Tired?

Electrodermal Responses in Vigilance Tasks using the Example of Tire Inspection at Apollo Vredestein

Patrick Porschen (s0118745)

Bachelor Thesis
Human Factors
Faculty of Behavioral Sciences

Supervisors:
Matthijs Noordzij
Rob van der Lubbe

Abstract

This study set out to examine the relationship between performance development and electrodermal activity (EDA) during a vigilance task. EDA responses were recorded at two locations, the palm and the wrist, with Q-Sensors. 15 student participants were instructed to observe moving arrays of lines for approximately 40 minutes of time. Their watch contained 16 deviant lines, which served as signals and had to be reported by pressing keyboard buttons. It was anticipated that a characteristic decline in performance, commonly referred to as vigilance decrement, would appear. This was not the case when detection performance of signals was considered. Reaction times decreased as a function of time on task. The reason for these changes in reaction times could not be ascertained by this study. Given the scarcity of supporting evidence for a vigilance decrement, a relationship of said phenomenon to the EDA measures could not be established. EDA was found to be more reactive at the palm location. Observed developments of the EDA recordings could be attributed to task or participant characteristics. Future iterations of the same experiment with changed parameters might be more successful in creating the fundamental feature of this study: a vigilance decrement.

Table of content

Introduction

Curiosity no longer is a sufficient impetus for scientific endeavours. Through the years, scientists had to leave the ivory tower and fit their ideas into the economic structures prospering around them. As a consequence, science, apart from the very formal brace of empiricism, has to face the rules of commerce in order to make its impact. Consequently, one could assume that any experiment would fit the original research questions a little better, if it was not constraint by the need to optimize its saleability. In line with this thinking, it will be a vital aspect of this writing to reduce this trade-off while trying to tackle an actual problem: the decrement in vigilance of tire-inspectors at Vredestein. It is assumed that a fairly new device, the Q-Sensor, could be beneficial in finding solutions to this problem by providing us with measures of electrodermal activity (EDA). At first, the problem itself shall be described and the concepts of vigilance and EDA will be considered. Then, scientific evidence will be reviewed in order to create a meaningful, yet attractive task, which is expected to forge a bridge between the relevant industry and scientific standards.

Vredestein and Vigilance

Vredestein, a manufacturer of a diversity of tires located in Enschede, the Netherlands, highly depends on the application of quality assurance in order to provide their customers with a safe and satisfying experience. In fact, small deviations from production standards can cause major problems when products eventually face real life demands. In the case of Vredestein, the quality of the surface textures has to be regarded as one predictor for later on-road safety of the manufactured tires. These textures can display problematic features which vary in size, place and appearance. Hence, they can be described as rather diverse in nature. This very lack of a clear cut problem-space makes human actors an integral part in identifying deficient products. Although the ability to tackle fuzzy problems with plausibility and swift learning makes us a good fit for detection tasks, performance of the workforce can be impeded by several factors. Unfortunately, no definite data is available for the omission rates of flawed products at Vredestein. Parasuraman (1986) derived an average omission rate of 30% [σ = 15.6%] from a multitude of industrial inspection tasks, while Harris (1966) reported rates as high as 80% for complex electronic equipment. These numbers make the case for scientific investigations into the very factors affecting the efficiency of the relevant workforce. Any potential solution would benefit from well-grounded psycho-physiological measurements, such as EDA, to evaluate and substantiate its progress and success.

In order to identify the relevant tires at Vredenstein, an X-ray technique is applied to produce pictures which facilitate a multilayered view of the objects. Subsequently, these scans are inspected by the workers based on the quality standards established by Vredestein. As production success is defined by the uniformity of tires, the control procedure implies that the workforce repetitively confronts a set of homogenous stimuli over and over again. The low frequency of flawed objects provides few opportunities for detection and, by that, low reinforcement to workers for keeping their attention at a high level. A task like this, requiring long periods of sustained attention to identify any number of critical events, is commonly referred to as vigilance task. In general, this situation poses high demands on the mental resources of the workforce, leading to a decline in attention and, consequently, in performance. It is easy to see how this effect, also known as vigilance decrement, can lead to undesired effects at Vredestein: if defective tires hit the market, the probability of harmful events increases. In order to examine the underlying factors, it is necessary to develop strategies and techniques for making occurrences of vigilance decrement quantifiable and tangible. Therefore, it is purpose of this study to explore the possible application of electrodermal activity (EDA) measures in vigilance task settings, considering the example of the Vredestein procedure.

A pilot laboratory study carried out by Eelke de Jonge in 2012 did not indicate a relationship between the EDA-responses of participants and subjective workload. Derived from the classical visual search task of Neisser (1964), subjects were asked to identify the letter Z which had been placed in various lists of 150 letters. During the experiment the Q-Sensors picked up their EDA-response. The experiment consisted of six blocks, each 10 minutes long, and asked subjects to rate their mental workload on multiple occasions during the testing. While it was evident that the perceived workload increased as a function of time, the EDA measures displayed no reciprocity. As it appears more plausible that the nature of the experiment lead to these results than the inadequacy of the EDA-measure, the study at hand will apply a modified task which is assumed to model the Vredestein task more accurately.

Attention and Vigilance

In her writing "Visual Attention: The past 25 years" Marisa Carraco (2011, p. 1486) describes attention as "a selective process, which is usually conceptualized as being related to limited cognitive and brain resources." Focusing on the visual aspects of perception, the spotlight metaphor of attention might serve as an adequate, yet intuitive access into the topic. Here, attention can be regarded as a torch to be moved across the perceivable, leaving irrelevant aspects in the dark. The early ideas of William James (1890) characterized two different mechanisms of attention: an involuntary part, nowadays commonly referred to as exogenous or transient attention, and a voluntary part, usually known as endogenous or sustained attention. In this line of thought, individuals are capable of deliberately controlling their focus, while also being prone to automatic and involuntary orienting responses, when confronted with unexpected stimuli. As Carrasco (2011) reports, this distinction still holds today as it is supported by a large body of scientific evidence.

In the light of vigilance tasks, it appears evident that sustained attention is an essential condition for on-task success.

Mackworth (1948) pioneered the field of vigilance research by instructing people to observe a prepared clock for extended periods of time and indicate whenever a second was skipped by the clockhand. Throughout the scientific history of vigilance studies, two main perspectives emerged, trying to hypothesize the origin of the vigilance decrement. Quite contrastive, one of them emphasizes the role of mental "underload", while the other one stresses the impact of mental "overload".

Proponents of the "underload" framework can be found in Manly et al. (1998) who examined the relevant phenomenon assuming that the decline in performance is due to a mindless state of the individual. They designed a go-no-go task, known as the Sustained Attention to Response Test (SART), which was intended to promote automatic processing in subjects. Reversing the classic vigilance scenario, participants were instructed to respond to non-targets and withhold their response when targets appeared. Reaction time served as a measure for the level of automatic processing: as the reaction schemata become well-established, reaction times decrease. This was often the case in intervals preceding a relevant stimulation. By this, they argue, failures to detect the relevant stimuli are a consequence of automatic processing which, in turn, is associated with mindlessness. In tradition with this line of thought, Smallwood et al. (2004) used the SART framework to examine the relationship between task unrelated thoughts and performance. Task unrelated thoughts, operationalised as "the frequency with which an individual's attention departs from the current situation" (Smallwood et al., 2004, p. 2) and conceptually related to mindlessness, was found to be associated with the vigilance situation. Detached from the SART scenario, Pattyn et al. (2008) argue in support of an "underload" view, based upon psychophysical measures and subjective reports of the participants during a 1.5 hours vigil.

The role of subjective evaluations in form of interviews and scales can be regarded as a source of the "underload-overload"-controversy. The aforementioned body of evidence draws extensively on measures as the Cognitive Failure Questionnaire (CFQ) by Broadbent et al. (1982), which is thought to identify individuals prone to display absent mindedness. In this test, subjects have to respond to questions such as "Do you daydream when you ought to be listening to something?" as a basis for the actual scoring. On the other hand, as Helton et al. (2005) report, we find a convolute of research utilizing workload measures to stress the importance of mental overload in the vigilance context. The NASA Task Load Index (NASA-TLX) can be regarded as a prominent instrument to assess the workload individuals perceive in a given situation or task. Helton et al. (2005) state that there is well-funded evidence that workload-scores obtained during a vigil surpass scores of other cognitively demanding tasks. For example, Warm, Dember and Hancock (1996) used the instrument at hand to indicate high mental workloads in relevant situations. In response to Manly et al.'s (1998) SART research, Grier et al. (2003) were able to indicate high levels of workload in exactly the same task situation. In line with this research, Warm, Parasuraman and Matthews (2008) argue that these findings ought to be interpreted in the framework of attentional resource theory. That is to say that a drainage of the re-energizing capacities of an attentional supervisory system leads to a decline in detection perfomance. Whether this system has to be regarded as a single resource or a system of multiple resources that deplete remains a matter of debate.

Many authors acknowledge several problems tied to the use of subjective measures (Pattyn et al., 2008; Helton et al., 2005). While all measures (measures of workload as well as measures of boredom and mindlessness) might be regarded as valid tools on their own, a legitimate question could be raised targeting the very heart of the debate. Do we have to regard an "underload" and "overload" view of the topic as truly dichotomous in nature? One might argue that both phenomena are representations of at least two very different cognitive aspects affected at the very same time during a vigilance scenario. As a consequence, one might consider the possibility that the quality of a subject's report is primed by the investigativ direction of the research at hand. While this stresses the importance of identifying meaningful psychophysical measures, it also reminds us of possible pitfalls of circular argumentation while validating them by means of subjective reports.

Electrodermal Activity

The reasons why one should opt for measures of electrodermal activity in an industrial setting are manifold. Discovered in the last decades of 19^{th} century, scientists immediately recognized its potential in serving as a psychophysical measure. This has led to a rather clear understanding of the concepts which are at work during measurement and a large research body to draw on. In addition to this, new devices as the Q-Sensor developed by Affectiva, which has the size of a matchbox, are easily applied and nearly unobtrusive; a quality that does not hold for many other investigation techniques.

Electrodermal activity is captured by measuring skin conductance while applying a very mild current between two electrodes placed on the skin. The electrical conductivity, expressed in the unit of Siemens (S), is the reciprocal of electrical resistance (R) usually expressed in the unit of Ohm (Ω). As a result, Ohm's law, describing the relationship between resistance (R), voltage (V) and current (I), serves as the conceptual basis for the phenomenon at hand. When currents are applied to different materials, different amounts of energy or work [here: voltage] are needed to make the current "move" through the conductive matter. In the case of skin, the efficiency of carrying current is relative to the amount of sweat held by its sweat glands. This is due to the fact that sweat, with its high amounts of dissolved salt, conducts the current significantly better than the tissue itself. Besides its well-known thermo-regulative capacities, a close connection to the sympathetic branch of the autonomic nervous system has been identified as the crucial factor determining the actual activity of the sweat glands and, hence, the overall conductivity of the skin (Wallin, 1981). Responses of the sympathetic nervous system, also known as flight-or-fight responses, such as an increase in heart rate and blood-flow, are widely regarded as preparative measures of the body to engage in physical and often emotionally involving activities. Multiple cerebrations have been linked to the regulation of sweat gland activity (Boucsein, 1992). Here, correlations between prefrontal cortical activity and changes in EDA have to be regarded as the most relevant observation for the topic at hand due to its involvement in the regulation of sustained attention (Wilkins, Shallice & McCarthy, 1987). In essence, it is this sympathetic mechanism that serves as the conceptual background for the experimental framework at hand. When subjects are confronted with sensory and/or especially emotionally meaningful stimuli, shifts in the level of skin conductivity can be observed. These phasic increases in conductivity are referred to as skin conductance responses (SCRs) which are projected onto the absolute level of conductance, called the skin conductance level (SCL). This SCL can be regarded as the tidal trend over time of the measure at hand and commonly achieves values between 2 μS and 20μS (Dawson, Schell & Filion, 2000), whereas the superimposed SCR amplitudes typically amounts to 1-3μS. Additionally, non-specific SCRs (NS-SCRs) constitute a third characteristic of EDA measurements which accounts for approximately 1-3 SCRs per minute. As the name implies, this status is assigned to SCRs when no specific trigger for the reaction can be identified. These different aspects of EDA protocols constitute the background for different research paradigms.

Traditionally, palms and fingers were regarded as the most appropriate loci for taking EDA measurements, which was explained in the light of the high densities of sweat glands in this region; a crucial factor for electrodermal reactivity (Freedman et al., 1994). Searching for an explanation for this phenomenon, it was reasoned by Darrow (1937) that this palmar sweating response and the high density of glands are related to a better grip on objects. As these places cannot be regarded as suitable choices for real life recordings, van Dooren, de Vries and Janssen (2012) set out to identify locations which are less affected by everyday activities. They took measurements at 16 bodyparts during a movie screening and identified the feet and shoulders as equally appropriate spots for EDA recordings.

Research indicated that people differ in their magnitude of EDA responses. Two different groups of individuals could be identified by Lacey and Lacey (1958), which were characterized due the magnitude of their EDA responses. Subjects which displayed higher levels of EDA fluctuations were labelled EDA labile whereas less reactive people are considered EDA stabile. It was observed that EDA labile individuals are likely to react more rapidly in vigilance settings, but are more prone to failure.

Given the Vredenstein problem and its ergonomic nature, it does not appear necessary to engage in endeavours targeting the precise mechanisms underlying the vigilance decrement. From a more pragmatic and solution-driven point of view, searching for possible correlations in psychophysical measures and performance breakdowns could suffice in finding and evaluating ways set out to stabilize detection rates.

It should be the goal to create a situation which supports the applicability of the Q-Sensor in the Vredestein scenario. It is obvious that the most promising conceptualizations have to draw on the plethora of research which is available. All relevant variables have been reviewed extensively while trying to establish a vigilance taxonomy (Teichner, 1974; Parasuraman, 1986), leaving us with a rich toolbox to work with. While recombining the building blocks of the experiment, however, one has to keep in mind that a plausible set up which resembles the de facto situation at the workplace in many ways would increase the benignity of all non-scientists involved. Accordingly, this study sets out to create a task which unifies these facets in order to create a "product" which complies with industrial as well as scientific requirements. Given an adequate task, the experiment will seek to answer the following research questions:

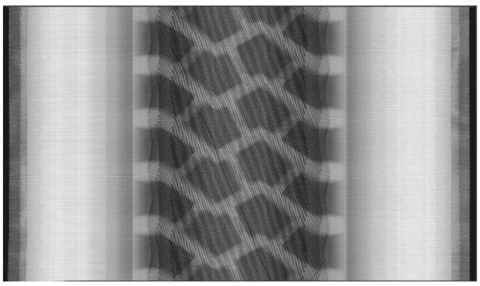
Research Question 1: Does a vigilance decrement, operationalised as a change in reaction time and/or a decrease in detection performance, occur on the relevant task?
Research Question 2: If this is the case, is it possible to observe psycho-physiological correlates to these aspects of performance by applying an electrodermal activity measurement with the Q-Sensor?

Methods

Designing the task

Just as de Jonge's research (2012) did not provide an indication of vigilance decrement and its correlates, Parasuraman (1986, p. 24) states that "in any situation requiring sustained performance on a task, there may or may not be a vigilance decrement". Given this uncertainty, it is evident that one should opt for the most conservative approaches when it comes to determining the relevant variables. Consequently, in order to increase the probability of success, the experiment is designed along the lines of other experiments and guidelines, which have already indicated a vigilance decrement, taking parts and variables of other experiments into consideration which appear to be the best fit for the purpose at hand. This follows from the intention to design closely to the actual task and, by that, to recreate the quite unique setting. At this time, there is no task available which can be regarded as an adequate representation of this setting.

As indicated before, workers at Vredestein repetitively face similar X-ray images trying to find tires which do not comply with production standards. In light of vigilance research, tasks which require an individual to identify a signal (the defect) against a non-signal background (the tire) are regarded as simultaneous-discrimination tasks (Parasuraman & Mouloua, 1987). In contrast to this, tasks which provide the individual with a signal or non-signal stimulus separately are called successive-discrimination tasks. Parasuraman and Mouloua (1987) point out one main difference regarding their potential to evoke a vigilance decrement: as successive-discrimination tasks pose higher demands on short-term memory, they tend to be more reliable in producing the relevant responses. At this point, one could argue that de Jonge's (2012) simultaneous-discrimination task did not challenge participants enough to show a decrement over time. In order to compensate for this, one has to draw on properties like stimulus representation rate and/or stimulus discriminability to make the task more challenging.



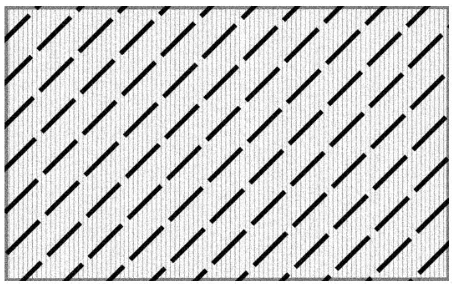


Fig. 1 Left: An X-ray image as used in quality control. Right: Non-signal stimulus used in the experiment.

The structure of the experiment was developed along the lines of Teichner's (1974) review of vigilance methodologies. In his writing, he strongly recommends the use of pre-test measures as referral point for the participants' possibly decremental performance curve. This means that a short measurement of the initial percentage of detection needs to be implemented.

The length of the relevant testing phase, which immediately follows the pre-test, was also determined. His analysis reveals that for signals with an initial detection rate above 40% the decremental effect vanishes after 35 minutes. At this moment, no further decline in detection performance could be observed across all 37 experiments reviewed. Half of the final loss was reached after 15 minutes. Consequently, it could be argued that a length of 10-20 minutes would suffice in indicating and investigating the vigilance decrement. This is not the case for at least two reasons. First, it is the goal of this study to examine the electrodermal response during the watch and to describe the relation between performance loss and the relevant psycho-physiological aspects. Hence, a full description needs to assess the full range of the phenomenon which is only covert by a 35 minute vigil. Secondly, Teichner's decrement functions are expressed as the best fitting line of several experiments. As a consequence, it appears reasonable to expect some variation from these values.

It can be argued that a modified version of a line length comparison task developed by Parasuraman and Mouloua (1987) can be regarded as a visually adequate representation of the Vredestein task; this is indicated in Figure 1. For the purpose at hand, a full-screen version of the task has to be created which conserves the visual-search aspect of the original task. Accordingly, the new scenario requires participants to observe a moving grid of lines and identify "defective" lines amongst them. These defective lines are approximately doubled in length and can appear at virtually every location on screen. The movement of the grid is achieved by looping 6 basic frames into a simple animation which carries the lines diagonally across the screen. This dynamic presentation of the stimuli prevents a "pop-out-effect" of the longer lines which would decrease task-difficulty significantly in a static display (Treisman & Gelade, 1980).

It is evident that the visual search adds complexity to the task. As a consequence, Parasuraman and Mouloua's (1987) original parameters like a stimulus presentation time of 150ms cannot be projected onto the modified scenario. Therefore, a presentation time of 2.6s (23 frames per minute) was determined as adequately challenging on the basis of trial and error. 920 frames are displayed throughout the whole experiment, 120 during the pre-test and 800 during the testing phase. The according time-frames amounted to 5:12min (=2.6s x 120) and 34:40min (=2.6s x 800), resulting in an overall time of 39:52min of effective testing.

During the pre-test, 10 defective lines (signal rate= 8.33%) were displayed, while 18 (signal rate= 2.25%) could be detected throughout the testing phase. A higher signal rate in the pre-test implies a more accurate estimate of task-difficulty. Out of the 18 defective lines, only 16 will be considered in the analysis. Defects closer to the middle of the screen may be detected quicker and more often than targets in the periphery. To correct for this, all 16 considered targets are located in the periphery with approximately the same distance from the centre. Two defects were still located in the centre of the screen to convey the impression of spatial randomness.

For analytical purposes, the testing condition consists of four blocks, each containing 200 frames and 4 defects. These defects appear at the same time in each block but on different locations. It is assumed that participants will not recognize this scheme and the impression of an unsystematic session with randomly occurring defects is conveyed.

Reactions in the first 300ms after stimulus onset were not registered by the software in order to prevent carry-over responses from the previous display.

Due to the high contrast of background (white) and bars (black), a Herman grid illusion (Herman, 1870) could be observed between the lines. In order to prevent subjects from this annoying phenomenon, grey lines were added to the background to reduce contrast and successfully diminished the issue.

Participants

8 male and 7 female participants, aged between 22 and 28 years (M=24), were tested in separated chambers at the library of the University of Twente. As no incentives were available, motivation for participation has to be characterized as a voluntary act of friendship. All of the participants followed programmes of higher education, had normal or corrected-to-normal vision and indicated an appropriate level of wakefulness in the beginning of the experiment by self-assessment. As caffeine could influence detection performance, participants were asked whether the consumed caffeine in a period of 2 hours prior to the experiment. For three subjects this was the case. All other participants did not consume any caffeine the whole day.

Apparatus

The experiment was presented on a 14” laptop [Intel i5-2410@2,3GHz; 4GB RAM] running Open Sesame 0.26 in Windows 7 as stimuli presentation and response collecting software. Consequently, the behavioural data was logged via the internal keyboard of the laptop. The EDA-responses were registered by two different Q-Sensors per participant, which are identical in function but were placed at different locations. All sensors were worn at the left hand. The Q-Sensor Pod collected data from the palm of the subject, while the Q-Sensor Curve was placed on their wrists, also facing in the palmar direction. Both sensors were held in place by purpose-built elastic bands. No discomfort was reported by participants during the experiment. The wireless design is powered by a lithium polymer accumulator which has a capacity of 130mAh. Data was stored on the internal flash memory with the highest possible sampling rate of 32Hz. Both data sets were synchronised by logging both, the behavioural and EDA-data, in relation to the system time.

Procedure

The experiment took place in the silenced section of the library of the University of Twente. To create an environment which provides few opportunities for external distraction, participants were seated in separated chambers. The only object in the participants' field of vision was the laptop running the experiment. Before the actual experiment started, subjects were equipped with the Q-Sensors. They had to perform a warming-up procedure (i.e. climb stairs), which ensured proper conductivity between skin and the electrodes of the sensor by providing initial moisturising.

Right after the warm up, subjects were seated in front of the testing device. The initial explanations were followed by two short illustrations of defective and normal arrays. Participants were encouraged to indicate whether they understood the nature of the experiment or not. After the instruction, the pre-test was administered. Participants had to press the L-button if the array was normal or the P-button if the array was defective. Responses were given with the index and middle finger of the right hand. The pre-test was followed by a black screen which informed participants that the actual experimental session was about to begin and that they had to prepare themselves for a 40 minute long period of continuous testing. Additionally, it was explicitly stated that the experiment has a definite ending and was not a test of the participants' willpower or endurance. The left hand had to be kept still.

Results

Behavioural data

Detection Performance. The first research question relates the presence of a vigilance decrement to the measurements of detection performance and reaction time. The detection performance of participants shall be considered first. When all 800 stimuli from the main testing scenario are considered, the overall proportion of correct responses, considering signals and non-signals, amounts to 87.8%. The worst three performers exhibit a hit rate of 77% while the best reached a 98% hit rate. When the data was segmented into four intervals of 200 stimuli, a Chi

Square Test indicated no significant differences in overall performance [X^2 (3, N = 11700) = 7.005, p = .072].

When the detection of signals, i.e. defective arrays, is considered, 55% of the targets were hit during the pre-test. Detection performance spread from 12.5% to 100% with a standard deviation of 25.8%. During the testing, performance dropped to an average hit rate of 36.7% with a minimum detection rate of 6.3% and a maximum of 75% (σ = 25.9%). For analytical purposes, data was divided into four 9 minute segments. Each segment contained 4 signals. The mean hit rates for the four segments across all 15 participants developed as follows: 38.3%; 36.7%; 31.7%; 40%. A Chi Square Test did not indicate significant differences between these segments [X^2 (3,N = 240) = 1.005, p = .8]. A visual inspection of individual detection rates indicated no clear pattern of a declining performance for any individual. What is evident, however, are great differences in individual performances. The average hit rate amounted to 5.5 signals identified per session with σ = 4.14. On basis of this distribution, scores equal or above 9 hits are referred to as good performances for the remainder of this writing. Five participants identified more than 9 signals correctly, whereas seven individuals hit 3 or less signals. 2 hits constitute the mode of the distribution. Hit rates were equally distributed across genders. It is noteworthy that participants 10, 11 and 12 drank coffee prior to the experiment. Those participants scored approximately one standard deviation above average.

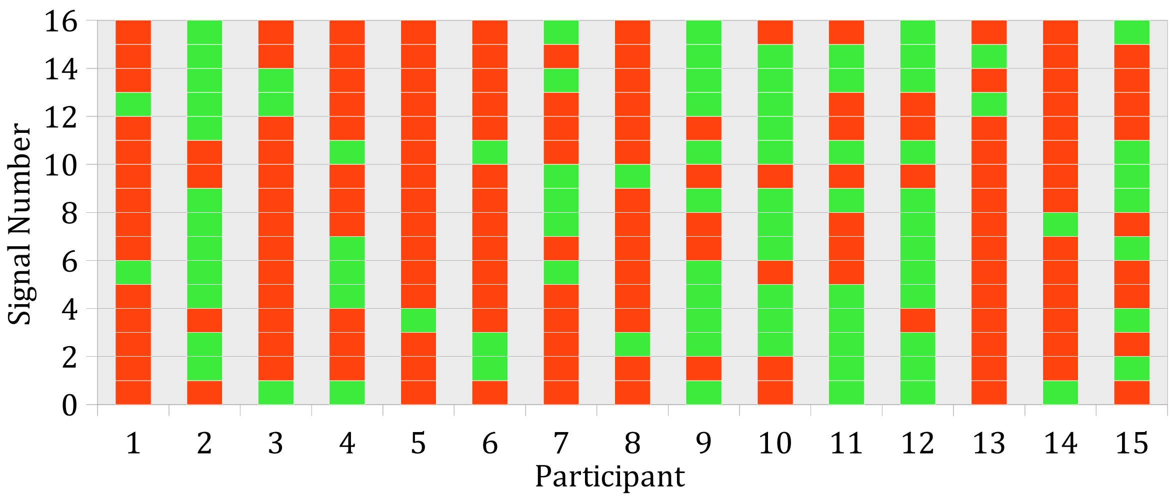


Fig. 2 Visualisation of hits (green) and misses (red) for all participants and all detection opportunities.

Reaction Times. A repeated measures ANOVA was applied to all participants reaction times recorded during the experiment. 800 data points per participant were split into 6 blocks of 6 minutes. Malchy's test of sphericity indicated a violation of the assumption of sphericity [X^2 (5) = 27,32, $p < .01$]. Hence, a Greenhouse-Geisser correction was applied to adjust the degrees of freedom ($\varepsilon = .457$). Results implied no significant differences between blocks [$F(1.37, 19.184) = .993$, $p = .359$]

For a graphical representation, reaction times were averaged per 10 responses. Visual inspection of the average reaction time of all participants plotted against time on task revealed a dampened oscillating pattern, which stabilizes approximately at stimulus 560, i.e. minute 25. From then, the curve lacks any observable homogeneity. The maxima of single oscillations are highly predictable and can be described by the estimated function $f(x) = 1417.627 - .437x$ with an R^2-value of .95. Reaction times fluctuate around a mean of 1160ms ($\sigma = 325.9$).

An inspection of participants' curves does not affirm that this declining trend does hold for all individual measures. The development of reaction times throughout the experiment was rather diverse. In four cases, reaction times increases, as indicated by a regression coefficient greater than .2. Four participants displayed relatively constant reactions ($.2 > \beta > -.2$), and seven developments were negative in nature ($-.2 > \beta$). Gender did not predict these results. Performance on the experiment did not correlate significantly with reaction times ($r = .26$). A property that is preserved throughout the individual measures is the oscillation, although less predictable in nature. To attain an estimate of the frequency, intersections with the mean value lines were counted. The average count of intersections amounted to 35.2 ($\sigma = 6.1$). Average amplitudes were assessed by computing the standard deviations of individual reaction times. These amounted to 315.13 ($\sigma = 80.34$).

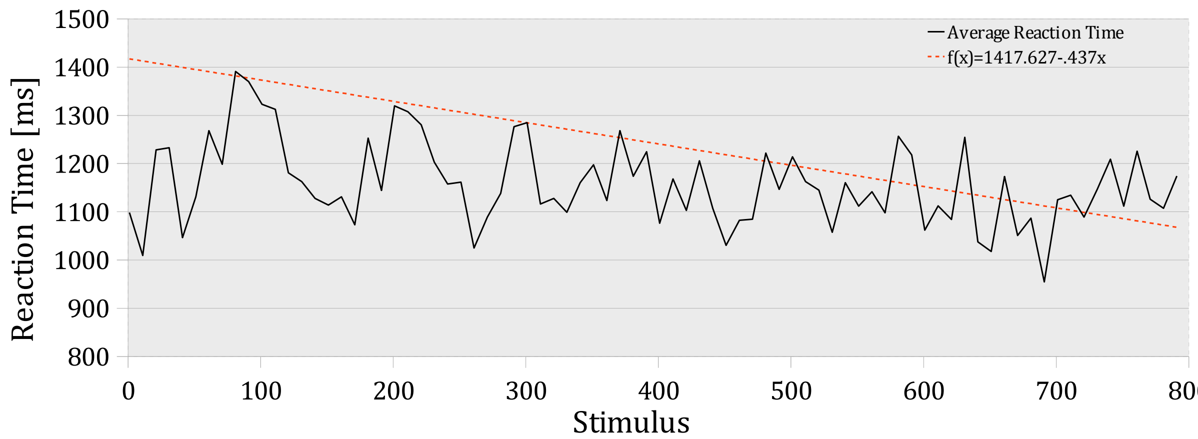


Fig. 3 Grand mean of participants' reaction times.

Further visual inspection indicated a relationship between performance on task and reaction times. A correlation coefficient of .72 confirmed the assumption that more hits are related to longer reaction times. Additionally, the plotted reaction times imply a relationship between performance and the distribution of values around the mean. Explicitly, bad performers appear to be far more inconsistent in their response times, expressed by higher standard deviations of their reaction times. A correlation analysis provides statistical support for this observation. On average, a better performance lead to lower standard deviations of the 15 reaction times ($r = -.74$).

EDA-Data

Data Preparation. The EDA data was processed in MATLAB applying functionalities of the simpleEDA package, which was specifically developed for the evaluation of electrodermal and muscular activity (Schleicher, 2005). In essence, the applied scripts were specified to detect all SCRs per participant. Additionally, the average amplitudes per minute were computed.

The approach draws on two necessities in order to characterise a course of the EDA-curve as a SCR. Firstly, the slope, mathematically utilised as the first derivation of the Q-Sensor data, has to exceed a pre-defined value, which is determined iteratively. When this condition is met, these points of interest are examined in the Q-Sensor EDA-curve. If the detected aberrations exceed an amplitude of 0.01µS they are classified as an SCR. If any number of SCRs is detected within an interval of 700ms, only one is added to the count. Determining the relevant values for the slope has to be regarded as an iterative process which can be improvemed by visually checking each automatically detected SCR due to their resemblence to the prototypical SCR shape (Boucsein, 2012). This check was skipped due to time constraints in the bachelor's thesis.

The consecutive data analysis took place in IBM SPSS Statistics 20. One subject was excluded from the analysis because no data was recorded by the Q-Sensor.

Number of SCRs. The grand mean of SCRs counts was plotted for the wrist and the palm location (Fig. 4). It is evident that measurements conducted at the palm ($M = 7.01$) display more SCRs than the wrist recordings ($M = 0.44$). Additionally, palm recordings appear to be more reactive, i.e. display a less monotone curve shape. The curve tendency is identical for both graphs in the first 20 minutes of the experiment. A sharp decline in the first minutes is followed by a period of oscillation within a range of 2 SCRs. While the wrist curve holds steady, the number of SCRs at the palm location increases and peaks at minute 25. Then, the number declines steadily.

For the analysis, the data was grouped into 6 blocks of 6 minutes. A repeated measures ANOVA, considering location of measurement (2 levels) and the number of SCRs per block (6 levels), implies significant differences between blocks [$F(5, 13) = 3.23$, $p = .011$]. Bonferroni corrected, pairwise comparisons identified block 1 as the major source of variance. The impression that the Q-Sensor registered more SCR when mounted at the palm is confirmed [$F(1, 13) = 70.36$, $p < .001$] with an average difference of 33.99 SCRs.

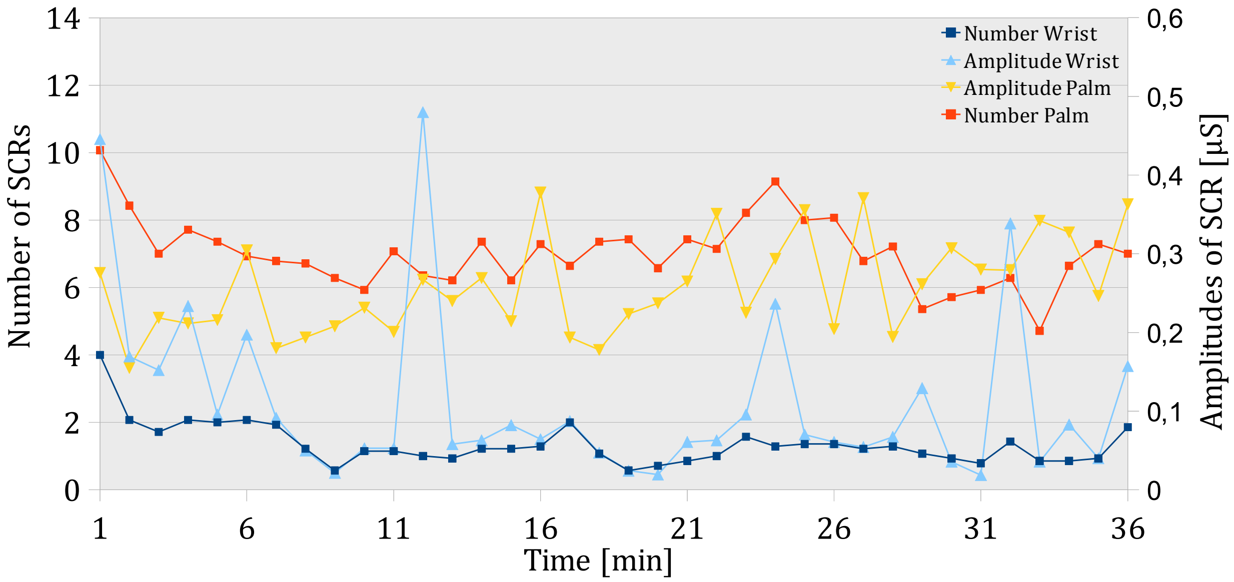


Fig. 4 Grand mean of the number of SCRs and their amplitudes recorded at the wrist and palm locations.

An investigation of the individual palm recordings implied observable differences among participants. SCR counts ranged from an average 2.5 per minute to 11 per minute.

A k-means cluster considering the absolute count of SCRs per participant identified two cluster centres (172.29 SCRs; 333 SCRs) with group sizes of 7 and 7 participants. Further inspection of the distribution of SCR counts across individuals does not prompt a strict grouping of participants as implied by the k-means cluster. A stem and leaf plot of the relevant data reveals that 6 participants' total SCR counts are captured by the 300 leafs, 4 by 200, 3 by 100 and 1 by 0.

1,00	0 .	9
3,00	1 .	149
4,00	2 .	1225
6,00	3 .	024559

Fig. 5 Stem leaf plot displaying the distribution of SCR counts at the wrist location.

The number of recorded SCRs is not correlated to the number of identified signals ($r = 0.11$, $n = 14$,

p = .97) and not correlated to participants' reaction times ($r = 0.2$, $n = 14$, $p = .5$).
Furthermore, the visual inspection of the plotted individual data per minute did not reveal any clear structure underlying all recordings. The individual curves do not resemble or express the overall trends observed in the grand mean SCR curve for the palm. All responses collected indicated some reaction. With a $\beta > -.15$, a curve estimation indicated a slight negative tendency for most graphs (10 out of 14) of the palm recordings in the course of time. Comparing the standard deviations of the individual SCR counts suggests that the reaction range is comparatively stable across participants amounting to approximately 3 SCR. Neither the number of SCR nor the recorded profiles differentiated good performers from the rest of the group. A paired-samples t-test was conducted investigating the relationship between the average number of SCRs during minutes of signal appearance and absence. The testing did not yield significant results for the minutes of signal appearance ($M = 1.26$, $\sigma = .42$) and signal absence ($M = 1.43$, $\sigma = .79$) at the wrist [$t(17) = -1.02$, $p = .33$]. Additionally, no significant difference between minutes of signal appearance ($M = 6.89$, $\sigma = 1.01$) and signal absence ($M = 7.15$, $\sigma = 1.05$) was indicated at the palmar location [$t(17) = -.96$, $p = .35$]. An ANOVA was used to examine the number of SCRs per minute while a stimulus was presented and detected and while a stimulus was presented and not detected. Testing for homogeneity of variances revealed that this assumption was not sufficiently met, as indicated by Levene's test [wrist location ($p < .01$); palmar location ($p = .07$)]. A Welch test indicated significant differences between both conditions for the wrist sensor [Welch's $F(1, 161.69) = 4.94$, $p = .03$], but not for the palm sensor [Welch's $F(1, 169.55) = .66$, $p = .42$]. However, it to should be noted that the absolute difference in numbers of SCRs reached similar dimensions (wrist = .523 SCRs per minute, palm = .403 SCRs per minute).

Amplitudes of SCRs. The grand means of SCR amplitudes were plotted for the wrist and the palm location (Fig. 4)**.** The recordings at the palm ($x = .257$) appear to be higher than those conducted at the wrist ($x = .112$). In the course of the experiment, SCR amplitudes at the palm increase from approximately .23 to .30, while the wrist location amplitudes decrease from approximately .23 to .07. A repeated measures ANOVA with 2 levels, considering the two locations of measurement and the six blocks, was applied to the amplitudes of the SCRs. Malchy's test of sphericity indicated a violation of the assumption of sphericity [X^2 (14) = 28.71, $p = .013$]. Hence, a Greenhouse-Geisser correction was applied to adjust the degrees of freedom ($\varepsilon = .534$). Results implied no significant differences between blocks [$F(2.668, 13) = .87$, $p = .454$]. The observed difference between recording locations was validated [$F(1,13) = 10.99$, $p < .01$).

When amplitudes of the palm measurements are plotted, it is evident that eleven participants show a comparable reaction range between .2μS and .8μS. Three participants' amplitudes did seldom exceed a value of .2µS.

A paired-samples t-test was conducted investigating the relationship between the average amplitudes of SCRs during minutes of signal appearance and absence. The testing did not yield significant results for the minutes of signal appearance (M = .12, σ = .11) and signal absence (M = .11, σ = .11) at the wrist [t(17) =.17, p = .87]. Contrary, a significant difference between minutes of signal appearance (M = 27, σ = .62) and signal absence (M = .24, σ = .57) was indicated at the palmar location [t(17) = 2.175, p = .04]. The impact of signal detection when a stimulus was presented was tested with an ANOVA. Testing for homogeneity of variances revealed that this assumption was not sufficiently met, as indicated by Levene's test [wrist location (p = .02); palmar location (p = .03)]. A Welch test indicated significant differences for the wrist sensor [Welch's F(1, 239.815) = 4.7, p = .03] but not for the palm sensor [Welch's F(1, 115,367) = 1.39, p = .24].

Corrected Data Analysis

As seen above, single individuals can impose a great risk to analytical procedures if their influence is disproportionate due to extreme values. Therefore, a z-standardisation was applied to all SCR-related measures.

The implications for the number of SCRs are not affected by the modulation. Still, a significant difference amongst blocks is indicated by a multivariate repeated measures ANOVA [F(5,13)=4.961, p<.01]. Again, the first block could be identified as a major source od variation b y means of a Bonferroni corrected, pairwise comparison.

Another repeated measures ANOVA examinig differences in amplitudes of the SCR across the 6 blocks was carries out. Malchy's test of sphericity indicated a violation of the assumption of sphericity [X^2 (14) = 26.39, p = .026]. As a result, the Greenhouse-Geisser correction was applied to adjust the degrees of freedom (ε = .655). No significant differences were indicated for the 6 blocks [F(3.276,13) = 2.04, p=.117].

Discussion

On first sight, the task at hand appears to be rather undemanding: find a deviant line amongst other lines. Anecdotally, many participants held the assumption that the task would be boring and unchallenging in nature before they participated in the experiment. Only few attributed a exhaustive aspect to the situation to come. This changed dramatically after the 40 minutes long session, when reactions underscoring the “cruelty” and demanding aspect of the vigilance scenario were quite common. Although not captured methodologically, microsleep and losing track of time appeared to be prevalent occurrences amongst subjects. The question remains, however, to what extend these sensations impacted task-performance and are captured by the data and applied recordings.

Following the example of Mackworth (1948), the presence of a vigilance decrement is operationalised as the mere count of hits or misses as the participant spends time on the task. In the present case, this conservative approach yielded no clear results supporting claims of a decrement over time. The sharp drop between the hit rates during the pre-test and the test phase are most likely an artefact of task design, more specifically signal frequency (pre-test signal-rate = 8.33% against testing signal rate = 2.25%), and not relevant to the research question. It can be argued, however, that a pre-test detection rate of 55% for signals is rather low; maybe too low to produce meaningful results in the context of vigilance. Detection percentages immediately drop to 30-40% percent when signals are presented less often; a number associated with no decremental occurrences in the review of Teichner (1974). But even when only individuals are considered who identified more than 60% of the defective arrays during the first 9 minutes, no decline in performance can be observed. Unfortunately, the proportion of these four good performers is too small to draw any valid conclusions.

It is evident that most variable values had to be determined by intuition for this very specific new task. It was already anticipated while designing the task that the study at hand would most likely fail on its first iteration. The taxonomy of vigilance research can provide us with a scientifically grounded way to calibrate the experiment at hand. Consequently, task difficulty ought to be decreased by adjusting variables as stimulus presentation time, line lengths or signal rate.

Addressing the differences in individual performances, caffeine could influence detection performance in a positive way. For example, Temple et al. (2000) were able to observe a positive effect of caffeine in a 12 minute vigilance task. Obviously, the present study did not focus on caffeine as an explanatory variable. Neither the precise amounts consumed by participants were assessed nor the definite time when consumption occurred. Due to the known effects and the setting in which the testing took place (a university library), it would have been unreasonable to ignore caffeine as a possible confounding variable. Consumers perform clearly above average. Hence, it might be sensible to control for this variable more stringent or exclude caffeine from the setting completely for testing purposes in future research.

A more fundamental issue when it comes to on-task performance are motivational aspects. Motivation might be regarded as a major contributor to between- and within-subjects differences in preserving one's attentional focus. This was already acknowledged in Teichner's (1974) writing and Parasunaram's (1986) review of vigilance research. It is trivial that individuals who lack the motivation to sustain their attention will be more prone to task unrelated thoughts or other distractions and, in consequence, will make more errors. On a conceptual level, it is rather hard to identify the causal network between motivation and other cognitive aspects. Explicitly, a lag of motivation may be caused by a depletion of energetic resources of the attentional apparatus or the other way around. Answering this question of causal direction seems to be of special importance in real world settings such as Vredestein. Different causes for bad performance might call for different interventions. Psychophysical measures as the Q-Sensor appear to be promising candidates to eradicate this ambiguity. Further exploration of the device at hand could link distinctive EDA-profiles to different cognitive processes or states.

However, on the between-subject level, it is evident that the performance distribution amongst subjects is rather disruptive. Five people performed approximately one standard deviation ($\sigma = 4.14$) above average ($M = 5.5$), while five people scored two hits or less. In line with signal detection theory, it can be argued that performance should approach a normal distribution, if mental capacities constituted the major determinant. It is conceivable that the phlegmatic attitudes some students display in university experimental settings is a contributing factor to this disruption. The combination of experimental apathy and an exhausting task could amplify this effect. Quite contrary, motivation can be assumed to be sustained to some degree by factors as legal obligations of the workforce and possible job loss in case of failure at Vredestein. Therefore, one should consider the introduction of incentives and feedback as a motivational aid in order to create a more valid testing framework. Any measure of motivation appears essential in this situation. Possible influences of the labile – stabile dichotomie on task performance will be reviewed later in this writing.

The research question conjectures that reaction times might display characteristics which could be interpreted as some form of vigilance decrement. In contrast to detection performance which is given by the mere count of successes, the interpretation of reaction times poses more of a challenge. While the dichotomous nature of failure and success only supports one conclusion, it cannot be said that a longer or shorter reaction time is better or worse per se. For example, a shorter reaction time may reflect learning effects or some form of "letting go" mentality. Reaction times do not seem to be correlated to task performance. Hence, there is no indication of a global speed accuracy trade-off related to the changes in reaction times.

The most prominent feature is constituted by the oscillating patterns of all individual recordings and the grand mean. In a first approach of explaining this phenomenon, one has to consider that a repetitive pattern like this might actually be an artefact of any software malfunction or flawed experimental set up. Open Sesame was rather new when the experiment was carried out.

Combined with the rather repetitive experimental set-up, a systematic problem could possibly be manifested as the pattern at hand. As there is no reference value for the frequencies of the oscillations, it is hard to determine empirically whether the outcome of the experiment can be interpreted in the light of such. Intuitively, however, the number of oscillations across participants appears to be relatively constant, increasing the plausibility of a software problem. In order to eliminate this explanation, it would be advisable for any rerun of the experiment to implement any form of supplementary measure of reaction time.

Accepting the premise that a software problem is of no importance to the outcome, the points in time when important stimuli are presented can be regarded as second explanation. One can assume that attention ceases while no signals are presented, giving rise to a more automated state of mind. In this state, subjects would be inclined to commit button presses more rapidly. Hence, the task could be regarded as a constant fluctuation between sensitization and habituation. When a signal is detected, participants' perceptual sensitivity is increased due to the unexpected stimulus and arrays are scanned more thoroughly. Consequently, reaction time would be increased. When there are no signal presented for a longer period, excitation of the relevant structures would decrease and lead to a habituation process increasing reaction times. If this explanation was correct, it would imply that a floor effect occurs at some point, stabilising reaction times at an individual minimum. This is not represented in the recordings. It is possible that signal occurrences prevent participants from reaching that floor or that the explanation at hand is not valid. Additionally, one would expect that these reactions appear in synchronicity with signal presentation. This is clearly not the case as the peaks of individual amplitudes do not temporally match each other and wave periods are rather diverse.

While it is not necessarily the case that the underlying methodology is a key determinant in producing these pattern, it could also be a merely individual phenomenon which manifests as a consequence of the testing situation. One could imagine a task which merely instructs participants to press a button whenever a signal appears, without any emphasis on quick reactions. The Vredestein task could have evolved into a "press-as-you-will"-scenario for somewhat unmotivated individuals. In consequence, the recordings would display some form of inner pacemaker depending loosely on the actual happening on screen. A part of the results indicated that better performances were associated with lower variability, i.e. smaller amplitudes of reaction times. It could be argued that the four good performers suppressed this pace making by motivation or other intellectual capabilities. At this point, it becomes evident that any solution to the problem is merely speculative. In order to come to a well-grounded, empiric resolution of the phenomenon at hand more evidence has to be collected. The observation that the grand mean develops negatively as a function of time implies that there is *some* development as the experiment progresses. As already mentioned, both explanations, the underload as well as the overload hypothesis, provide a reasonable framework for this phenomenon. A less intensive investigation of the arrays consumes a smaller amount of time. The research question did not relate the presence of a vigilance decrement to any specific result. Hence, it could be argued that the outcomes at hand at least justify a further investigation of the described phenomena.

So far, it has to be concluded that the first research question was not answered as definite as it might be needed in order to develop a deeper understanding of the relationship between a vigilance decrement and EDA measures. This implies a general investigation of said signals with less focus on the situation.

Examining the recordings of the two Q-sensors, the claim that the palmar locations provide higher resoluted recordings than the wrist location is confirmed. In general, palm EDA is more reactive, displays more SCRs and higher amplitudes. It was already mentioned that the high density of sweat glands at the inside of the hand may be the reason for this effect (Freedman et al., 1994). The grand means of both recordings develop identical for most of the experiment. The first 9 - 10 minutes after the beginning of the experiment, the number of SCRs drops by approximately 4 SCRs. It could be argued that this is due to a carryover effect induced by the pretest and/or the anticipation of the long session to come, which was mentioned again between pre-test and test. The individual palmar recordings of 9 participants displayed this huge drop in the first 10 minutes of the experiment. Quite contrary, a similar study by Munro et al. (1987) showed an incline in SCR number shortly after the experiment begins which is followed by a steady decline throughout the rest of the testing. De Jonge's results (2012), on the other hand, are in accord with the latter progression of the recordings at hand but not with the sharp decline at the beginning. Consequently, it is hard to interpret the results in the light of these works.

Literature prompted the expectation that two groups, EDA-labile and EDA-stabile, would be identifiable based on the subjects' reactivity. Regarding SCR counts, this does not seem to be the case as the distribution does not imply a strict grouping. De Jonge (2012) reported a similar result. Additionally, neither the deviation nor the decline of SCRs counts seems to differentiate individuals clearly from one another. The relevant framework implies a relationship between the number of SCRs and task performance. A correlation between these variables could not be identified. Contrary to this observation, two different profiles can be identified regarding the SCR amplitudes recorded at palm location. One group displayed a rather constant reaction range between .2µS and .8µS, while the other group remained below a .2µS threshold for the whole experiment. Again, the grouping could not be related to the observed differences in performance. It is obvious that the small sample size might have influenced the results, regarding SCR count as well as SCR amplitudes. Furthermore, the distribution of factors like age, social status and intellect cannot be projected onto the population in general. One can assume that selection bias has to be considered as an influence in student research. Considering the target group for future applications of this technology at Vredestein, it is advisable not only to enlarge the sample size, but to diversify the pool of subjects, too.

Blakeslee (1979) observed an orienting response in EDA recordings as a result of detecting relevant information, expressed as an increase in EDA amplitudes. Examining four datasets, i.e. the number of SCRs and their amplitudes at the palm and the wrist location, lead to somewhat mixed results. More specifically, an effect could be observed only at the wrist sensor. As indicated before, this location was shown to be far less reactive than the palmar counterpart, leaving us with inconclusive results. It can be argued, though, that the results of this analysis crucially depend on the way time is segmented in preparation for the ANOVA. While EDA data was processed per minute, signals were not exactly matched to these intervals but could have appeared at any second within this minute. As a consequence, it not necessarily the case that the full effect of a signal was captured within this very minute, inducing an uncertain amount of randomness to the analysis. Anyway, the fact that a signal leads to a decrease in amplitude size constitutes a major deviation from the established literature as Blakeslee's writing.

Conclusion

The present study was not designed to grant immediate success. Right from the start, the unification of vigilance research and Q-Sensor measurements could be regarded as a subject that might need several experimental iterations. As literature indicated, a vigilance scenario does not necessarily lead to an observable vigilance decrement. It is trivial that any correlates of a vigilance decrement can only be observed in the presence of the said effect. A vigilance decrement could not be observed when defined as a performance decline in the detection of signals. The present study offers a lot of starting points for methodological modifications along the lines of previous research, like adapting stimulus presentation time or line length of signals. It is evident that designing an adequate task has to precede the methodological integration of devices as the Q-sensor.

It might be the case that in a monotonous tasks setting, reaction times might serve as a predictive measure as well. In order to do so, a deeper understanding of the relationship between motivation, reaction time and task performance has to be developed. Due to the immediate nature of reaction time measures, those could be of equal value as EDA-measures. Measures of reaction time are much easier to conduct and rely on fewer premises than other psycho-physical measures. Consequently, it is possible that less research and, by that, less money would lead to comparable results.

In general, results of previous EDA research have not or only partially been reproduced by this study. As mentioned before, possible confounding factors should be handled more carefully in future research.

As a result, this study cannot be regarded as supportive evidence for the application of the Q-sensor in qualitiy management situations as the Vresdestein.

References

Blakeslee, P. (1979). Attention and vigilance: performance and skin conductance response changes. *Psychophysiology, Vol. 16*(5), p. 413-419.

Boucsein, W. (1992*). Electrodermal Activity*. New York: Plenum.

Broadbent, D. E., Cooper, P., Fitzgerald, P., & Parks, K. (1982). The cognitive failures questionnaire (CFQ) and its correlates. *British Journal of Clinical Psychology, Vol. 21*(1), p. 1-16.

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research, Vol. 51*(13), p. 1484-1525.

Darrow, C.W. (1937). Sensory, secretory, and electrical changes in the skin following bodily excitation*. Journal of Experimental Psychology, Vol. 10*(3), p. 197-226.

Dawson, M. E., Schell, A. M., & Filion, D. L. (2000). The electrodermal system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (Vol. 2, p. 200-223). Cambridge University Press.

De Jonge, E. (2012). *Aandacht gewenst! Een pilot studie naar de verandering in huidgeleiding tijdens vigilantie taken* (Unpublished bachelor's thesis). University of Twente, the Netherlands.

Freedman, L.W., Scerbo, A.S., Dawson, M.E., Raine, A., McClure, W.O., & Venables, P.H. (1994). The relationship of sweat gland count to electrodermal activity. *Psychophysioligy, Vol. 31*(2), p. 196-200.

Grier, R. A., Warm, J. S., Dember, W. N., Matthews, G., Galinsky, T. L., Szalma, J. L., & Parasuraman, R. (2003). The vigilance decrement reflects limitations in effortful attention, not mindlessness. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *Vol. 45*(3), p. 349-359.

Harris, D.H. (1966). Effect of equipment complexity on inspection performance. *Journal of Applied Psychology, Vol. 50*(3), p. 236-237.

Helton, W. S., Hollander, T. D., Warm, J. S., Matthews, G., Dember, W. N., Wallaart, M., Beauchamp, G., Parasuraman, R., & Hancock, P. A. (2005). Signal regularity and the mindlessness model of vigilance. *British Journal of Psychology, Vol. 96*(2), p. 249-261.

Hermann, L. (1870). Eine Erscheinung simultanen Kontrastes. *Pflügers Archiv für die gesamte Physiologie des Menschen und der Tiere, Vol. 3*(1), p. 13-15.

James, W. (1890). The principles of psychology. New York: Henry Holt.

Lacey, J. I., & Lacey, B. C. (1958). The relationship of resting autonomic activity to motor impulsivity. *Research publications - Association for Research in Nervous and Mental Disease, Vol. 36,* p. 144-209.

Mackworth N. H. (1948). The breakdown of vigilance during prolonged visual search. *The Quarterly Journal of Experimental Psychology, Vol. 1*(1), p. 6-21.

Manly, T., Robertson, I. H., Galloway, M., & Hawkins, K. (1999). The absent mind: Further investigations of sustained attention to response. *Neuropsychologia, Vol. 37*(6), p. 661–670.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, Vol. 44*(2), p.

314-324.

Munro, L. L., Dawson, M. E., Schell, A. M., & Sakai, L. M. (1987). Electrodermal lability and rapid vigilance decrement in a degraded stimulus continuous performance task. *Journal of Psychophysiology, Vol. 1*(3), p. 249-257.

Neisser, U. (1964). Visual search. *Scientific American, Vol. 210*(6), p. 94-102.

Pattyn, N., Neyt, X., Henderickx, D., & Soetens, E. (2008). Psychophysiological investigation of vigilance decrement: boredom or cognitive fatigue? *Physiology & Behavior*, *Vol. 93*(1-2), p. 369-378.

Parasuraman, R. (1986). Vigilance, monitoring, and search. In K.R. Boff, L. Kaufman, & J.P. Thomas (Eds), *Handbook of perception and human performance* (Vol. 2, p. 1-39). Oxford, England: John Wiley & Sons.

Parasuraman, R., & Mouloua, M. (1987). Interaction between signal discriminability and task type in vigilance decrement. *Perception & Psychophysics, Vol. 41*(1), p. 17-22.

Smallwood, J., Davies, J. B., Heim, D., Finnigan, F., Sudberry, M., O'Connor R., & Obonsawin, M. (2004). Subjective experience and the attentional lapse: task engagement and disengagement during sustained attention. *Consciousness & Cognition, Vol. 13*(4)*,* p. 657-690.

Teichner, W. H. (1974). The detection of a simple visual signal as a function of time of watch. *Human factors, Vol. 16*(4), p. 339-353.

Temple, J.G., Warm, J.S., Dember, W.N., Jones, K.S., LaGrange, C.M., & Matthews, G. (2000). The effects of signal salience and caffeine on performance, workload, and stress in an abbreviated vigilance task. *Human Factors: The Journal of the Human Factors and Ergonomics Society, Vol. 42*(2), p. 83-194.

Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, Vol. 12*(1), p. 97-136.

Van Dooren, M., De Vries, J.J.G., & Janssen, J. H. (2012). Emotional sweating across the body: Comparing 16 different skin conductance measurement locations. *Physiology & Behavior, Vol. 106*(2), p.298-304.

Wallin, B. G. (1981). Sympathetic nerve activity underlying electrodermal and cardiovascular reactions in man. *Psychophysiology, Vol. 18*(4), p. 470-476.

Warm, J. S., Dember, W. N., & Hancock, P. A. (1996). Vigilance and workload in automated systems. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (p. 183–200). Mahwah, NJ: Erlbaum.

Warm, J. S., Parasuraman, R., & Matthews, G.(2008). Vigilance requires hard mental work and is stressful. *Human Factors, Vol. 50*(3), p. 433-441.

Wilkins, A.J., Shallice, T., & McCarthy, R. (1987). Frontal lesions and sustained attention. *Neurologica, Vol. 25*(2), p. 359-365.