones and the sensors within, it would be possible to assess MWL without the need for external equipment or expert consulting. Users would simply have to place their smartphone on the sternum, and the phone would utilize its onboard accelerometer to obtain the SCG signals. The use of smartphones for obtaining SCG signals has already shown to be feasible by several studies (Tadi et al., 2016; Landreani et al., 2016).

Accelerometers attached to humans also show potential to be applied for other purposes than heart rate quantification. This includes determination of posture and different movements as reported by Lugade et al., 2015, which could provide feedback associated to ergonomics and physical activity (Attal et al., 2015). With the many applications of accelerometers, the technology could implement multiple applications, making for a more combined and holistic health- and performance monitoring system.

Conclusion

The results of the current study suggest that the traditional approach to assess MWL by HRV analysis is feasible with the implementation of SCG instead of the more setup intensive ECG system. This is however only pertinent if interpolation can effectively deal with the exclusion of data due to noise, as missing data leads to biased HRV results especially in the frequency domain, where several associations with MWL has been reported. The classification of MWL based on the SCG signal showed to be insufficient and no additional association between MWL and SCG based features was identified. Future studies could however examine additional features of the signal in both time and frequency domain to achieve this.

References

- APA, n.d., *Stress effects on the body*. American Psychological Association
<https://www.apa.org/helpcenter/stress-body>
Last seen 18/02/2019
- Attal, Ferhat; Mohammed, Samer; Dedabrishvili, Mariam; Chamroukhi, Faicel; Oukhellou, Latifa; Amirat, Yacine, 2015. *Physical Human Activity Recognition Using Wearable Sensors*, Sensors, Vol. 15, pp. 31314-31338.
- Bányai, Fanni; Griffiths, Mark D.; Király, Orsolya; Demetrovics, Zsolt, 2018. *The Psychology of Esports: A Systematic Literature Review*, Journal of Gambling Studies, Vol. 35, Iss. 2, pp. 351-365.
- Blitz, P. S.; Hoogstraten, J.; Mulder, G., 1970. *Mental Load, Heart Rate and Heart Rate Variability*, Psychol. Forsch., Vol. 33, pp. 277-288.
- Charles, Rebecca L.; Nixon, Jim, 2019. *Measuring mental workload using physiological measures: A systematic review*, Applied Ergonomics, Vol. 79, pp. 221-232.
- Hogervorst, Maarten A.; Brouwer, Anne-Marie; van Erp, Jan B. F., 2014. *Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload*, Frontiers in Neuroscience, Vol. 8, Art. 322.
- Holmes, S., 2001. *Work-related stress: a brief review*, The Journal of The Royal Society for the Promotion of Health, Vol. 121, No. 4, pp. 230-235.

- Jackson, Donald A., 1993. *Stopping Rules in Principal Component Analysis: A Comparison of Heuristical and Statistical Approaches*, Ecology, Vol. 74, No. 8, pp. 2204-2214.
- Javaid, Abdul Q.; Dowling, Sean; Etemadi, Mozziyar; Heller, J. Alex.; Roy, Shuvo; Klein, Liviu, Inan, Omer T., 2016, *Quantification of posture induced changes in wearable seismocardiogram signals for heart failure patients*, 2016 Computing in Cardiology Conference (CinC), Vancouver, BC, pp. 777-780., doi: 10.23919/CIC.2016.7868858
- Kim, H. G.; Cheon, E. J.; Bai, D. S.; Lee, Y. H.; Koo, B. H., 2018. *Stress and Heart Rate Variability: A Meta-Analysis and Review of the Literature*, Psychiatry Investigation, Vol. 15(3), 235–245. doi:10.30773/pi.2017.08.17
- Koo, Terry K.; Li, Mae Y., 2015. *A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research*, Journal of Chiropractic Medicine, Vol. 15, pp. 155-163.
- Landreani, F.; Martin-Yebra, A.; Casellato, C.; Frigo, C.; Pavan, E.; Migeotte, P-F.; Caiani, EG., 2016. *Beat-to-beat Heart Rate detection by Smartphone's Accelerometers: validation with ECG*, 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 525–528.
- Lee, Hyunwoo; Lee, Hana; Whan, Mincheol, 2017. *An Enhanced Method to Estimate Heart Rate from Seismocardiography via Ensemble Averaging of Body Movements at Six Degrees of Freedom*, Sensors, Vol. 18, Art. 238.
- Li, Yanjun; Tang, Xiaoying; Xu, Zhi, 2015. *An approach of heartbeat segmentation in Seismocardiogram by matched-filtering*, Proc. 7th Int. Conf. on Intelligent Human-Machine Systems and Cybernetics, Vol. 2, pp 47–51.
- Lippman, N.; Stein, K.M.; Lerman, B.B., 1994. *Comparison of methods for removal of ectopy in measurement of heart rate variability*, The American Journal of Physiology, Vol. 267, pp. H411-H418.
- Lugade, Vipul; Fortune, Emma; Morrow, Melissa; Kaufman, Kenton, 2015. *Validity of Using Tri-Axial Accelerometers to Measure Human Movement - Part I: Posture and Movement Detection*, Medical Engineering & Physics, Vol. 36, Iss. 2, pp. 169-176.
- Marandi, Ramtin Z.; Madeleine, Pascal; Omland, Øyvind; Vuillerme, Nicolas; Samani, Afshin, 2018A, *Reliability of Oculometrics During a Mentally Demanding Task in Young and Old Adults*, IEEE Access, Vol. 6, pp. 17500-17517.
- Marandi, R. Z.; Madeleine, P.; Vuillerme, N.; Samani, A., 2018B, *Heart Rate Monitoring for the Detection of Changes in Mental Demands During Computer Work*, World Congress on Medical Physics and Biomedical Engineering 2018, IFMBE Proceedings 68/2, https://doi.org/10.1007/978-981-10-9038-7_69
- Mariotti, A., 2015. *The effects of chronic stress on health: new insights into the molecular mechanisms of brain-body communication*, Future Sci. OA (2015) 1(3), FSO23 eISSN 2056-5623
- Mounsey, P. 1956. *Præcordial ballistocardiography*, British heart journal, Vol. 19(2), pp. 259–271. doi:10.1136/hrt.19.2.259
- Paukkunen, Mikko, 2014. *Seismocardiography: Practical implementation and feasibility* (Doctoral dissertation), Aalto University, Department of Electrical Engineering and Automation, Finland.
- Pan, J.; Tompkins, W., 1985, *A Real-Time QRS Detection Algorithm*, IEEE Transactions on Biomedical Engineering, Vol. BME-32, No. 3, pp. 230-236.
- Peltola, Mirja A., 2012. *Role of Editing of R-R Intervals in the Analysis of Heart Rate Variability*, Frontiers in Physiology, Vol. 3, Art. 148.
- Salo, Mirja A.; Huikuri, Heikki V.; Seppänen, Tapio, 2001. *Ectopic Beats in Heart Rate Variability Analysis: Effects of Editing on Time and Frequency Domain Measures*, Ann. Noninvasive Electrocardiol., Vol. 6, pp. 5-17.
- Sedghamiz, 2018. *BioSigKit: A Matlab Toolbox and Interface for Analysis of BioSignals*, Journal of Open Source Software, 3(30), 671, <https://doi.org/10.21105/joss.00671>
- Slavich, G. M., 2016. *Life Stress and Health: A Review of Conceptual Issues and Recent Findings*, Teaching of Psychology 2016, Vol. 43(4), pp. 346-355
- Stansfeld, Stephen; Candy, Bridget, 2006. *Psychosocial work environment and mental health - a meta-analytic review*, Environment & Health, Vol. 32, No. 6, pp. 443-462.

- Sørensen, K.; Jensen, A. S.; Hansen, J.; Søgaard, P.; Struijk, J. J.; Schmidt, S., 2017. *Seismocardiographic correlations to age, gender and BMI*, Paper ThAT5.4. 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, EMBC 2017, Jeju Island, Sydkorea.
- Sørensen, K.; Schmidt, S.E.; Jensen, A.S.; Søgaard, P.; Struijk, J.J., 2018. *Definition of Fiducial Points in the Normal Seismocardiogram*, Scientific Reports, Volume 8, Art.15455.
- Tadi, Mojtaba Jafari; Lehtonen, Eero; Koivisto, Tero; Pänkäälä, Mikko; Paasio, Ari; Teräs, Mika, 2015. *Seismocardiography: Toward heart rate variability (HRV) estimation*, 2015 IEEE International Symposium on Medical Measurements and Applications (MeMeA) Proceedings, pp. 261-266.
- Tadi, Mojtaba Jafari; Lehtonen, Eero; Hurnanen, Tero; Koskinen, Juho; Eriksson, Jonas; Pänkäälä, Mikko; Teräs, Mika; Koivisto, Tero, 2016. *A real-time approach for heart rate monitoring using a Hilbert transform in seismocardiograms*, Physiological Measurement, Vol. 37, pp. 1885-1909.
- Taebi, Amirtaha; Solar, Brian E.; Bomar, Andrew J.; Sandler, Richard H.; Mansy, Hansen A., 2019. *Recent Advances in Seismocardiography*, Vibration, Vol. 2, pp. 64-86.
- Taelman, Joachim; Vandeput, Steven; Vlemincx, Elke; Spaepen, Arthur; Huffel, Sabine Van, 2011. *Instantaneous changes in heart rate regulation due to mental load in simulated office work*, European Journal of Applied Physiology, Vol. 111, pp. 1497-1505.
- Van Daalen, G.; Willemsen, TM.; Sanders, K.; van Veldhoven, MJ., 2009. *Emotional exhaustion and mental health problems among employees doing "people work": The impact of job demands, job resources and family-to-work conflict*, Int Arch Occup Environ Health., Vol. 82(3), pp. 291–303. doi: 10.1007/s00420-008-0334-0.
- Wilson, Glenn F.; Russel, Cristopher A., 2003. *Real-Time Assessment of Mental Workload Using Psychophysiological Measures and Artificial Neural Networks*, Human Factors, Vol. 45, No. 4, pp. 635-643.

Worksheet

Preface

The present study was formed and presented by the 10th semester group 10211 as the Master thesis in Sports Technology at Aalborg University. The study was conducted in the period from 1st of February to June the 6th 2019. The study was supervised by Anderson de Souza Castelo Oliveira, whose knowledge, skills and feedback, the study group is very thankful for. Also, a big thanks to Samuel Emil Schmidt who assisted with the hardware setup and Ramtin Marandi for supplying the mental task GUI. The following paper contains a worksheet describing, presenting and discussing the process from initial problem, theory, test protocol, data analysis and results, in continuation of the previously presented scientific article. All material is presented in English.

Table of contents

Preface	2
Table of contents	3
Abbreviations	5
1.0 Theory	6
1.1 Mental workload	6
1.1.1 Performance measures	7
1.1.2 Subjective measures	8
1.1.3 Physiological measures	8
1.2 Cardiac cycle	10
1.3 Autonomous Nervous System	13
1.4 Electrocardiography	13
1.5 Seismocardiography	14
1.6 Accelerometer	16
1.7 Machine learning	17
1.7.1 Decision Trees	20
1.7.2 Nearest Neighbor	20
2.0 Method	21
2.1 Participants	21
2.2 Experimental setup	21
2.2.1 SCG accelerometer	21
2.2.2 ECG electrodes	22
2.2.3 A/D converter, amplifier and computer software	23
2.3 Mental task	23
2.4 Pilot trials	27
2.5 Experimental protocol	27
3. Data analysis	30
3.1 Performance metrics	30
3.2 Subjective ratings	31
3.3 Physiological measures	31
3.3.1 ECG cardiac cycle segmentation	32
3.3.2 SCG cardiac cycle segmentation	33

3.4 HRV analysis	36
3.5 Statistical analysis	38
3.6 Machine learning	38
4. Results	41
5. References	45
Appendix 1	48

Abbreviations

aHF	- Absolute power of the high-frequency band (0.15–0.4 Hz)
aLF	- Absolute power of the low-frequency band (0.04–0.15 Hz)
aTotal	- Absolute power of all frequency bands
EEG	- Electroencephalography
ECG	- Electrocardiography
HRV	- Heart Rate Variability
HRVTi	- Integral of the density of the NN interval histogram divided by its height index
LFHF	- Ratio of LF-to-HF power
meanHR	- Mean heart rate
Peak HF	- Peak frequency of the high-frequency band (0.15–0.4 Hz)
Peak LF	- Peak frequency of the low-frequency band (0.04–0.15 Hz)
pLF	- Percentage of the sum of aLF and aHF for the low frequency band
pHF	- Percentage of the sum of aLF and aHF for the high frequency band
RMSSD	- Root mean square of successive NN interval differences
SCG	- Seismocardiography
sdHR	- Standard deviation of heart rate
SDNN	- Standard deviation of NN intervals
TINN	- Baseline width of the NN interval histogram

1.0 Theory

1.1 Mental workload

The study of mental workload (MWL) established itself during the 1980s, as the development of technology imposed greater cognitive demands on individuals. Even though MWL has an intuitive appeal, a concordant definition of the concept is still lacking. A frequently used analogy is however often made between mental and physical load, where they are comprised of stress (task demands) and strain (impact on human). Stress furthermore comprises several factors associated to the demand, where strain relies on available resources to cope with the demands leading to a demand/resource balance operationalization (Young et al., 2015). The definition of MWL from Young & Stanton, 2005 tries to comprehend this demand/resource relationship, as MWL reflects *“the level of attentional resources required to meet both objective and subjective performance criteria, which may be mediated by task demands, external support, and past experience”* - Young & Stanton, 2005, chp. 39-1. This definition implies that MWL is affected by various factors, where cognitive load theory has been suggested to distinguish these factors from each other. Three categories of cognitive load have been distinguished, referring to either intrinsic, extraneous and germane cognitive load. Intrinsic cognitive load relates to the intrinsic nature of the task being processed such as task difficulty. Extraneous cognitive load relates to external factors influencing the induced load such as time pressure and noise. Germane cognitive load relates to load induced by scheme formation and automation to solve the task. All the aforementioned loads are perceived to be additive in regard to the demands of the work, where the outcome relates to demand/resource balance (Galy et al., 2012).

One of the reasons to study MWL is to identify suboptimal workloads represented (either underload or overload), by linking MWL and task performance (Figure 1).

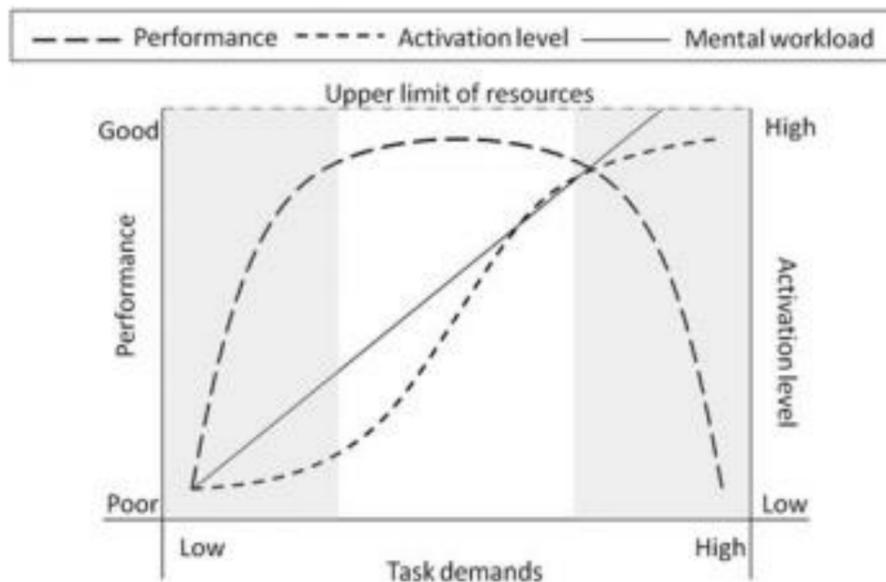


Figure 1: Relationship between activation level, workload and performance (Young et al., 2015).

As noticeable on Figure 1 both underload and overload are believed to be detrimental to performance, which solidify why performance measures are one of three main measures of MWL (Young et al., 2015). The other two relates to either physiological measures and subjective measures and are all often used in combination for the triangulation of measures to assess MWL (Charles & Nixon, 2019).

1.1.1 Performance measures

A frequently used metric of MWL is based on direct performance registration of the primary task to be dealt with. Primary task measures are useful when the performance can be related to other MWL levels, where parameters such as speed, accuracy, reaction time and error rate are often used performance measures (Cain, 2007). The assessment of MWL may also be conducted by evaluating the performance of a secondary task, as it relates to the spare mental capacity, not used by the primary task, and thereby a measure of the MWL (Young et al., 2015). Performance measures are highly dependent on the task to be performed, which complicate the overall comparison of the correlation between performance measures and MWL. Previous studies have proven that different MWL levels can be quantified by performance measures of the primary task (Marandi et al., 2018A; Marandi et al., 2018B)

1.1.2 Subjective measures

MWL can also be assessed using subjective measures such as self-reporting the degree of perceived MWL induced by the task (Young et al., 2015). The most used self-reporting scheme is the NASA Task Load Index (NASA-TLX) (Charles & Nixon, 2019), and consist of six factors represented by subscales exploring the mental demand, physical demand, temporal demand, own performance, effort and frustration level. NASA-TLX is a two-part evaluation consisting of determining weights of the aforementioned factors and the rating of each. The weights are based on pairwise comparisons of the different factors contribution to the perceived MWL of the specific task. The rating of each factor is given on a numerical scale (0-100), reflecting the magnitude of each factor in the specific task. Each scale is divided into 21 gradations encapsulated by bipolar descriptors, indicating the low and high end of the scale (NASA, 1986). The NASA-TLX has proven to be a useful tool to assess MWL (Marandi et al., 2018A; Marandi et al., 2018B).

1.1.3 Physiological measures

Several different physiological measures have been presented to assess MWL. Charles & Nixon, 2019 identified six key measures that is noninvasive nor requires medical procedures and expensive static equipment, which includes cardiac activity, respiration, skin-based measures, blood pressure, ocular measures and brain activity (Charles & Nixon, 2019).

One of the most commonly used physiological measures, have been reported to be cardiac activity (Charles & Nixon, 2019), where different measures associated with the cardiac cycle give insight into MWL. The heart rate variability (HRV) is a typically reported parameter, which can be analyzed in both the time and frequency domain. These measures are based on the time between heartbeats, also known as inter-beat intervals (IBI), and the variance within. A summary of the time domain features has been presented in Figure 2 and the frequency domain measures in Figure 3.

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

Interbeat interval, time interval between successive heartbeats; NN intervals, interbeat intervals from which artifacts have been removed; RR intervals, interbeat intervals between all successive heartbeats.

Figure 2: Summary of time domain features (Shaffer & Ginsberg, 2017).

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

Figure 3: Summary of frequency domain features (Shaffer & Ginsberg, 2017).

The frequency domain methods involve a power spectral density analysis, with three commonly used spectral components being very low frequency (VLF - 0.0033-0.04Hz), low frequency (LF - 0.04-0.15Hz) and high frequency (HF - 0.15-0.4Hz) (Shaffer & Ginsberg, 2017). The HF spectrum reflects the parasympathetic activity and is commonly referred to as the respiratory band due to heart rate variations related to the respiratory cycle. These heart rate variations are known as respiratory sinus arrhythmia, as inhalation accelerates heart rate and slows during expiration. The LF spectrum reflects baroreceptor activity where the detection of arterial blood pressure adjusts the HR. (Shaffer et al., 2014)

Several studies support the correlation between HRV and MWL in office settings. Both Cinaz et al., 2013 and Taelman et al., 2011 found that several different HRV measures in both time and frequency domain decreased accordingly to an increase of the induced MWL (Cinaz et al., 2013; Taelman et al., 2011). Marandi et al., 2018B furthermore showed that MWL induced from a mental computer task resulted in significant different cardiac measures between load levels (Marandi et al., 2018B).

1.2 Cardiac cycle

The cycle that spans from the contraction of atria to ventricular relaxation, is known as the cardiac cycle. Two distinct periods of the atriums and the ventricles are present within this cycle. The period of contraction that allows the heart to pump and circulate blood is called systole period. The period of relaxation where heart chambers refill with blood is called diastole period. The systolic and diastolic periods do not occur simultaneously for the atriums and ventricles, and this offset is essential for a well-functioning heart. The heart activity and the individual periods of the cardiac cycles can be examined by electrocardiography (ECG), by measuring the electrical signals from the heart. (OpenStax, 2013)

The heart utilizes the aspects of pressure and flow in order to move blood through the heart and provide circulation in the cardiovascular system. As fluids, both liquids and gasses, move from

higher pressurized to lower pressurized regions, the heart is able to move blood by pressurizing different chambers of the heart and thereby controlling the blood flow. (OpenStax, 2013)

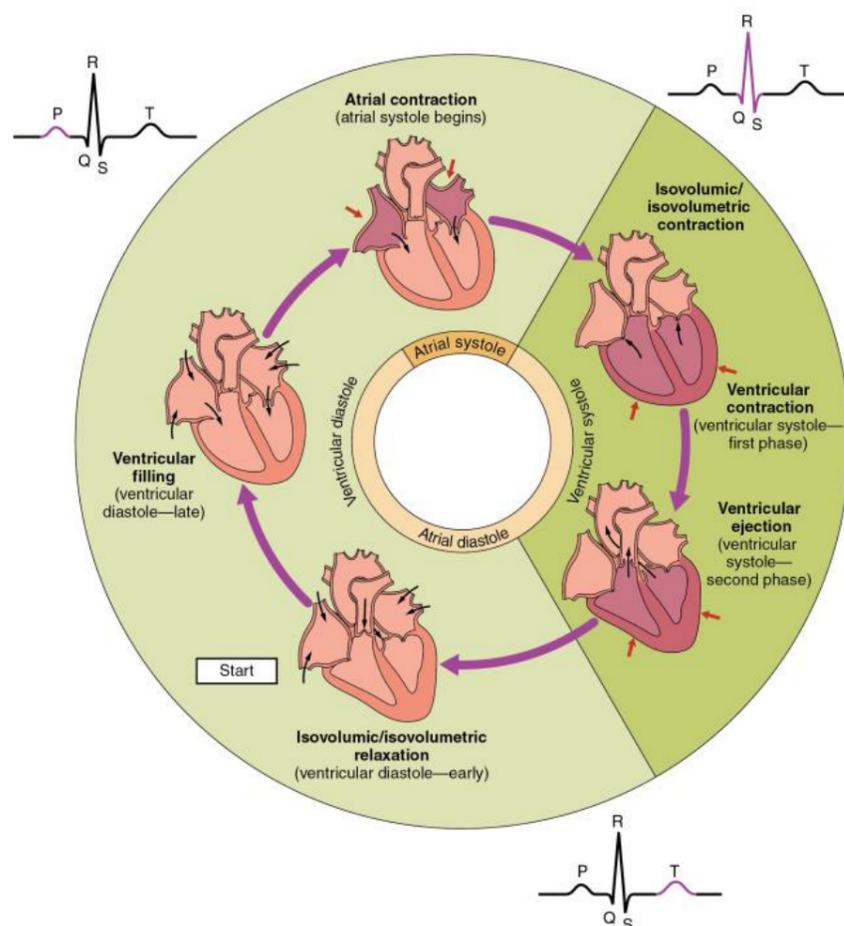


Figure 4: Overview of the Cardiac Cycle. The cardiac cycle begins with atrial systole and progresses to ventricular systole, atrial diastole, and ventricular diastole, when the cycle begins again. Correlations to the ECG signal are highlighted. (OpenStax, 2013)

When the heart muscles relax (diastole) blood will flow from higher pressurized veins to the atriums, building up pressure in the atriums and blood will passively flow to the ventricles through the atrioventricular valves, filling up 80% of the ventricles. This phase is called ventricular filling (Figure 4). The remaining 20% fills up when the action potential triggers the atrial contraction (atrial systole), forcing the remaining blood through the atrioventricular valves. This phase is represented by the P-wave of the ECG signal. The contraction of ventricles (ventricle systole) is divided into two phases. Initially when the ventricles start to contract, pressure will rise, but not high enough to force the semilunar valves to open. As pressure quickly rises, the pressure is now higher than that of the atria, resulting in closing of the atrioventricular valves and thereby preventing backflow. Since no flow is occurring, the volume of the ventricles remains

the same in this initial phase of the ventricle systole. This phase is called the isovolumic contraction. In the second phase of the ventricle systole, the ventricular ejection phase, the pressure within the ventricles has now reached a threshold beyond what the semilunar valves can withstand. Blood is now rapidly ejected into the pulmonary trunk and aorta. The ventricular systolic phase is represented by the QRS complex in the ECG signal. Once the ventricles start to relax (ventricular diastole), pressure within the ventricles will fall to a level, lower than what is present in the pulmonary trunk and aorta. Blood will start to flow back towards the heart (represented by the small dip in aortic blood pressure readings, (Figure 5) forcing the semilunar valves to close and prevent backflow into the heart. Since the atrioventricular valves are still closed at this point, the volume of the ventricles remains the same (isovolumic ventricular relaxation phase). As the ventricular muscles continue to relax, pressure within the ventricles will continue to fall, eventually lower than the pressure in the atriums. This phase is represented by the ECG T-wave. Pressure will force blood from the atria to the ventricles, thereby opening the atrioventricular valves. The semilunar valves are closed, and both the atriums and ventricles are in diastole, thereby completing the cardiac cycle. (OpenStax, 2013)

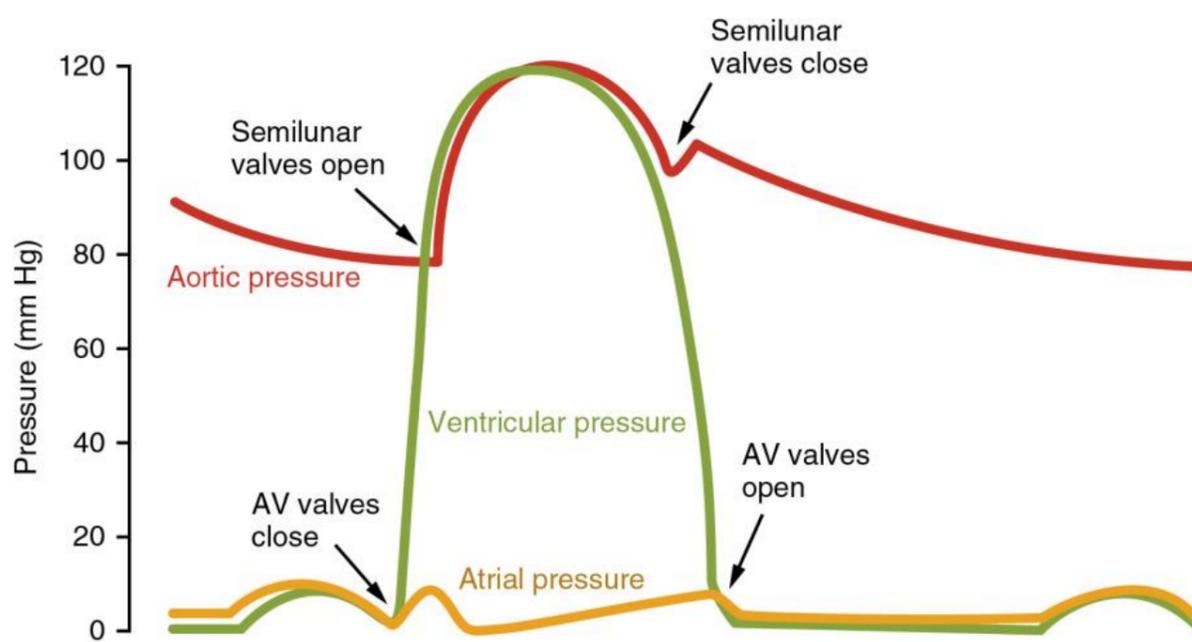


Figure 5: Blood pressure during one cardiac cycle, in different areas of the cardiovascular system surrounding the heart (OpenStax, 2013).

1.3 Autonomous Nervous System

The autonomous nervous system is a part of the efferent component of the peripheral nervous system, with the purpose of innervating the glands, the heart and smooth muscles and thereby regulation of a wide range of body functions (blood flow, heart rate, digestion, respiration etc.). The main function of the autonomous nervous system is to response to both internal and external changes and maintain body homeostasis. (Richter & Wright, 2013)

The autonomous nervous system is divided into two subsystems, the sympathetic and parasympathetic nervous system. The sympathetic part accelerates bodily functions (for example in a fight or flight situation), and is associated with for example increased workload (physical or mental) and stress. When exposed to situations where metabolic exertion is required, either during physical- or higher mental activity, sympathetic activity is increased. Heart rate increases, blood pressure rises, and adrenaline is released, while pupils are widened, and blood flow is directed towards skeletal muscles to increase the ability of the body to cope with the raised mental or physical demands. However, prolonged and raised sympathetic activity can lead to stress and diseases. The parasympathetic part works in the opposite way, responsible for relaxation, preparing the body for recovering. (Richter & Wright, 2013)

1.4 Electrocardiography

With placement of electrodes on specific locations of the body, it is possible to record the complex electrical heart-induced signals. These electric signal recordings are recorded by ECG. With analysis of the ECG, a detailed picture of normal and abnormal heart function can be computed. The ECG normally uses 3, 5 or 12 electrodes, with more electrodes providing more information. However, for obtaining HRV, only a 3-electrode setup is required. (OpenStax, 2013)

Recordings of a normal functioning heart contains multiple components and intervals that corresponds to important heart-induced electrical events, that describes the relationship between heart contraction and electrical signals (Figure 6). (OpenStax, 2013)

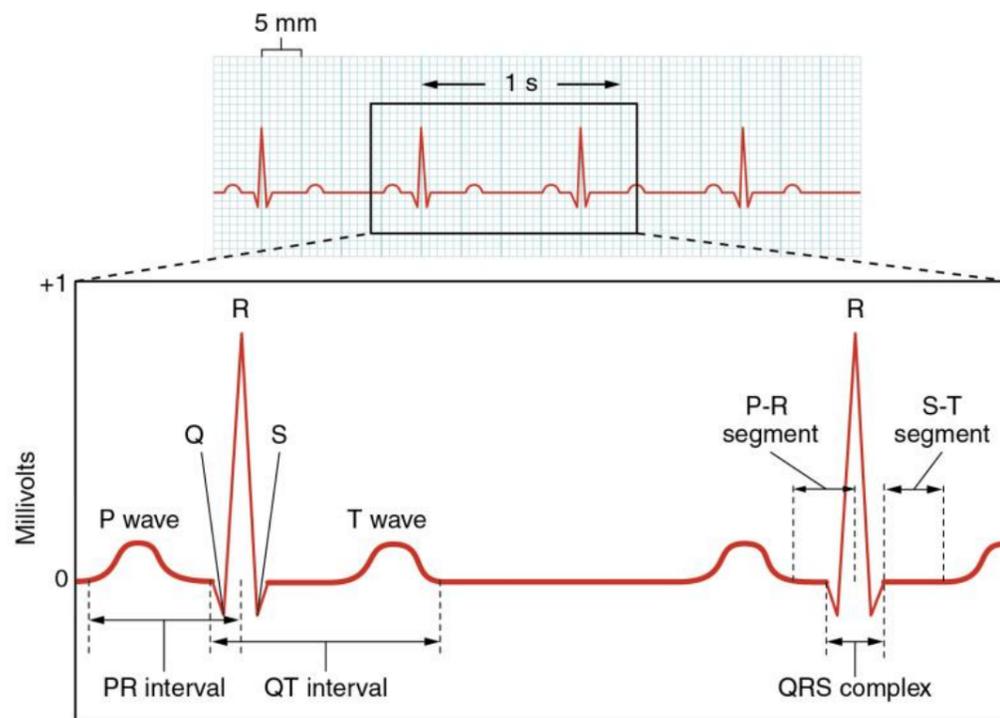


Figure 6: ECG from a normal operating heart. (OpenStax, 2013)

Within one cardiac cycle, five eminent events are present on the ECG: the P-wave, the QRS complex and the T-wave. Represented by the P-wave are the depolarization of the atriums, where the physiological contraction occurs approximately 25ms after the start of the P-wave. The QRS complex represents the depolarization of the much larger ventricles, that requires a stronger electrical signal. Following the QRS complex, the T-wave is caused by the repolarization of the ventricles. Repolarization of the atriums occur simultaneously with the QRS complex, making it invisible in an ECG. (OpenStax, 2013)

Interpretation of the ECG can reveal certain abnormalities in the cardiac function. Amplitude and duration for different segments and intervals, as well as vector analysis, are used to discover cardiac related issues, e.g. enlargement of the atria, represented by an amplified P-wave. Furthermore, ECG signals can be used to obtain the HRV. (OpenStax, 2013)

1.5 Seismocardiography

Using techniques derived from seismology, seismocardiography (SCG) is the measure of chest wall vibrations caused by heart induced movements, as the heart is contracting, ejecting blood

into the vascular system and closing and opening and closing of heart valves. By utilizing highly sensitive, low-noise accelerometers, commonly attached to the lower part of sternum, these mechanical vibrations can readily be detected. With implementation of tri-axial accelerometers, heart-induced accelerations in three axes can be obtained, and not only the most commonly used dorso-ventral component. (Inan et al., 2015; Salerno & Zanetti, 1990)

Compared to ECG, SCG provides new information of the cardiac cycle, related to the actual movements of the heart, and not only the electrical signals. Fiducial points indicating e.g. the mitral valve closing, aortic valve opening, and closing can be identified (Figure 7).

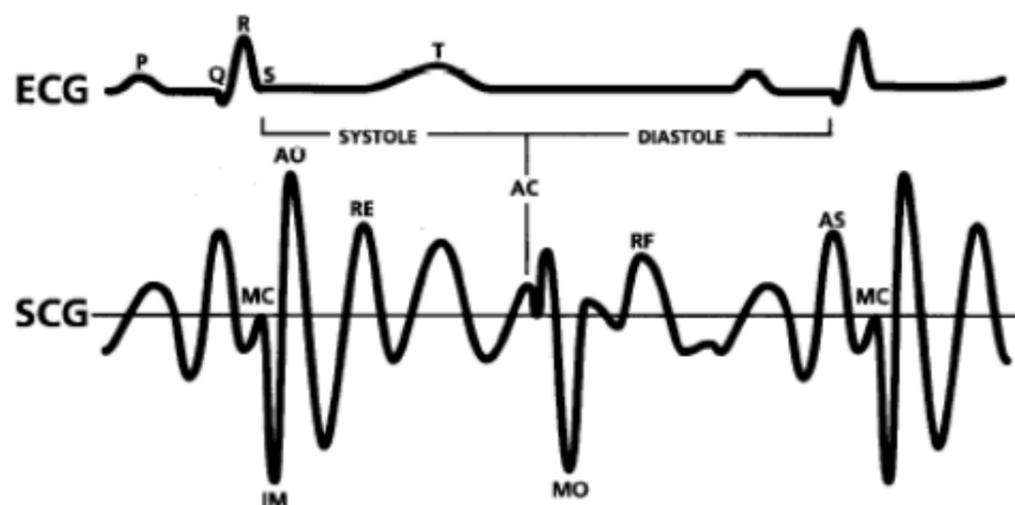


Figure 7: Dorso-ventral SCG signal compared to ECG signal showing fiducial points of MC= Mitral (Valve) Closure, AO =Aortic (Valve) Opening, RE = Rapid (Ventricular) Ejection, MO = Mitral (Valve) Opening, RF = Rapid (Ventricular) Filling, AS =Atrial Systole, AC =Aortic (Valve) Closure (Dinh, 2011)

With advances in accelerometer technology, the sensitivity of the accelerometers has greatly improved over the last 60 years since SCG were first investigated. Together with the decreasing weight of the accelerometers, the method of SCG is becoming more and more precise and applicable. (Inan et al., 2015)

1.6 Accelerometer

Accelerometers are basically inertial sensors that utilize force sensing to quantify linear acceleration. Accelerometers commonly involve a mechanical sensing element, which has a proof mass suspended with respect to a reference frame. Inertial force due to accelerations will cause deflection of the suspended proof mass according to Newton's Second Law, which can be measured electrically by principles of piezo resistance, piezoelectricity and differential capacitance as the most common. Piezoresistive accelerometers consists of a cantilever beam with incorporated piezo resistors arranged as a Wheatstone bridge, which results in proportionality between applied acceleration and voltage output. In a piezoelectric accelerometer the applied accelerations cause the sensing element to bend, which results in a change in output voltage. In a differential capacitive accelerometer, the displacement of the proof mass is measured capacitively, where the mass is encapsulated between two electrodes. As acceleration causes deflection of the proof mass and thereby changes in the electrical capacitance. Differential capacitive accelerometers are usually used when high sensitivity is required, due to a low noise level, which makes it ideal for seismocardiographic measures. (Yang & Hsu, 2010)

Accelerometers have been used for various purposes in the areas of sport and healthcare. This includes quantification of sport specific movements, which covers both gross body movement and interlimb kinematics by applying various numbers of accelerometers (Chambers et al., 2015). Accelerometers are similarly used to classify different movements and postures, which furthermore includes estimations on e.g. steps completed and thereby an indirect estimation of physical activity. In close relation, accelerometers have been used to estimate energy expenditure in various settings. Postural stability has also been subject to the implementation of accelerometers, where fall detection algorithms have been developed to identify remote fall incidents (Yang & Hsu, 2010). Other implementations of accelerometers include assessment of oral activities by a tooth prosthesis instrumented with an accelerometer and as previously mentioned, quantification of cardiac activity by means of ballistocardiography and seismocardiography (Tamura & Chen, 2018).

1.7 Machine learning

Machine learning comprises the ability to transfer a learning process to a computer, which definitely grants the ability to make decisions based on distinguishing inequality. This entails face recognition, biometric authentication, and document recognition among others (Dougherty, 2013). The purpose of applying machine learning are often to predict specific patterns in any given data with a supervised or unsupervised approach. Unsupervised learning implies that any given algorithm is trained to cluster the data without any pre definition of present clusters in the data. The purpose of applying unsupervised learning is thereby to discover unknown clusters, known as classes, which is left to be interpreted. Regarding supervised learning, classes has been preassigned to the data and used to train the algorithm to make predictions on future data. This approach ensure that the predefined classes entail relevance to the purpose of applying machine learning (Dougherty, 2013). A specific area within supervised learning is classification, where a classification algorithm is trained to cluster the data into the predefined classes. A schematic overview of classification is illustrated in Figure 8.

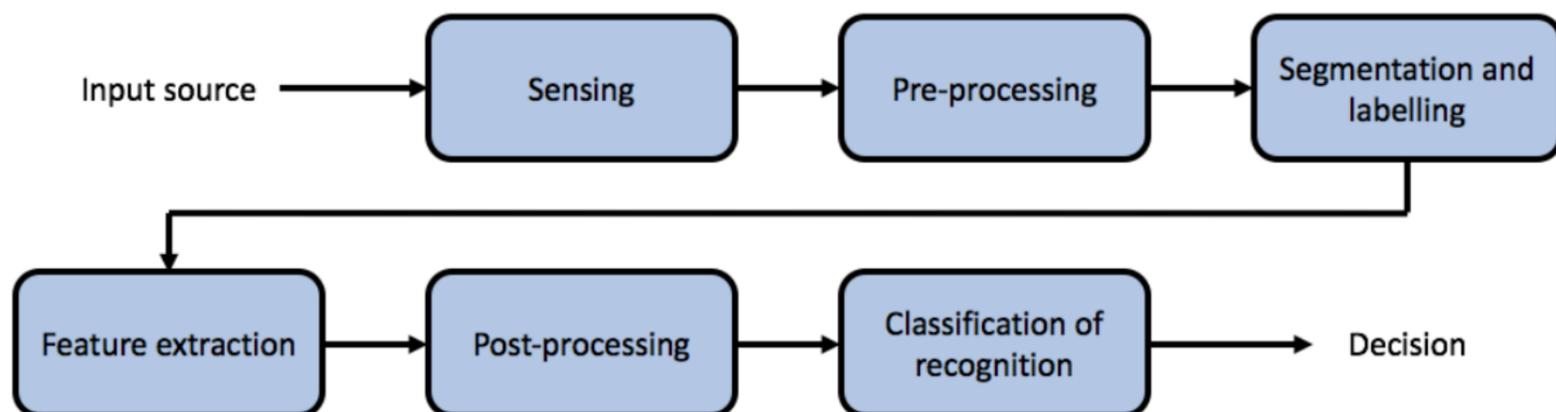


Figure 8: Schematic overview of a classification process (Dougherty, 2013, pg. 4)

Labelling of classes is the first step directly related to classification, and the term labelling noise should be considered in that process. Labelling noise relates to incorrect labelling of classes with the risk of invalidating the results from a classification process. The risk is however limited with robust definition of the classes included (Kubat, 2017). The next step in the classification process is feature extraction, with the purpose of assigning single valued features to represent each set of class labelled data. The outcome of the feature extraction has a great impact on the

classification performance, so several aspects should be considered. First of all, the included features should characterize each class with large inter-class distance and small intra-class difference (Dougherty, 2013). It should furthermore be secured that all relevant features are included while still excluding irrelevant or redundant features (Kubat, 2017). Irrelevant features should be avoided due to the “curse of dimensionality” which states that the amount of data needed to train the classifier increases as the number of features increases. In relation to a finite amount of data, the performance of the classifier increases as the number of features increases until a certain point in dimensionality followed by a decrease of performance as illustrated in Figure 9.

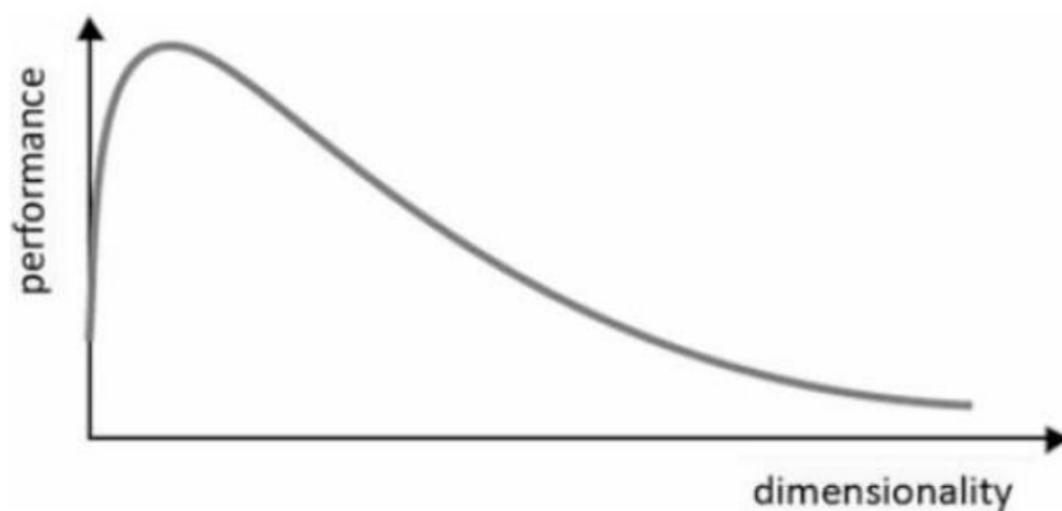


Figure 9: Classification performance in relation to dimensionality (Dougherty, 2013, pg. 124).

This relationship is due to overfitting, which implies that the classifier has been trained to be too specific and therefore not flexible in regard to classifying new data. Overfitting can also be related to the specificity of the classification algorithm as illustrated in Figure 10, where two boundary separation techniques (linear and polynomial) are compared. Instance B seems superior in regard to instance A to classify the white and black dots correctly, but instance B might be performing poorer than instance A when faced with unseen data due to the specificity of instance B.

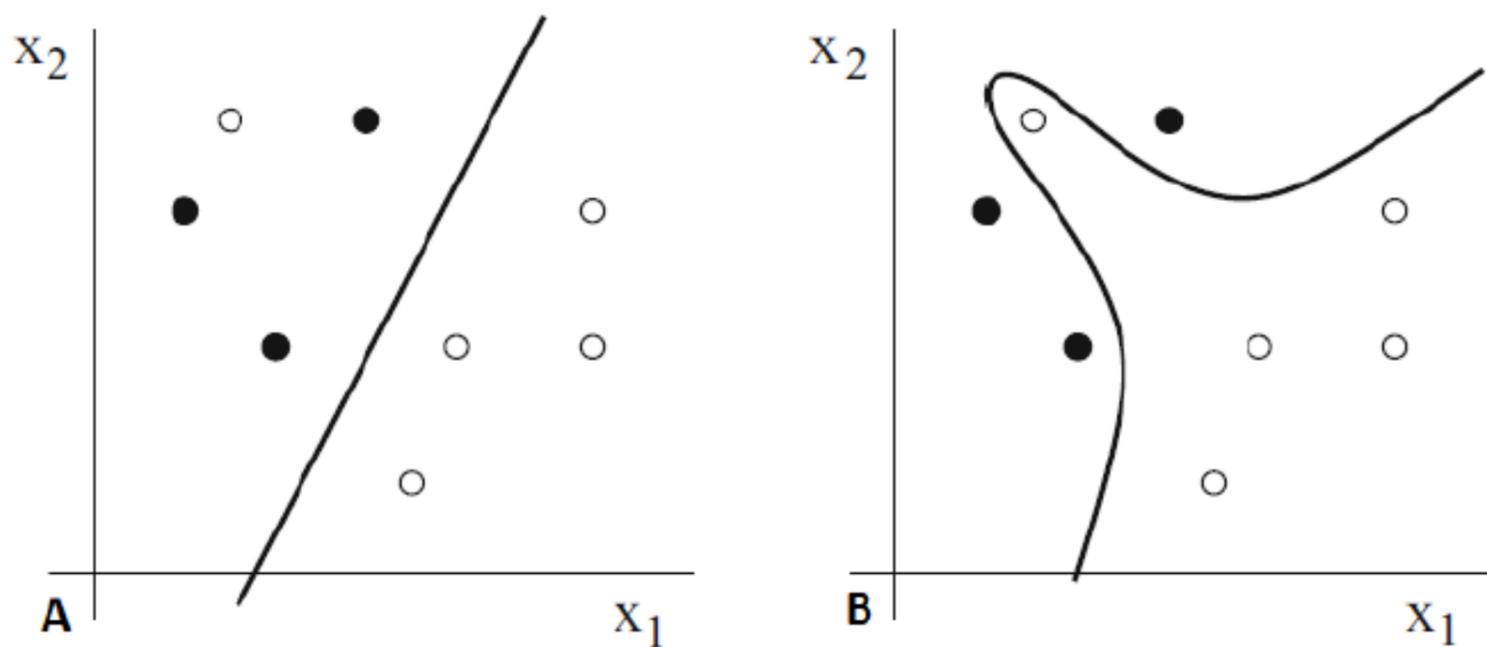


Figure 10: Feature x_1 and x_2 is generally linearly separable as shown in instance A with one false classification due to noise. Instance B illustrates overfitting of the polynomial boundary due to the specificity of the algorithm (Kubat, 2017, pg. 82).

Several classification algorithms exist, but none are superior in regard to performance as the right choice of classification algorithm is very case determined. The evaluation of classification algorithms is often a twofold process which includes a validation evaluation based on the data used for training and an evaluation of the true error based on unseen data referred to as the test set. The purpose of validation is to provide an initial estimate of the performance. This validation is either based on a k-fold cross validation or a simple hold out method. K-fold cross validation separates the training data into k number of folds, where one fold is used for validation and the remaining for training. This procedure is then repeated k times meaning that all the training data has been used to train and validate the performance of the classifier. The hold out validation reserves a portion of the training data for validation purposes. The process evaluating the true error is however crucial, as evaluating the true error based on the validation process, would result in a biased true error (Dougherty, 2013).

1.7.1 Decision Trees

Decision trees works by conducting several tests and thereby narrowing the field of possible classes to be assigned. Every decision tree is initiated by a root statement that relates to a feature, which then leads to other test-containing nodes related to another feature. Eventually, leaf nodes are reached which relates to different classes and the data are assigned to this specific class. Decision trees are generally easy to interpret as every test node is directly relates to the different features used. (Kubat, 2017)

1.7.2 Nearest Neighbor

Nearest neighbor classification algorithm is based on similarity, as the unseen data is classified to the most similar class based on the features involved. The similarity is quantified by the Euclidean distance in an n-dimensional space which is specified by the number of features included. The data is thereby classified to be the same as its nearest neighbor and thereby the minimum Euclidean distance between the unseen data and any data used for training. The Euclidean distance between two instances of data in n dimensions, $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$, is given by equation 1:

$$d_E(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

The number of nearest neighbors is often valuable to consider due to noise. By increasing the number of nearest neighbors to be considered, the data is classified to the class that is represented the most in its k nearest neighbors. The nearest neighbor algorithm furthermore solidify why irrelevant and non-normalized data might impair the performance of the classification. (Kubat, 2017)

2.0 Method

2.1 Participants

All participants (12 males) (Age = 26 ± 1 years, body mass = 81.8 ± 9.3 , BMI = 24.9 ± 1.8 kg/m²) had normal or corrected to normal vision and were right handed computer mouse users. Participants were all non-smokers, non-drug addicts and had no known mental or heart diseases. Participants were instructed to abstain from alcohol (24h), caffeine (12h), painkillers and sleep medicine (24h), prior to the experiment. All participants were furthermore instructed to sleep a minimum of 7 hours prior to the experiment (reported sleep = $7.4 \pm 1,1$ hours). All experimental trials were conducted between 9am - 3pm. All participants signed a declaration of consent (Appendix 1) prior to the experiment, and were informed about their rights to withdraw from the experiment at any time. The participants were to be excluded if they were unable to complete the entire experimental protocol.

2.2 Experimental setup

The hardware setup consisted of a combined ECG and SCG measuring system. A five electrodes setup for the ECG, as well as a single accelerometer for the SCG, was connected to the same amplifier/AD-converter for synchronized data-logging. The hardware was connected via USB to a computer, running a data-logging software.

2.2.1 SCG accelerometer

For the recording of SCG signals, a small $\pm 2g$ low-noise ($5\mu g/vHz$) capacitive sensing accelerometer (Silicon Designs model 1221) with a sensitivity of 2000 mV/g was used. The accelerometer was encased in a small (10x10x8mm) lightweight 3D-printed PLA box and was placed on the lowest part of sternum with double adhesive tape. Furthermore, the wire was secured to the participants chest forming a small loop, to limit wire movement noise, and made sure not to overlap with the ECG electrode wires (Figure 11).

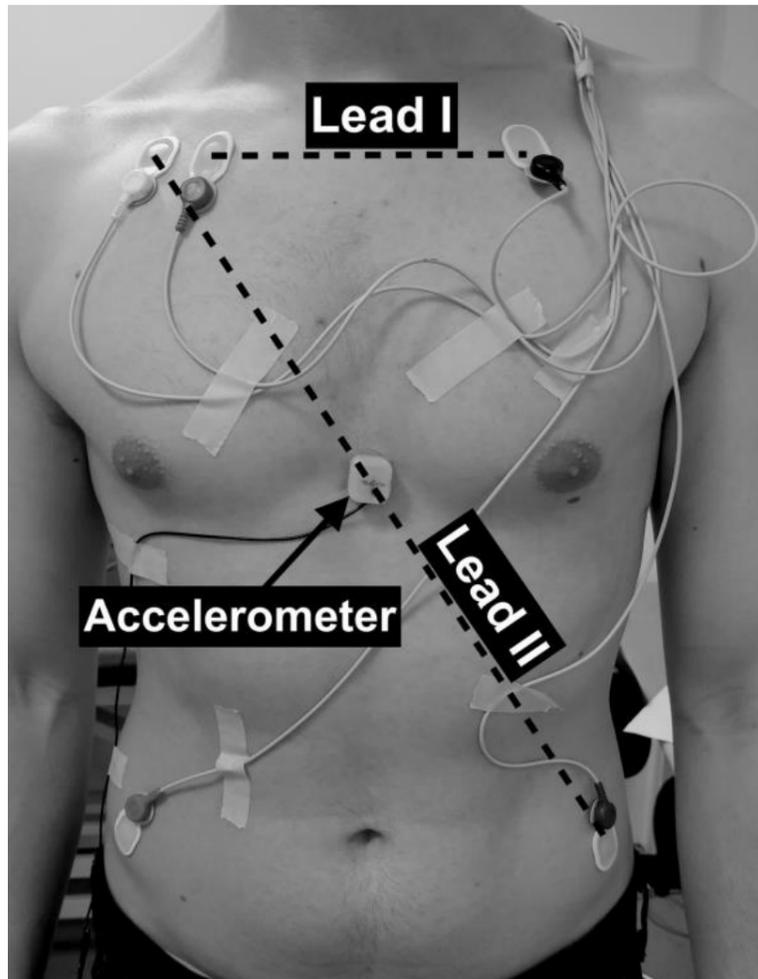


Figure 11: ECG electrodes and SCG accelerometer placement.

2.2.2 ECG electrodes

A two-lead ECG, with Lead I and II configuration using five electrodes (Ambu[®] Neuroline 720, Ag/AgCl wet electrode) placed with one acting as ground (right leg), two as Lead II configuration (right arm (negative) and left leg (positive)) and two as lead I configuration (right arm (negative) and left arm (positive)) in accordance with Einthoven's triangle (Figure 12).

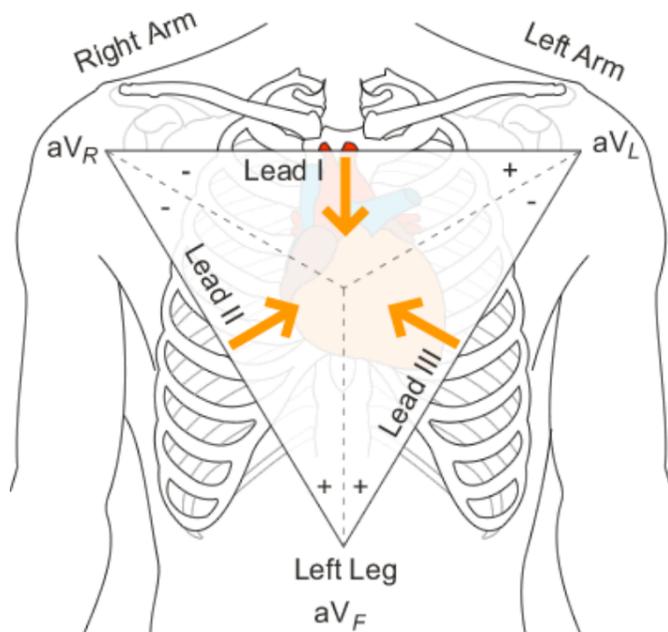


Figure 12: Einthoven's triangle of ECG electrode placement.

The electrodes were placed on the torso right under the clavicle for the arms and above the anterior superior iliac spine for the legs, to minimize electromyogenic noise. Furthermore, electrode wires were securely taped in place, forming a small loop to reduce wire motion. (Figure 11)

2.2.3 A/D converter, amplifier and computer software

An IWork 214 4-channel data recorder was used to sample both SCG and ECG signals. Sampling frequency was set to 1000 Hz. The data recorder was connected to a MacBook Pro running OS X 10.14.2. For the preview and recording of data, the IWork LabScribe V3.62 was used, able to export data as MatLab files for later analysis.

2.3 Mental task

A graphical user interface (GUI) running a connect-the-dots game (WAME 1.0) developed at Aalborg University (Marandi et al., 2018A), was used to induce the mental load. The task has been designed in accordance with other standard models of computer work (Samani et al., 2009, Birch et al., 2001) and consisted of cyclic computer operations where the participants had to recreate a specific pattern in a certain order. Each cycle involved a memorization period (MP), washout period (WP), and replication period (RP) (Figure 13).

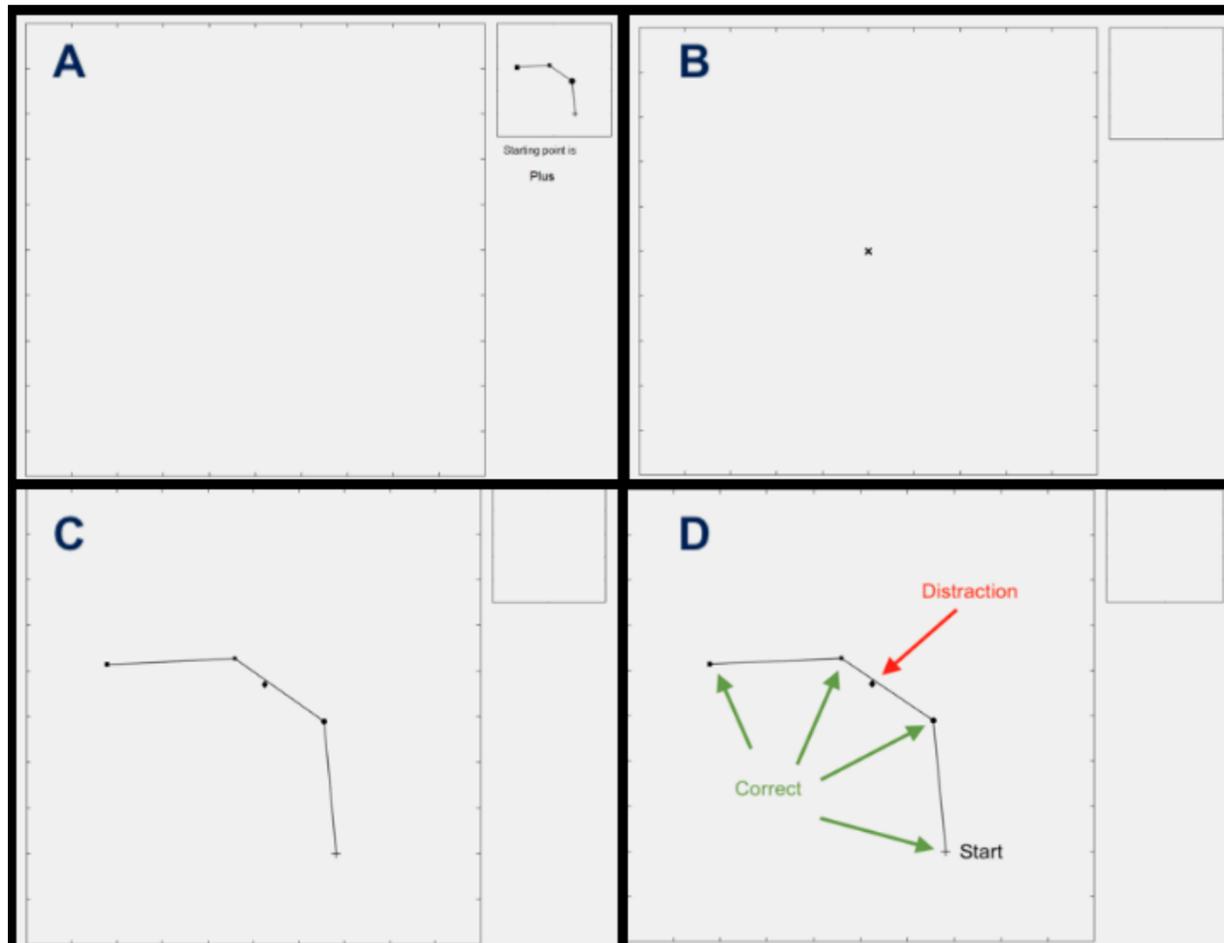


Figure 13: Mental task GUI. **A:** memorization period, **B:** washout period, **C:** replication period, **D:** replication period with starting-, distraction- and correct pattern points indicated.

The GUI was displayed on a Dell E193FP 19-inch LCD monitor with 1280x1024 resolution and 75Hz refresh rate, and was placed approximately 55cm from the participants eyes and the center of the screen was furthermore placed approximately 15° degrees below the horizontal line of sight (Marandi et al., 2018A) (Figure 14).

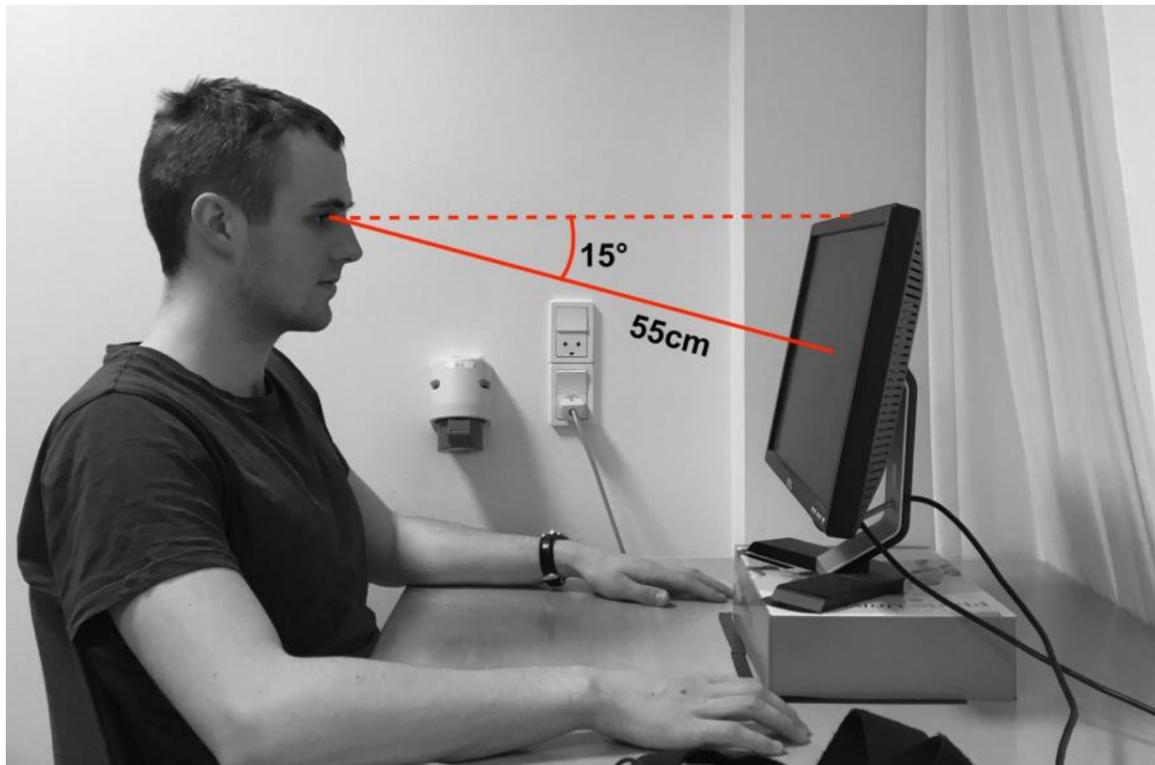


Figure 14: Mental task workstation with eyesight angle and distance illustrated.

The work panel contained a template panel with an appurtenant textual instruction and a replication panel, which subtended respectively 5° and 20° of visual angle in both horizontal and vertical direction. The GUI subtended approximately 27° of visual angle horizontally and 22° vertically. The area in which the participants were to complete the computer task was cordoned off to minimize the exposure of distracting elements in the surrounding environment and thereby standardize the extraneous cognitive load related to environment (Figure 15).



Figure 15: Mental task workstation with surrounding curtain to minimize distraction.

Each cycle was initiated by the MP, where a specific pattern composed by a series of points was showed on the template section (Figure 13, A). Each point could take various shapes i.e. plus, asterisk, circle, triangle, square, diamond, pentagram, and a short text indicated the starting point for the pattern. The MP was followed by the WP where the pattern in the template panel disappeared and a cross located in the center of the replication panel was used as indication of the WP (Figure 13, B). The mouse cursor was furthermore made invisible to avoid any prepositioning prior to the RP. The RP was initiated by a scaled version of the pattern points appearing on the replication panel, and the participant was now able to connect the points in the correct sequential order to replicate the same pattern shown during the MP (Figure 13, C). To indicate that the first point of the pattern was correctly clicked, the point was enlarged by a factor of two. Whenever the participant clicked on the points in the right order, a line was drawn to connect the newest correctly clicked point to the previous one, otherwise no line appeared. The number of pattern points (PP) to be connected and the geometrical complexity of the patterns were changed to increase the intrinsic cognitive load and thereby inducing three different levels of mental loads referred to as low (PP = 4), medium (PP = 5) and high level (PP = 6). The geometrical complexity was changed such that the angles between any connecting lines were tightened with increased mental load (Marandi et al., 2018A). All MWL levels of the mental task also included a distraction point (DP), which were to be avoided in the replication of the pattern (Figure 13, D). The patterns were predefined and randomly generated, where the order of these patterns was randomized for each participant to minimize the germane cognitive load. A time constraint of each period was predetermined and in accordance with Marandi et al., 2018A being 2.06s, 2.34s, and 2.62s for MP and WP, and 4.11s, 5.06s, and 6.02s for RP in low, medium, and high levels of mental load respectively. The purpose of the level dependent time constraint was to standardize the extraneous cognitive load.

2.4 Pilot trials

For early changes and adaptations, pilot trials were conducted. The initial assumptions and experimental procedures were tested for practical implementation and ease of use. Furthermore, the complete experimental setup was tested for bugs and for familiarization purposes. The pilot trials were conducted on two individuals, that corresponds to the participant characteristics as described in 2.1 Participants. During the pilot trials it was discovered how important wire placement was, to reduce noise on the ECG signal. More specifically, as the ECG record electrical impulses, the wire from SCG cannot be in contact with the ECG wires, as this strongly interfered with the ECG signal.

2.5 Experimental protocol

The experimental trials followed the protocol as shown in Table 1.

Phase	Time	Description
Introduction	5 min	Introduction & signing declaration of consent
Hardware setup	10 min	Skin preparation, placement of electrodes, respiratory band & accelerometer, preview signals
Traing session	2 x 5 min	Familiarisation with mental task on low & high level
Nasa TLX weighting	5 min	Subjective weighting of NASA TLX parameters
Rest	10 min	Resting
Mental task	5 min	Matlab GUI mental task WAME 1.0, mental task level randomised
Subjective rating	1 min	Nasa TLX subjective performance rating
Mental task	5 min	Matlab GUI mental task WAME 1.0, mental task level randomised
Subjective rating	1 min	Nasa TLX subjective performance rating
Mental task	5 min	Matlab GUI mental task WAME 1.0, mental task level randomised
Subjective rating	1 min	Nasa TLX subjective performance rating

Table 1: Experimental protocol.

After participants had arrived, received information about the study and signed the declaration of consent, height and weight was measured, and sleep and substances ingested prior to the experiment (alcohol, caffeine, painkillers, sleep medicine etc.) were reported. The workstation at which the participants would be sitting, when completing the mental tasks, were adjusted to each participants height, to ensure a standardized visual distance and angle, as described in paragraph 2.3 *Mental task*. Participants randomly drew an ID number matching a mental task

level sequence. The sequence of the three different mental load tasks was counterbalanced across the participants (Table 2).

Mental task level sequence order			
Participant	Trial 1	Trial 2	Trial 3
Participant 3	Low	Medium	High
Participant 2	Low	High	Medium
Participant 6	Medium	High	Low
Participant 1	Medium	Low	High
Participant 5	High	Medium	Low
Participant 4	High	Low	Medium
Participant 9	Low	Medium	High
Participant 8	Low	High	Medium
Participant 12	Medium	High	Low
Participant 7	Medium	Low	High
Participant 11	High	Medium	Low
Participant 10	High	Low	Medium

Table 2: Counterbalanced mental task level sequence orders for all participants.

One researcher was responsible for acquiring and scheduling participants, while the other was responsible for constructing the sequence matrix and running the MatLab GUI. This ensured no biased experiments, keeping the mental task sequence blinded from the participant responsible researcher.

Next, the skin underneath each ECG electrode was prepared to lower the skin impedance. This procedure included removal of hair, light abrasion of the skin surface and cleansing using alcohol wipes. It was ensured that the alcohol was fully evaporated before application of electrodes (Crawford & Doherty, 2011). The ECG electrodes and SCG accelerometer was placed on the participant according to paragraph 2.2.1 and 2.2.2, while ensuring that wires from each system were not in contact.

The mental task was explained to participants and a training session consisting of at least 2 x 5 min (5 min low level and 5 min high level, repeated until familiar) was completed, to familiarize participants with the task, while equipped with the apparatus. After the familiarization period was completed, participants weighted the different parameters of the NASA-TLX test, based on

their experience with the task during the familiarization period. Next, participants rested for 10 min, before completing the first mental task level determined by the participant ID based sequence. After the mental task, participants were instructed to complete the NASA-TLX subjective rating. This was repeated until all 3 mental task levels (each consisting of 5 min) had been completed and rated, after which the trial was completed. The experimental protocol was completed twice on each subject on 2 different days, with at least 7 days in between. However, participants did not weigh the NASA-TLX parameters on the second day, as all ratings use the weighting from the first familiarization period.

3. Data analysis

3.1 Performance metrics

An overall performance metric (OP) was computed for each task completed by each participant to address their dexterity. This OP metric was computed as the ratio of two other performance metrics, which quantifies how accurate and how fast the participant performed each task.

To account for the participant's clicking speed the mean reaction time (MRT) was defined. The MRT was computed in three different ways, depending on the degree of completion of the replication (Equation 2). If all the points in the pattern was correctly clicked, the time intervals (TI) between the correct clicks (CC) and the first click in respect to the task onset time were averaged with respect to the number of pattern points (PP). If only some of the PP were correctly clicked, the remaining time of the replication period (RTRP) was added to the summation of time in between correct clicks and averaged with respect to the number of PP. If no PP were correctly clicked, MRT was equal to the replication period.

$$MRT = \begin{cases} \frac{\sum_{i=1}^{CC} TI_i}{CC}, & \text{Completed pattern} \\ \frac{\sum_{i=1}^{CC} TI_i + RTRP}{CC + 1}, & \text{Partially completed pattern} \\ RP, & \text{No correct clicks} \end{cases} \quad (2)$$

The MRT was normalized with respect to the minimum of MRT across all participants (0.5120s). The parameter related to accuracy, selective attention (SelA), was defined as the ability to keep focused on a set of actions despite any distracting stimuli (Equation 3).

$$SelA = \frac{CC}{IC + PP + DC} \quad (3)$$

The SelA acquires the highest value when the number of CC is equal to the number of PP and where no incorrect clicks (IC) and clicks on the distraction point (DC) was performed.

The OP was defined as the ratio between SelA and MRT, where a value of 1 account for the highest performance and 0 accounts for the lowest performance (Marandi et al., 2018A).

3.2 Subjective ratings

Overall weighted ratings were computed for each task completed by each participant. A weighting of the six different parameters (mental demand, physical demand, temporal demand, performance, effort, frustration) were computed based on the pairwise comparisons of the parameters after the familiarization period. This weight was multiplied to the rating from the numerical scale associated with each parameter. These adjusted ratings were summed and divided by 15, which correspond to the total number of weightings, and results in a total weighted rating associated to each task.

3.3 Physiological measures

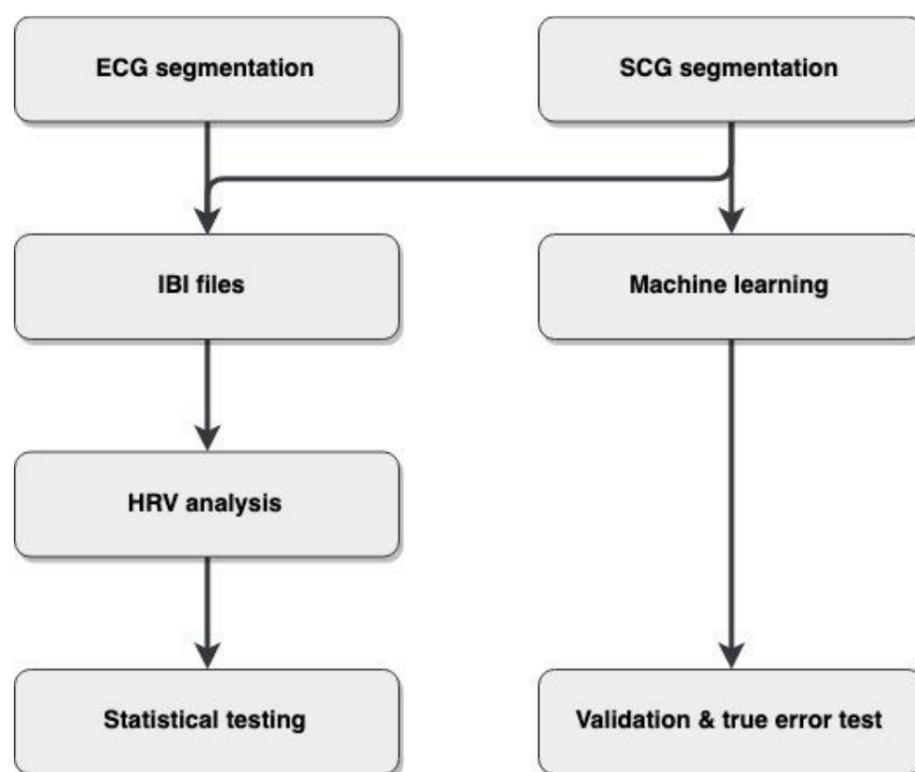


Figure 16: General steps involved in the data analysis of physiological measures.

Both ECG and SCG were included as physiological measures of MWL, and the general steps involved in the data analysis has been illustrated in Figure 16. These steps include segmentation of the ECG signal, segmentation of the SCG signal, HRV analysis, agreement between ECG and SCG HRV measures, statistical testing and machine learning.

3.3.1 ECG cardiac cycle segmentation

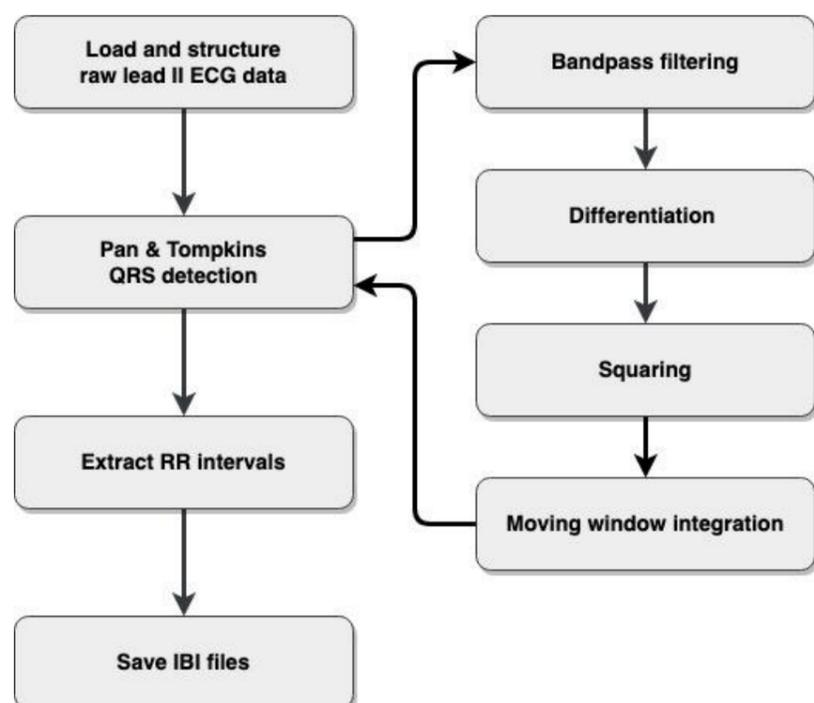


Figure 17: Processing steps involved in the ECG segmentation and creation of IBI files.

For the analysis of ECG based HRV, a MatLab script was constructed where the computational steps involved are illustrated in Figure 17. Firstly, raw Lead II data was extracted from the data matrix and structured based on the ID of each participant, test day and load level. Secondly, the MatLab toolbox BioSigKit (Sedghamiz, 2018) was implemented, using the Pan-Tompkins algorithm (Pan & Tompkins, 1985) for R peak detection, which includes several individual steps. The raw ECG data are filtered using a 3rd order Butterworth filter with $f_{c,low} = 5Hz$ as the low cut off frequency and $f_{c,high} = 15Hz$ as the high cut off frequency to eliminate baseline wander and high frequency noise (Sedghamiz, 2018). The filtered ECG signal are then differentiated, squared and ultimately integrated using a moving window integration with a window of 30 samples (Sedghamiz, 2018). The R peaks are then annotated if larger than $\frac{1}{3}$ of the max value of the filtered ECG signal (Sedghamiz, 2018). The IBI are computed as the time between adjacent R peaks and are ultimately saved as IBI files and later used for the HRV analysis.

3.3.2 SCG cardiac cycle segmentation

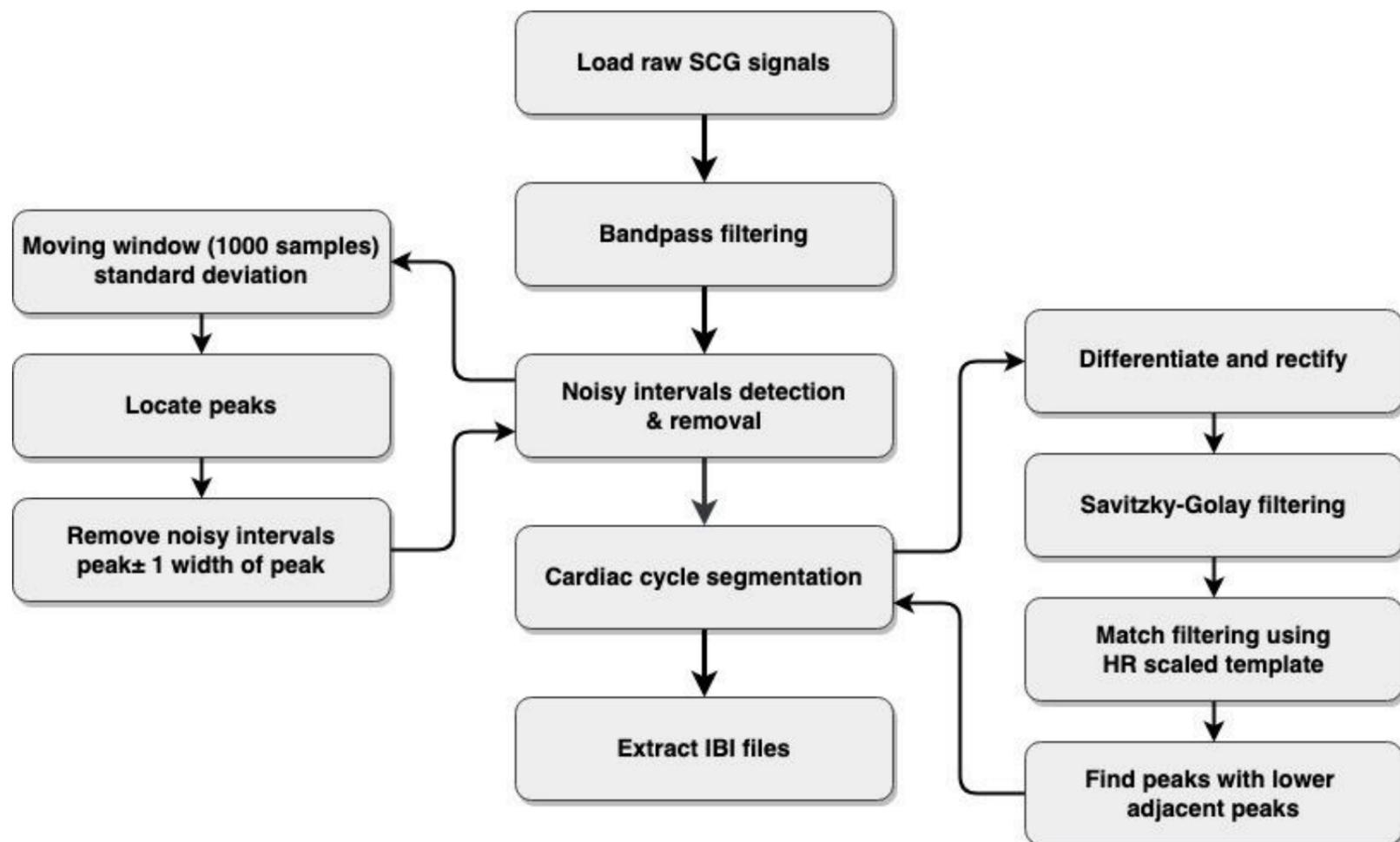


Figure 18: Processing steps involved in the SCG segmentation and creation of IBI files

For the purpose of running HRV analysis based on SCG, another MatLab script was constructed, involving several steps as illustrated in Figure 18. The goal was to automatically detect and remove noisy areas of the signal, and to segment individual cardiac cycles of the SCG signal without the use of the concurrent ECG measurements.

For the auto-detection and cancellation of noisy areas in the SCG signal (Figure 19, A) a moving standard deviation was calculated (window length=1000 samples) (Figure 19, B). The locations and widths of the noisy areas peaks were found using a minimum peak prominence of 0.3V, and the SCG was zeroed out around these peaks (± 1 width of the peak) (Figure 19, C) for later data analysis.

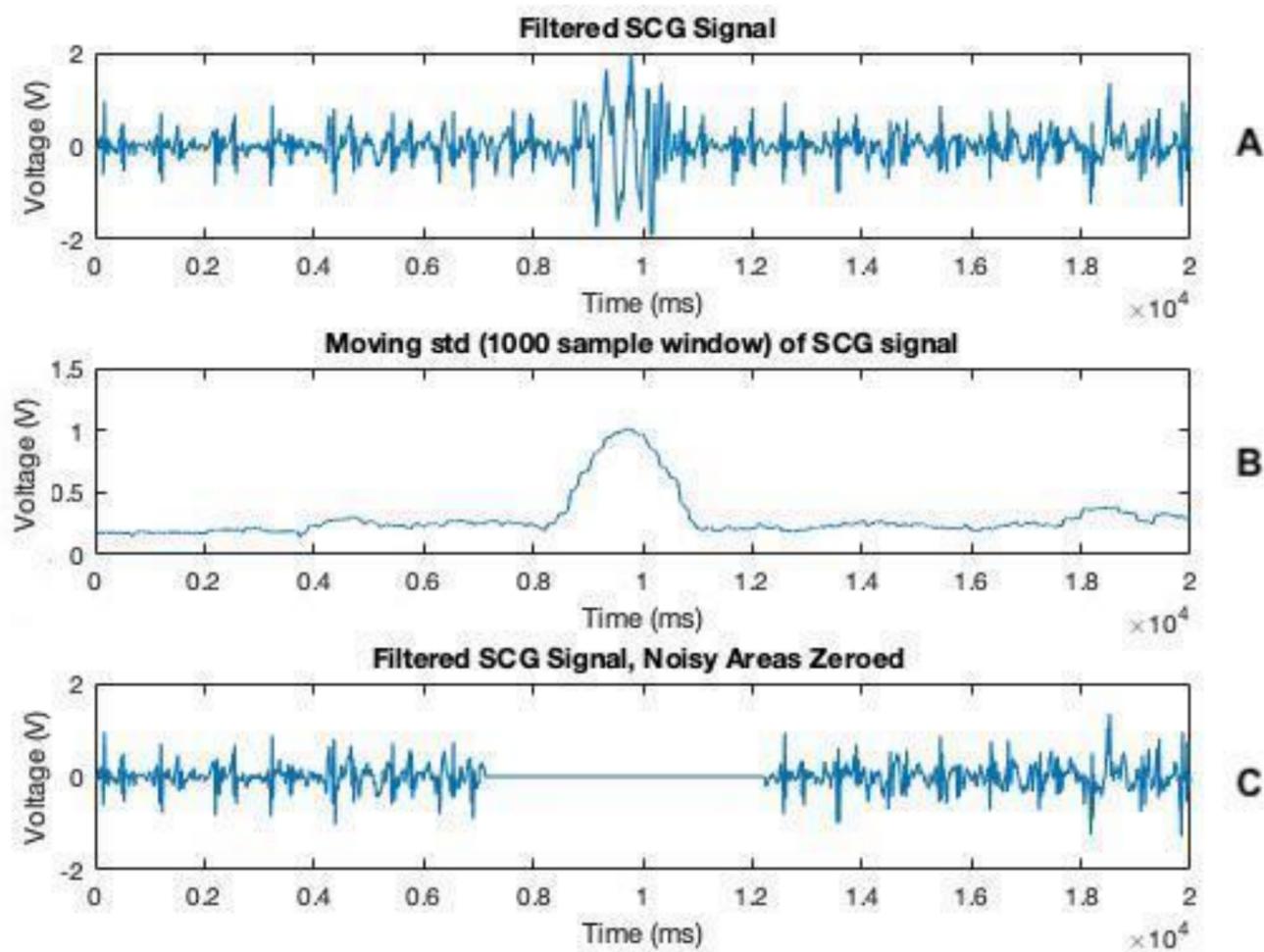


Figure 19: Steps involved in the automatic noise detection/cancelation of SCG signal. A) filtered SCG signal, B) moving std with a 1000 sample window, C) filtered SCG signal with zeroed noisy areas.

For the cardiac cycle segmentation, SCG data was filtered using a 4th order band pass Butterworth filter ($f_{c,low} = 0.5\text{Hz}$, $f_{c,high} = 80\text{Hz}$). The SCG signal was then differentiated, rectified and filtered using a 2nd order Savitzky-Golay filter with a window length of 101 samples for segmentation purposes (Figure 20, A,B,C).

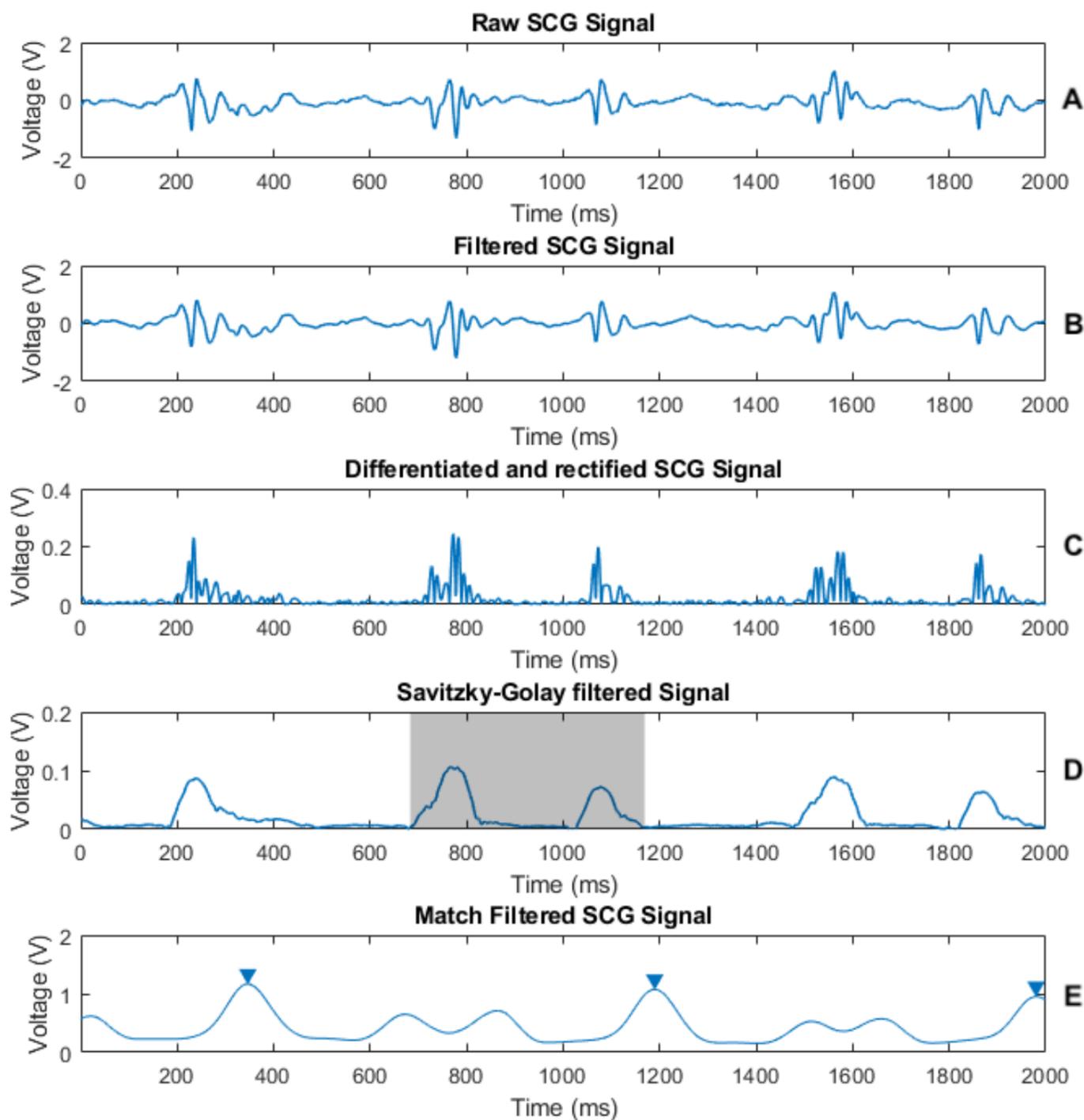


Figure 20: Steps involved in the automatic cardiac cycle segmentation of SCG signals. A) raw SCG signal, B) filtered SCG signal, C) differentiated and rectified SCG signal with template illustrated by grey area, D) Savitzky-Golay filtered signal, E) match filtered SCG based on template.

Next, a template was defined from the Savitzky-Golay filtered SCG signal (Figure 20 D, indicated by greyed area). The differentiated, rectified and Savitzky-Golay filtered SCG signal was then match filtered using the time inverted and heart rate scaled template, resulting in figure 20, E,

where peaks with lower adjacent peaks were located, resulting in cardiac cycle segmented SCG signals.

The auto segmented SCG based IBI were validated against ECG based IBI using a Bland-Altman plot, based on data with no noise removal from 2 random participants.

3.4 HRV analysis

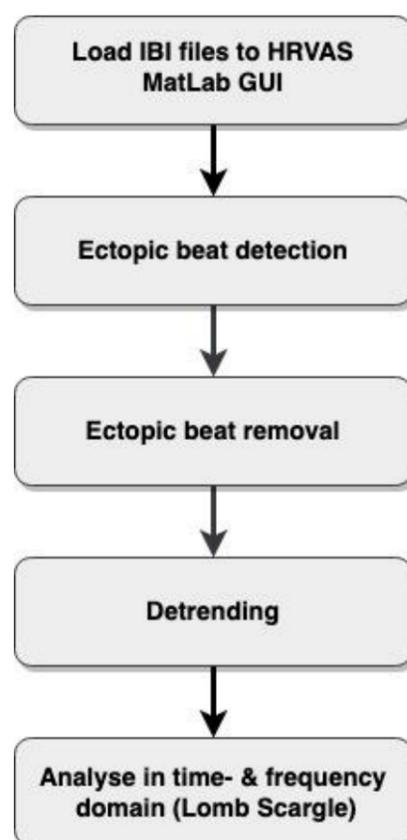


Figure 21: Processing steps involved in HRV analysis.

As illustrated in figure 21, the IBI files were preprocessed to remove ectopic beats, detected by a threshold of 3 standard deviations, as pure removal of IBI has been proved to perform superior to e.g. linear and cubic spline interpolation (Lippman et al., 1994). The IBI are now characterized as Normal-to-Normal intervals (NN-intervals) as any abnormalities in the IBI has been removed. The NN-intervals was furthermore detrended for low frequency trends, using the Wavelet Packet method.

After preprocessing, the signal was analyzed in both the time- and frequency domain. The HRV time domain features included were: average heart rate (MeanHR), standard deviation of NN intervals (SDNN), root mean square of successive NN interval differences (RMSSD), baseline

width of the NN interval histogram (TINN), standard deviation of heart rate (sdHR), and the integral of the density of the NN interval histogram divided by its height (HRVTI). The HRV frequency domain features included were: the absolute power of the low-frequency (0.04–0.15 Hz) band (aLF), absolute power of the high-frequency (0.15–0.4 Hz) band (aHF), absolute power of all frequency bands (aTotal), percentage of the sum of aLF and aHF for the low frequency band (pLF), percentage of the sum of aLF and aHF for the high frequency band (pHF), the ratio of LF-to-HF power (LF/HF ratio), peak frequency of the low-frequency band (Peak LF), peak frequency of the high-frequency band (Peak HF).

Traditional analysis of frequency domain measures (Autoregressive and Fast-Fourier Transform based techniques) requires evenly sampled data, which is naturally not the case for NN-intervals. Also, for the SCG based HRV analysis, removal of noisy areas furthermore contributes to the problem of acquiring evenly sampled NN-intervals. Interpolating and resampling unevenly sampled data can distort power spectral estimates, which can result in loss or distortion of information. (Fonseca et al., 2013). Instead, the Lomb Scargle method was applied as it does not require evenly sampled data.

Since the current studies involves short recordings of 5 min, only the low frequency (LF: 0.04-0.15Hz) and high frequency (HF: 0.15-0.04Hz) bands are of interest, since the very low frequency band (VLF: 0.0033-0.04Hz) requires at least 5 min long recordings (24h is optimal) and the ultra-low frequency band (ULF: 0-0.0033Hz) requires 24 hour recordings. (Shaffer & Ginsberg, 2017). Due to the short length of the recordings, certain time domain measures (SDANN, SDNNI) have furthermore been excluded from the data analysis.

The above mentioned were conducted in a batch process, handling all sequences from all subjects, leading to exported spreadsheets for SCG and ECG recordings, with all time- and frequency domain measures, from which agreement between ECG and SCG based HRV analysis and statistical testing was conducted.

3.5 Statistical analysis

All measures were statistically tested in SPSS, using a two-way ANOVA with repeated measures, followed by a pairwise comparison for load levels, using the Bonferroni correction, with a significance level of $\alpha = 0.05$. This includes the performance measures obtained by the mental task MatLab GUI, subjective ratings from the NASA-TLX, as well as time and frequency domain measures from both the ECG and SCG based HRV analysis. Mauchly's test of sphericity were implemented, and if violated, corrected for, using the Greenhouse Geisser correction.

The agreement between SCG- and ECG based HRV analysis was assessed using the Intraclass Correlation Coefficient (ICC), with a mixed effects ICC with single measures, ICC(3,1). The ICC values were interpreted based on the following guideline adopted from Koo & Li, 2015; $ICC < 0.5$ are indicative of poor agreement, $0.5 < ICC < 0.75$ are indicative of moderate agreement, $0.75 < ICC < 0.9$ are indicative of good agreement and $ICC > 0.9$ are indicative of excellent agreement. ICC values for all HRV measures, used in the statistical analysis, were calculated.

3.6 Machine learning

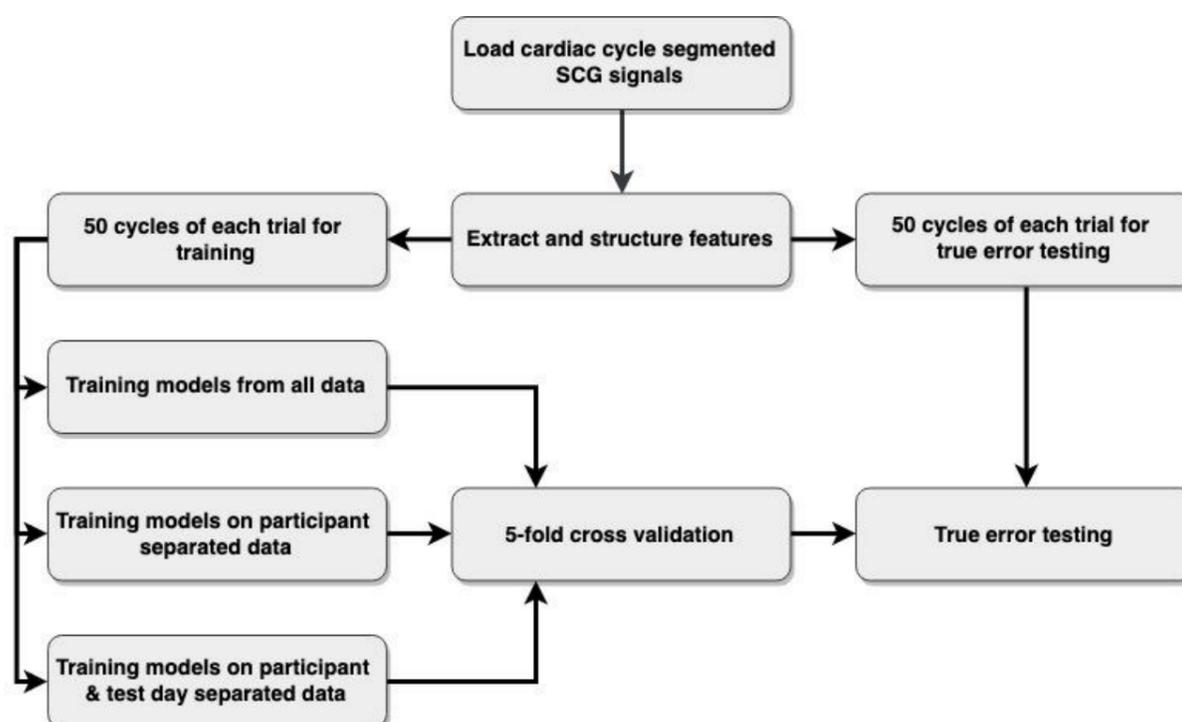


Figure 22: Schematic overview of the steps included to classify MWL levels using machine learning.

Machine learning was implemented to evaluate the ability to discriminate between MWL levels based on the segmented SCG signals as illustrated in Figure 22. Three different approaches to this classification process was selected, meaning that the classification algorithms were trained on the entire dataset for one approach, while the other two approaches were divided into participant and participant on the respective test day respectively. Different parts of the data were allocated for training and test purposes where 50 cardiac cycles were included from each mental workload level, on each respective day, for each participant. This means that each approach to the classification process, all together, participant separated, and participant and test day separated, contained different sample sizes being 3600, 300 and 150 samples respectively.

A total of 22 features were computed for each SCG segmented cardiac cycle including: mean, standard deviation, integral, median, variance, range, skewness, kurtosis, length, RMS, systolic max, location of systolic max, diastolic max, location of diastolic max, time from systolic max to diastolic max, first systolic min occurring before systolic max, location of systolic min, first diastolic min occurring before diastolic max, location of diastolic min, time form systolic min to diastolic min, time from systolic min to max, and time from diastolic min to max. The systolic and diastolic features were found by searching within 40-70% and 70-100% of the signal respectively. Some selected features have been visualized in figure 23.

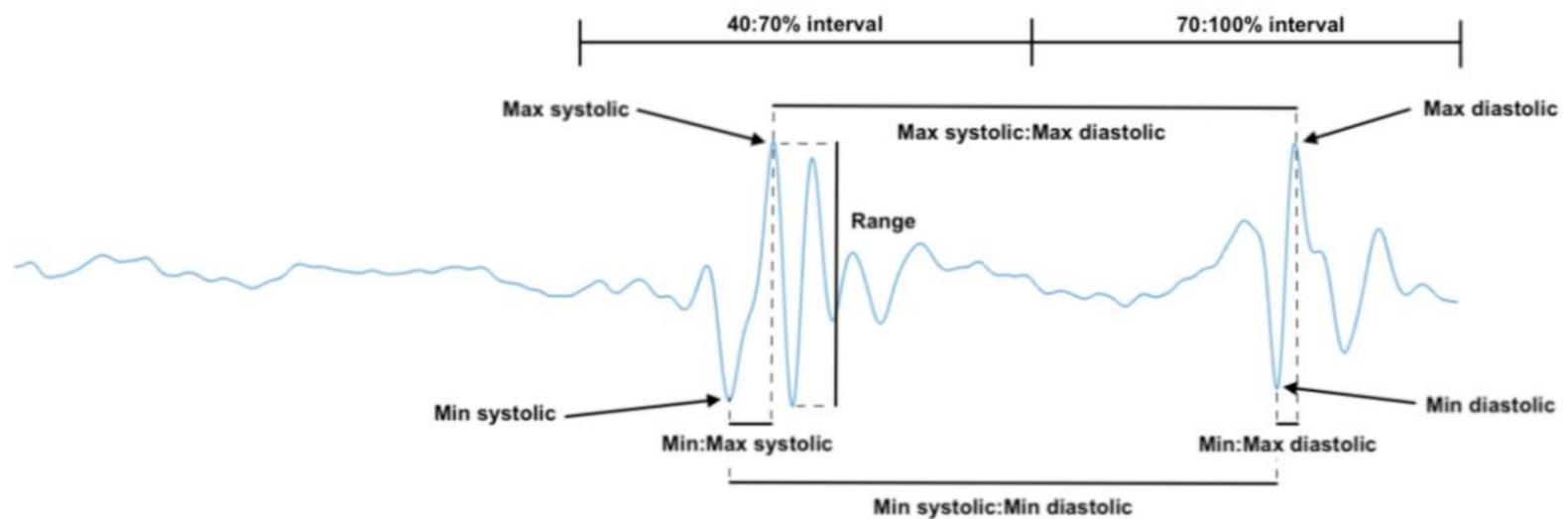


Figure 23: Illustration of certain features of the SCG signal for one cardiac cycle.

All features were rescaled to obtain values in the range 0-1 by equation 4:

$$y_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (4)$$

Where y denotes the scaled version of the feature x . A Principal Component Analysis was performed in order to reduce the dimensionality of the features while retaining any potential inter-class variation (Dougherty, 2013). The number of principal components included for further analysis was based on that the principal components comprised 95% of the total variance (Jackson, 1993). Decision Trees and K-Nearest Neighbor were chosen as the included classification algorithms and trained to predict the mental workload level. The algorithms were validated using a 5-fold cross validation and the superior variant of each classification algorithm were chosen for estimating the true error based on the testing data set.

4. Results

Table 3 provides an overview of the measures (ECG and SCG based HRV, performance measure, subjective rating) computed to assess MWL. The statistical analysis reveal that significant differences were found for ECG based Peak LF, SCG based aHF, aTotal and Peak LF, performance, and subjective rating. The appurtenant p-values of the two-way ANOVA with repeated measures and the post hoc test have been presented in Table 4. These results reveal that ECG and SCG based Peak LF and performance score significantly decreased with increments of MWL, the subjective ratings significantly increased with increments of MWL. SCG based aHF and aTotal and subjective rating significantly decreased while performance score significantly increased from the first to the second test day. Interactions were found for SCG based aHF, aTotal, pLF, pHF and LF/HF ratio and for ECG based aHF. The computed ICC ($\pm 95\%$ CI) between ECG and SCG based HRV measures has been summarized in Table 5. All time domain features show good to excellent correlation, except for HRVTI which showed moderate to good agreement, while the frequency domain features ranged from poor to excellent agreement (Table 5). The cross-validation accuracies and true error of the two included algorithms classifying the three different levels of mental workload has been presented in Table 6. This analysis generally showed poor classification accuracy, while a tendency towards slightly higher accuracies for the participant and day specific approaches occurred. The most superior classification models associated to each different algorithm were coarse decision tree (4 splits) and a medium KNN (10 neighbors). The validation of the cardiac cycle auto segmentation is presented as a Bland-Altman plot between ECG and SCG based IBI (Figure 24). The bias was found to be 0.006ms, upper- and lower limits ± 28 ms and no trend was discovered, representing a good agreement between methods when visually inspecting the location of datapoints in the plot.

		Day 1			Day 2		
		Low	Medium	High	Low	Medium	High
ECG Based HRV Analysis	MeanHR (bpm)	65.40±11.20	66.97±11.21	65.86±11.02	68.62±9.71	69.22±10.10	69.37±11.09
	HRVTI (ms)	10.85±2.17	11.85±1.89	11.36±1.68	11.55±1.27	11.16±1.96	12.11±1.60
	RMSSD (ms)	52.62±20.39	51.98±20.24	51.87±19.08	52.58±27.22	51.87±19.08	47.87±21.63
	TINN (ms)	250.6±104.9	273.9±97.0	263.8±80.8	271.0±155.8	272.0±125.3	267.7±93.7
	sdHR (bpm)	4.927±1.942	5.436±1.956	5.573±2.067	4.955±1.782	5.491±1.653	5.155±1.204
	SDNN (ms)	70.80±29.27	74.09±25.39	75.96±22.90	66.89±30.44	71.22±26.97	67.93±25.00
	aLF (ms ²)	0.008±0.002	0.008±0.002	0.008±0.003	0.009±0.004	0.008±0.004	0.007±0.004
	aHF (ms ²)	0.003±0.002	0.003±0.002	0.005±0.004	0.003±0.001	0.003±0.002	0.002±0.002
	aTotal (ms ²)	0.011±0.003	0.012±0.004	0.013±0.007	0.012±0.005	0.011±0.006	0.009±0.006
	pLF (%)	71.86±12.93	70.57±8.87	69.21±11.91	72.46±9.09	74.65±6.62	76.64±8.17
	pHF (%)	27.36±12.97	27.25±9.88	29.93±11.95	26.43±8.99	24.11±6.55	22.21±8.01
	LF/HF ratio	3.302±1.722	3.041±1.426	2.913±1.863	3.218±1.604	3.360±1.080	4.071±2.126
	Peak LF (Hz) *	0.120±0.000	0.100±0.000	0.089±0.003	0.120±0.000	0.100±0.000	0.090±0.000
	Peak HF (Hz)	0.259±0.068	0.216±0.053	0.245±0.069	0.240±0.064	0.241±0.073	0.228±0.071
SCG Based HRV Analysis	MeanHR (bpm)	65.19±11.27	66.72±11.22	65.46±10.95	68.54±9.79	68.97±10.11	69.26±11.16
	HRVTI (ms)	11.02±2.00	12.15±2.94	10.79±1.97	10.78±1.13	10.72±1.87	11.81±2.40
	RMSSD (ms)	53.94±21.42	52.24±19.05	53.88±19.25	53.79±27.72	53.88±19.25	48.83±20.34
	TINN (ms)	253.8±118.2	267.3±110.4	237.1±81.9	254.6±124.9	257.0±110.4	267.0±122.8
	sdHR (bpm)	4.827±1.941	5.200±1.972	5.282±1.723	4.946±1.949	5.418±1.669	5.009±1.319
	SDNN (ms)	69.60±28.85	71.29±24.50	73.27±21.38	66.39±31.52	70.71±26.85	65.86±24.73
	aLF (ms ²)	0.011±0.003	0.015±0.005	0.016±0.009	0.012±0.006	0.012±0.006	0.013±0.007
	aHF (ms ²)*	0.005±0.003	0.007±0.005	0.009±0.007	0.005±0.004	0.005±0.003	0.004±0.003
	aTotal (ms ²)*	0.016±0.005	0.022±0.009	0.025±0.015	0.018±0.009	0.016±0.009	0.017±0.010
	pLF (%)	70.62±14.43	68.23±9.94	65.95±12.86	70.68±9.59	73.49±7.93	74.59±7.24
	pHF (%)	28.83±14.50	29.30±11.77	33.14±12.33	28.66±9.79	25.52±7.48	24.28±7.14
	LF/HF ratio	3.192±1.812	2.928±1.757	2.456±1.610	2.868±1.359	3.153±1.043	3.347±1.090
	Peak LF (Hz) *	0.120±0.000	0.100±0.006	0.090±0.006	0.119±0.005	0.100±0.004	0.092±0.006
	Peak HF (Hz)	0.266±0.064	0.230±0.044	0.219±0.053	0.234±0.067	0.226±0.071	0.243±0.077
Performance score *	0.512±0.094	0.467±0.090	0.417±0.101	0.560±0.093	0.505±0.107	0.466±0.110	
Nasa TL-X rating *	53.000±17.228	63.212±11.278	69.879±9.347	46.879±14.102	53.091±18.815	64.394±12.704	

Table 3: ECG and SCG based HRV results, performance measures and subjective ratings in mean±1std. Statistically significant differences between one or more groups/days, are marked in bold and with a *.

		One-way ANOVA, repeated measures					Pairwise comparison			
		Day		Load level		Day*Load Level	Low - Med.	Med. - High	Low - High	
ECG Based HRV Analysis	MeanHR (bpm)	$F_{1,10} = 3.2$	$p = 0,105$	$F_{2,20} = 2.3$	$p = 0,128$	$F_{2,20} = 1.3$	$p = 0.286$	-	-	-
	HRVTI (ms)	$F_{1,10} = 0.2$	$p = 0,684$	$F_{2,20} = 0.7$	$p = 0,494$	$F_{2,20} = 1.5$	$p = 0.254$	-	-	-
	RMSSD (ms)	$F_{1,10} = 0.2$	$p = 0,681$	$F_{2,20} = 0.6$	$p = 0,535$	$F_{2,20} = 0.4$	$p = 0.665$	-	-	-
	TINN (ms)	$F_{1,10} = 0.1$	$p = 0,755$	$F_{2,20} = 0.2$	$p = 0,801$	$F_{2,20} = 0.5$	$p = 0.644$	-	-	-
	sdHR (bpm)	$F_{1,10} = 0.4$	$p = 0,539$	$F_{2,20} = 1.9$	$p = 0,171$	$F_{1,3,12,5} = 0.4$	$p = 0.584$	-	-	-
	SDNN (ms)	$F_{1,10} = 1.9$	$p = 0,196$	$F_{2,20} = 0.9$	$p = 0,418$	$F_{2,20} = 0.3$	$p = 0.722$	-	-	-
	aLF (ms ²)	$F_{1,10} = 0.0$	$p = 0,959$	$F_{2,20} = 1.0$	$p = 0,384$	$F_{2,20} = 2.4$	$p = 0.120$	-	-	-
	aHF (ms²)*	$F_{1,10} = 1.6$	$p = 0,238$	$F_{2,20} = 0.2$	$p = 0,839$	$F_{1,3,12,6} = 3.7$	$p = 0.045*$	-	-	-
	aTotal (ms ²)	$F_{1,10} = 0.3$	$p = 0,598$	$F_{2,20} = 0.2$	$p = 0,802$	$F_{2,20} = 3.1$	$p = 0.069$	-	-	-
	pLF (%)	$F_{1,10} = 1.9$	$p = 0,198$	$F_{1,3,12,7} = 0.1$	$p = 0,848$	$F_{2,20} = 2.6$	$p = 0.097$	-	-	-
	pHF (%)	$F_{1,10} = 1.9$	$p = 0,203$	$F_{2,20} = 0.2$	$p = 0,795$	$F_{2,20} = 3.0$	$p = 0.072$	-	-	-
	LF/HF ratio	$F_{1,10} = 1.3$	$p = 0,276$	$F_{2,20} = 0.3$	$p = 0,739$	$F_{2,20} = 2.1$	$p = 0.150$	-	-	-
	Peak LF (Hz) *	$F_{1,10} = 1.0$	$p = 0,341$	$F_{1,10} = 3477$	$p = 0.000*$	$F_{1,10} = 1.0$	$p = 0.341$	$p = 0.000*$	$p = 0.000*$	$p = 0.000*$
	Peak HF (Hz)	$F_{1,10} = 0.0$	$p = 0,84$	$F_{2,20} = 1.4$	$p = 0,269$	$F_{2,20} = 2.6$	$p = 0.100$	-	-	-
SCG Based HRV Analysis	MeanHR (bpm)	$F_{1,10} = 3.4$	$p = 0,096$	$F_{2,20} = 1.9$	$p = 0,183$	$F_{2,20} = 1.7$	$p = 0.210$	-	-	-
	HRVTI (ms)	$F_{1,10} = 0.2$	$p = 0,699$	$F_{2,20} = 0.4$	$p = 0,666$	$F_{2,20} = 3.0$	$p = 0.070$	-	-	-
	RMSSD (ms)	$F_{1,10} = 0.1$	$p = 0,725$	$F_{2,20} = 0.5$	$p = 0,628$	$F_{2,20} = 0.9$	$p = 0.420$	-	-	-
	TINN (ms)	$F_{1,10} = 0.1$	$p = 0,753$	$F_{2,20} = 0.2$	$p = 0,799$	$F_{2,20} = 1.7$	$p = 0.209$	-	-	-
	sdHR (bpm)	$F_{1,10} = 0.0$	$p = 0,877$	$F_{2,20} = 1.2$	$p = 0,309$	$F_{2,20} = 0.5$	$p = 0.592$	-	-	-
	SDNN (ms)	$F_{1,10} = 1.4$	$p = 0,266$	$F_{2,20} = 0.6$	$p = 0,585$	$F_{1,3,13,7} = 0.7$	$p = 0.472$	-	-	-
	aLF (ms ²)	$F_{1,10} = 1.4$	$p = 0,259$	$F_{2,20} = 1.3$	$p = 0,294$	$F_{2,20} = 2.6$	$p = 0.101$	-	-	-
	aHF (ms²)*	$F_{1,10} = 7.5$	$p = 0.021*$	$F_{2,20} = 1.7$	$p = 0.211$	$F_{2,20} = 6.7$	$p = 0.006*$	-	-	-
	aTotal (ms²)*	$F_{1,10} = 9.2$	$p = 0.013*$	$F_{2,20} = 1.7$	$p = 0.217$	$F_{2,20} = 3.8$	$p = 0.039*$	-	-	-
	pLF (%)*	$F_{1,10} = 2.1$	$p = 0,174$	$F_{2,20} = 0.0$	$p = 0,952$	$F_{2,20} = 3.9$	$p = 0.036*$	-	-	-
	pHF (%)*	$F_{1,10} = 1.8$	$p = 0,214$	$F_{2,20} = 0.4$	$p = 0,708$	$F_{2,20} = 5.2$	$p = 0.015*$	-	-	-
	LF/HF ratio*	$F_{1,10} = 0.4$	$p = 0,524$	$F_{2,20} = 0.2$	$p = 0,854$	$F_{2,20} = 4.1$	$p = 0.033*$	-	-	-
	Peak LF (Hz) *	$F_{1,10} = 0.1$	$p = 0,756$	$F_{1,3,12,7} = 111.4$	$p = 0.000*$	$F_{2,20} = 0.4$	$p = 0.698$	$p = 0.000*$	$p = 0.000*$	$p = 0.000*$
	Peak HF (Hz)	$F_{1,10} = 0.1$	$p = 0,812$	$F_{2,20} = 1.1$	$p = 0,361$	$F_{1,2,12,1} = 1.9$	$p = 0.196$	-	-	-
Performance score*	$F_{1,11} = 8.6$	$p = 0.014*$	$F_{2,22} = 34.6$	$p = 0.000*$	$F_{2,22} = 0.3$	$p = 0.713$	$p = 0.013*$	$p = 0.001*$	$p = 0.000*$	
Nasa TL-X rating*	$F_{1,10} = 9.1$	$p = 0.013*$	$F_{2,20} = 26$	$p = 0.000*$	$F_{2,20} = 0.8$	$p = 0.449$	$p = 0.011*$	$p = 0.007*$	$p = 0.000*$	

Table 4: F and p values of two-way ANOVA with repeated measures tests, with pairwise comparison (Bonferroni) for load levels. Bold font and * indicates significant difference.

SCG & ECG agreement		
Variable	ICC correlation	95% CI
MeanHR (bpm)	0,999	0.998-1
HRVTI (ms)	0,684	0.532-0.794
RMSSD (ms)	0,986	0.975-0.992
TINN (ms)	0,919	0.869-0.950
sdHR (bpm)	0,978	0.958-0.988
SDNN (ms)	0,987	0.975-0.993
aLF (ms ²)	0,226	-0.036-0.458
aHF (ms ²)	0,347	0.047-0.574
aTotal (ms ²)	0,210	-0.038-0.436
pLF (%)	0,901	0.807-0.945
pHF (%)	0,903	0.803-0.948
LF/HF ratio	0,814	0.695-0.887
Peak LF (Hz)	0,927	0.884-0.955
Peak HF (Hz)	0,670	0.511-0.784

Table 5: Two-way mixed single measures Intraclass Correlation Coefficient (ICC) measuring absolute agreement, between ECG- & SCG based HRV analysis.

Cross validation accuracy		
Data sample	Coarse decision tree	Medium KNN
Single day	48.1%	47.9%
Participant	45.4%	46.1%
All	34.5%	37.3%

True error		
Data sample	Coarse decision tree	Medium KNN
Single day	37.3%	38.8%
Participant	36.8%	37.1%
All	35.4%	34.9%

Table 6: Classification accuracies in percentage (%) from the initial 5-fold cross validation, and true error testing of trained models fed with new data.

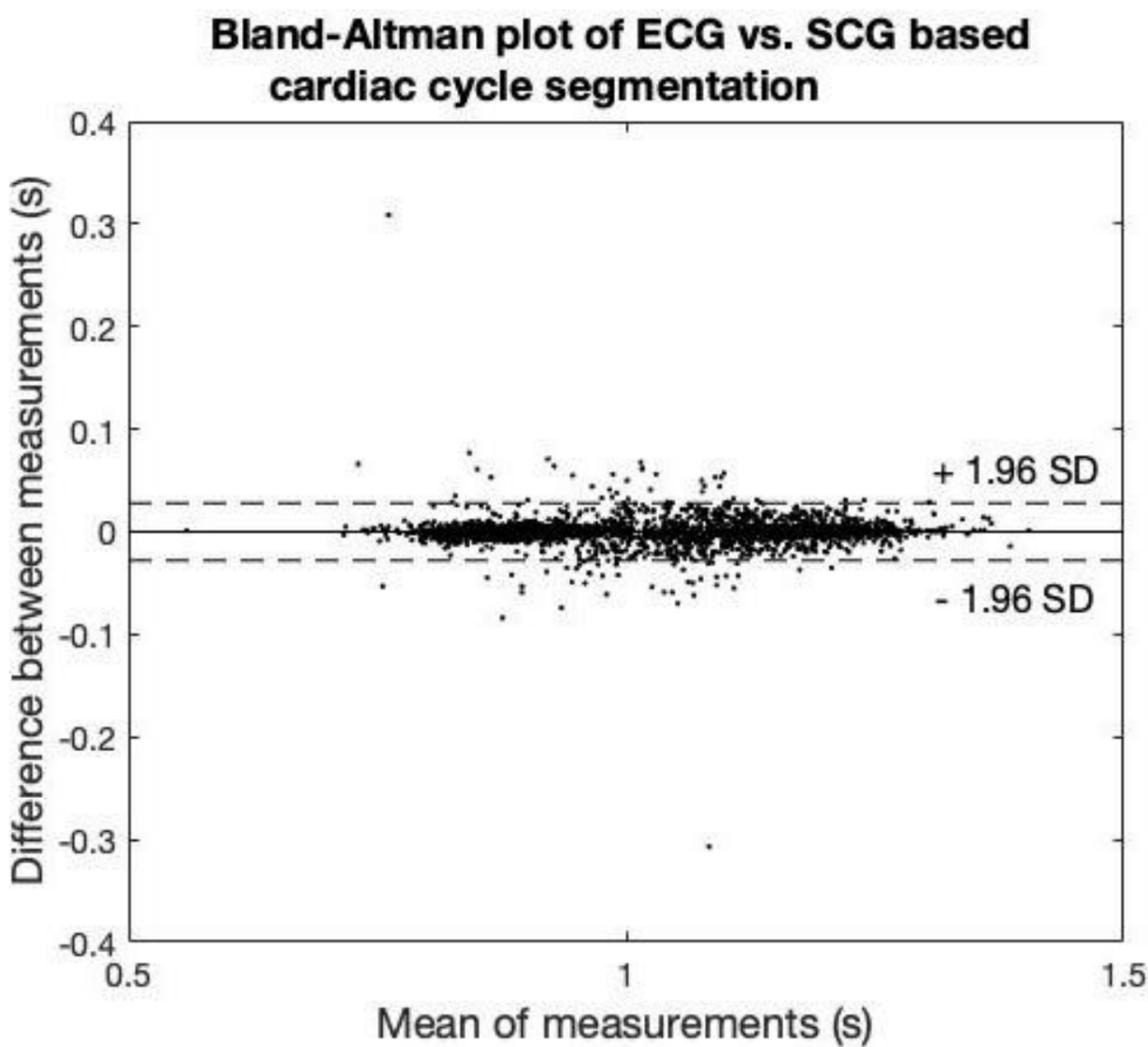


Figure 24: Bland-Altman plot illustrating agreement between ECG and SCG based IBI, n=2750.

5. References

Birch, Laila; Arendt-Nielsen, Lars; Graven-Nielsen, Thomas; Christensen, Hannse, 2001. *An investigation of how acute muscle pain modulates performance during computer work with digitizer and puck*, Applied Ergonomics, Vol. 32, pp. 281-286.

Cain, Brad, 2007. *A Review of the Mental Workload Literature*. Defense Research and Development, Toronto, Canada.

Cinaz, Burcu; Arnrich, Bert; Marca, Roberto La; Tröster, Gerhard, 2013. *Monitoring of mental workload levels during an everyday life office-work scenario*. Personal and Ubiquitous Computing, Vol. 17, Iss. 2, pp. 229-239.

Chambers, Ryan; Gabbett, Tim J.; Cole, Michael H.; Beard, Adam, 2015. *The Use of Wearable Microsensors to Quantify Sport-Specific Movements*, Sports Medicine, Vol. 45, pp. 1065-1081.

Charles, Rebecca L.; Nixon, Jim, 2019. *Measuring mental workload using physiological measures: A systematic review*. Applied Ergonomics, Vol. 79, pp. 221-232.

Crawford, Jacqui; Doherty, Linda, 2011. *Practical Aspects of ECG Recording*. M&K Update Ltd.

Dinh, A., 2011, *Design of a Seismocardiography Using Tri-Axial Accelerometer Embedded with Electrocardiogram*, Proceedings of the World Congress on Engineering and Computer Science 2011 Vol II

Dougherty, Geoff, 2013. *Pattern Recognition and Classification*, Springer International Publishing, pp. 1-7, 9-26, 123-141, 143, 157-176.

Fonseca, D. S.; D'Affonsêca Netto, A.; Ferreira, R. B.; Miranda de Sá, A. M., 2013, *Lomb-scargle periodogram applied to heart rate variability study*, 2013 ISSNIP Biosignals and Biorobotics Conference: Biosignals and Robotics for Better and Safer Living (BRC), DOI: 10.1109/BRC.2013.6487524

Galy, Edith; Cariou, Magali; Mélan, Claudine, 2012. *What is the relationship between mental workload factors and cognitive load types?*. International Journal of Psychophysiology, Vol. 83, pp. 269-275.

Inan, O.T.; Migeotte, P.F.; Park, K.S.; Etemadi, S.; Tavakolian, K.; Casanella, R.; Zanetti, J.; Tank, J.; Funtova, I.; Prisk, G.K.; Rienzo, M.D., 2015, *Ballistocardiography and Seismocardiography: A Review of Recent Advances*, IEEE journal of biomedical and health informatics, VOL. 19, NO. 4, JULY 2015

Jackson, Donald A., 1993. *Stopping Rules in Principal Component Analysis: A Comparison of Heuristical and Statistical Approaches*, Ecology, Vol. 74, No. 8, pp. 2204-2214.

Koo, Terry K.; Li, Mae Y., 2015. *A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research*, Journal of Chiropractic Medicine, Vol. 15, pp. 155-163.

Kubat, Miroslav, 2017. *An Introduction to Machine Learning*, Springer International Publishing, 2nd edition, pp. 1-16, 43-62, 65-87, 113-133.

Lippman, N.; Stein, K.M.; Lerman, B.B., 1994. *Comparison of methods for removal of ectopy in measurement of heart rate variability*, The American Journal of Physiology, Vol. 267, pp. H411-H418.

Marandi, Ramtin Z.; Madeleine, Pascal; Omland, Øyvind; Vuillerme, Nicolas; Samani, Afshin, 2018A, *Reliability of Oculometrics During a Mentally Demanding Task in Young and Old Adults*. IEEE Access, Vol. 6, pp. 17500-17517.

Marandi R. Z.; Madeleine P.; Vuillerme N.; Samani A., 2018B, *Heart Rate Monitoring for the Detection of Changes in Mental Demands During Computer Work*, World Congress on Medical Physics and Biomedical Engineering 2018, IFMBE Proceedings 68/2, https://doi.org/10.1007/978-981-10-9038-7_69

NASA, 1986. *NASA TASK LOAD INDEX (TLX) v. 1.0*. Human Performance Research Group, NASA Ames Research Center.

OpenStax, 2013, *Anatomy and Physiology*, OpenStax CNX

<https://opentextbc.ca/anatomyandphysiology/>

Last seen 03/06/2019

Pan, J.; Tompkins, W., 1985, *A Real-Time QRS Detection Algorithm*, IEEE Transactions on Biomedical Engineering, VOL. BME-32, NO. 3, pp. 230-236.

Richter, M.; Wright, A. R., 2013, *Autonomous Nervous System (ANS)*, Encyclopedia of Behavioral Medicine, Springer, New York, pp. 165-168, <https://doi.org/10.1007/978-1-4419-1005-9>

Salerno, D.M.; Zanetti, J., 1990, *Seismocardiography: A new technique for recording cardiac vibrations. Concept, method, and initial observations*, Journal of Cardiovascular Technology 9(2): 111-118 9. 111-118.

Samani, Afshin; Holtermann, Andreas; Sogaard, Karen; Madeleine, Pascal, 2009. *Active pauses induce more variable electromyographic pattern of the trapezius muscle activity during computer work*, Journal of Electromyography and Kinesiology, Vol. 19, Iss. 6, pp. E430-e437.

Sedghamiz, 2018. *BioSigKit: A Matlab Toolbox and Interface for Analysis of BioSignals*. Journal of Open Source Software, 3(30), 671, <https://doi.org/10.21105/joss.00671>

Shaffer, Fred; McCraty, Rollin; Zerr, Christopher L., 2014. *A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability*. Frontiers in psychology, Vol. 5, Article 1040.

Shaffer, F.; Ginsberg, J.P., 2017, *An Overview of Heart Rate Variability Metrics and Norms*, Front. Public Health 5:258, doi: 10.3389/fpubh.2017.00258

Taelman, Joachim; Vandeput, Steven; Vlemincx, Elke; Spaepen, Arthur; Huffel, Sabine Van, 2011. *Instantaneous changes in heart rate regulation due to mental load in simulated office work*. European Journal of Applied Physiology, Vol. 111, pp. 1497-1505.

Tamura, Toshiyo; Chen, Wenxi, 2018. *Seamless Healthcare Monitoring: Advancements in Wearable, Attachable, and Invisible Devices*, Springer International Publishing AG, Cham, Switzerland.

Yang, Che-Chang; Hsu, Yeh-Liang, 2010. *A Review of Accelerometry-Based Wearable Motion Detectors for Physical Activity Monitoring*. Sensors, Vol. 10, pp. 7772-7788.

Young, Mark S.; Brookhuis, Karel A.; Wickens, Christopher D.; Hancock, Peter A., 2015. *State of science: mental workload in ergonomics*. Ergonomics, Vol. 58, pp. 1-17.

Young, Mark S.; Stanton, N. A., 2005. *Mental workload*. Handbook of Human Factors and Ergonomics Methods, Chap. 39.

Appendix 1

School of Medicine and Health, AAU



Participant declaration of consent

Assessment of mental workload using seismocardiography - a machine learning approach.

Please make sure you have read and understood the following, before signing the declaration. If anything is unclear, please ask for further elaboration.

Protocol: During the experiment, electrocardiography (ECG) and seismocardiography (SCG) will be recorded during your completion of a mental demanding tasks. This means that five electrodes and one accelerometer will be attached to your skin on your upper body, using self adhesive material. This means that if needed, chest hair will be shaved off in these specific spots. These measures will provide health data regarding your heart function, however we do not have the expertise to estimate your health and the well-being of your heart. You will furthermore be asked to evaluate your own mental workload.

All electronic equipment will be connected through an electrical fuse box, to ensure no shortening and electric shocks can occur.

Terms: Your participation in this research is entirely voluntary. It is your choice whether to participate or not. You may change your mind later and stop participating even if you agreed upon earlier.

The collected data (performance measures, ECG and SCG) will be stored safely on a local hard drive, until no later than the 1st of July 2019, after which all data will be deleted.

No compensation for your participation will be provided.

It is the responsible researchers and not Aalborg University who takes responsibility for handling your personal data.

Researchers and contact info:

Jacob Guy Diemar
Mail: jdiema14@student.aau.dk
Phone: +45 23 42 47 84

Mikkel Jul Hansen
Mail: mjha14@student.aau.dk
Phone: +45 25 30 47 92

By signing this declaration of consent, i agree to the above mentioned terms, and are well aware that i can withdraw from the study at any given time.

Date: _____ Signature: _____