MAUU5900

MASTER THESIS

in

Universal Design of ICT

May 2019

<Universal Design Of Smartphone User interface Interaction in Stressful Situations>

<Saja Andersson>

Department of Computer Science

Faculty of Technology, Art and Design

OSLOMET

Preface

The main target group for this paper is web designers and for those interested in cognitive psychology in relation to human-computer-interaction. The main topic in this paper is cognitive load and stress while interacting with mobile technology and how they affect one and the other. Seeing there is no current solution on a subjective approach in the assessment of stress, this study aims to investigate whether creating an individual stress template can be a solution for a more accurate stress detection, this by allowing the users themselves describe a stressed scenario. Cognitive load is also surveyed when it's initially low and further evaluated when exposed to acute stress. This is mainly due to getting a greater understanding of how these operates with each other but also individually.

Oslo, May 2019

Master's thesis for the Master's programme for Universal Design of ICT, OsloMetropolitan University

Saja Andersson

Abstract

With today's use of digital devices, it's easier to push pressure on our cognitive resources and thereby increases the possibility of an overload, which eventually reaches to a negative impact on our brain. This thesis provides insights on how the overloads occur, yet also how it can be reduced. The amount of our cognitive resources has a limited threshold and when this is exceed, stress occurs. This means that a person might start with a near-exceeded threshold when interacting with a user interface and a cognitive overload can be imposed by a regular website. Therefore it is important for designers and developers to understand how we interact dynamically, how our cognitive processes are affected and how design can contribute to an increased load and stress level. While cognitive load and stress correlates, this study does not assume that these have a compulsory relationship between them or that they can't occur independently. We therefore decided to adopt two different physiological measurements in order to distinguish them. This was done by using Heart Rate Variability (HRV) as an index of stress assessment and by Galvanic Skin Response (GSR) as an index of cognitive load, the latter being designed and developed by us for this project purpose. Previous research has investigated how high values of cognitive load contributes to an increased stress level, yet there are still limitations on how stress affects our cognitive load when it is initially low. That is mainly what this project attempts to address. Stress is also very much based on individual preferences and while similar studies lack this kind of an approach, this study presents an experimental design of creating an individual stress template for this assessment. Studies in this field is also usually limited to a controlled laboratory environment, whereas this project proposes an experimental method which investigates and determine when cognitive load and stress reaches a high/low level, while interacting on mobile interface in a real-life scenario. In our results we discovered that GSR is highly sensitive to various reactions and it often corresponded to the HRV values. The GSR was clearly affected by emotional states such as arousal, laughter and deep sighs but it appeared that the values decreased while the participants stopped interacting and responded alone during interaction in silence and with no emotional response. This result nevertheless indicates that the GSR is affected by a mental effort and while controlled thoroughly, it can be served as an index of cognitive load. For the

HRV, the extracted features mainly produced inconsistent results but by exploring the differences between heartbeats (R-R Interval) and heart rate (HR), a lowered R-R and a higher HR could be linked to the amount of stress level and emotional states. A lower HRV usually means that a person experience more stress and in our results all of the participants had lower HRV during the simulated stress scenario, which indicates that our manipulation of creating a stress template was successful. However, our results also revealed that our stress manipulation was clearly disturbed by a "first impression effect" which affected the GSR values. For this reason we suggest that using a training session before starting the experiment would be more beneficial. We acknowledge that this study has some limitations both in terms of our GSR equipment but also our own interpretation ability of the physiological equipment but we do believe that the current experimental design of this study can be considered as a good starting point and we hope that this thesis can serve as a resource for those conducting further research in this field.

Table of content

Table of figures

1 Introduction

While today's use of technology has increased the ease of our life, it has also increased our way to constantly be available and to always, do more. We are able to respond quicker to emails, we get easier access to necessary information and thanks to our wearable devices, we can get access anytime and anywhere.
With today's use of digital devices, the possibilities of performing more than one task or activity at the same time, or rapidly switch between tasks have never been greater.
It is what we call multitasking. We walk in the street and at the same time we talk on the phone, respond to emails, send text messages, update our social media or check the time table of our next bus or train. Many of today's youth have grown up with digital media and they seem to be multitasking at a much greater frequency than ever before. A recent study discovered that college students were multitasking twice as much compared to information workers, who earlier had the highest frequency of multitasking (Mark, Wang & Niiya (2014). With our use of smartphones we can more easily get access to new information than we were able to 20-30 years ago and just in the last decade, getting access to the internet is almost unlimited.
The general public has become increasingly dependent on the Web to gather information and in 2016 statistics showed that internet usage is being dominated by smartphones (Heisler, 2016). Okoshi et al. (2015) states that there is an explosion of information available for people to respond to which causes a constantly information-seeking need, and further contributes to an information overload. This is due to the fact that the growing amount of information does not correspond with the amount of information that a person's attention resource has available, and when this threshold is exceeded, stress occurs. To handle this we divide our attention and share it between different resources by using push notifications (Okoshi et.al, 2015). Instead of actively seeking, we push new information to us. As for the growing number of mobile and other wearable devices, push notifications can arrive from a variety of applications and services and on multiple applications installed on each device.
Yet, while this approach solves our limited attention resource, the distraction from the excessive number of push notifications and the feeling of being constantly interrupted leads to an interruption overload, which again is a key component towards stress

(Okoshi et.al, 2015). Friedmann (2006) describes this as the “malady of modernity”, where we all are afflicted with chronic multi-tasking and continuous partial attention, induced by the internet and our handheld devices. Mark, Gudith & Klocke (2008) claims that some interruptions may be perceived as beneficial when the actual interruption contributes to the same context of the current task, but as proven by their own study, the cost of the disruptions still remains the same. In their study they found out that whether the interruptions were from similar or different context, the participants still experienced more stress, higher frustration and higher workload (Mark, Gudith & Klocke, 2008). With our constantly increasing information-seeking need it contributes to an interruption overload, which again is a sub-component of information overload and all of this of course ease our way of multitasking which again, increases our stress level. This way of interacting with technology and our belief in multitasking depends not only on how our society is built up, it also has a physiological explanation of how our mind interprets multitasking. This type of behavior activates our reward-seeking centres of the brain and it actually creates a dopamine addiction in our minds, which very effectively make us choose multitasking over and over again (Levitin, 2015). That is why most of us do it and why we might think that we are designed for it, even though we’re not. The increase of stress level has shown to deplete the nutrients in our brain and in long term, it can have deleterious effects upon broad aspects of our cognition (McEwen, Sapolsky, 1995). When we are exposed to stress our brain tells our body to produce cortisol in order to react instantly, this as a part of our biological defence system. If our body produces a high level of cortisol during a long period of time, it alters our memory. As stated by Klemm (2016), when the cortisol binds to the cells it disrupts the memory process and new information will not be stored correctly, meaning that the learning process will also be degraded (Klemm, 2016). Both our memory process and our learning process is by scientists explained as our **cognitive load**. It means the total amount of mental effort being used in the working memory (memory process) and the amount of cognitive resources required for a person to complete a certain task (learning process). These are important processes that can be extremely degraded by the effect of multitasking and stress (Niculescu, Cao & Nijholt, 2009) and (Harris, Vine and Wilson, 2017). When the cognitive load exceeds a person’s cognitive ability, the mental effort increases and it causes mental stress (Lyu, Luo, Miao, Wang, Shi & Kameyama, 2015).

While information overload leads to an interruption overload, this in turn makes us multitask more, it put higher demands on our cognitive load and finally, it increases our stress levels. When the body is stressed, our sympathetic nervous system (SNS) sends signals to release cortisol and the changes in these signals have long been linked to the amount of cognitive load and stress a person experiences. When the load reaches a certain level of demand, stress appears. However, even though stress and cognitive load correlates, this study does not assume that these have a compulsory relationship between them or that they can't occur independently, as Niculescu et.al (2009) also states in their study. Stress can appear in many different forms, whether it's acute or chronic and it is usually divided into two different factors, internal or external. Internal stress are factors that comes from within, like your attitudes, expectations, and perceptions of things in life. External stressors are factors that a person can't control either its environmental, actions or caused directly by others. In comparison to our context, one assumption could be that the external stressors come from our modern technology, while the internal lies in the way we react and respond repeatedly to the external factor. Perception of stress is based on individual response, however, extensive research has discovered many commonly factors that causes stress reactions in our body, as explained by the effect of multitasking. The external factors which can also be described as an acute stress, means a condition where one becomes stressed due to a specific situation. How does this affect our interaction behaviour? And more importantly, how does this affects our cognitive load, when it is initially low? Previous research has investigated on how high values of cognitive load contributes to an increased stress level, yet there are still limitations on how stress affects our cognitive load when it is initially low. That is mainly what this project attempts to address.

To the best of our knowledge, studies in this field is also usually limited to a controlled laboratory environment and often investigated while a person interacts with a stationary device. This study tries to take it one step further by conducting an experiment in a real-life environment. Fuglerud et.al (2015) states that in order to get the best and most reliable results users needs to be studied in the actual field and in natural surroundings, and thereby this project proposes an experiment intended to determine when cognitive load and stress reaches a high/low level, while interacting on mobile interface in a real-life scenario. This thesis also aims to integrate some scientific evidence that has emerged regarding the association between cognitive load and stress, and also between smartphone and web users.

1.1 Structure of the Thesis

This paper is organized with an introduction along with problem statements, research questions and deeper explanations of the presented cognitive loads. This is followed by a literature review in ch.2 along with our prototype in ch.3 and methodology in ch.4. Furthermore, we present our analysis methods in ch.5 and ethical considerations in ch.6 before we continue to our results in ch.7, which is discussed in ch.8 and finally reaches a conclusion in ch.9. Following the conclusion is the acknowledgement, reference list and the appendices which consists of the programming code of the GSR equipment, interview guide, questionnaire, and consent form from the participants.

1.2 The overloads

Instead of seeking new information self initiated, we allow ourselves to be interrupted by new information, and we ease our way of multitasking. This circle is the key component towards cognitive overload. The frequency of multitasking is normally measured by the amount of switching tasks but the effect of it is usually found by looking at interruptions. Interruptions that both can be self initiated (e.g switching task voluntarily) or thrown to us by push notifications. While some interruptions can ease the burden of our working memory, like calendar reminder, other interruptions increases the load (e.g email notification). Mark, Gudith & Klocke (2008) claims that interruptions makes us work faster and that it does not affect the level of quality in the performed tasks. They continue with stating that interruptions even can be beneficial to us. In Bailey & Konstan (2006) study, the user performance was from 3-27% slower when not interrupted, which they claim was likely due to users needing more time to reorient and thereby induced a higher mental demands at the point of interruption. This was discovered by measuring the completion time and error rate of the performed task. Similar problems are addressed by Czerwinski, Horvitz & Wilhite (2004) who highlights the difficulty of switching back to our main task followed by an interruption.
Conaill & Frohlich (1995) states that it’s the complexity of the new task that makes an interruption disruptive. Which, of course, is logical since a higher complexity also requires a higher mental effort and probably a longer recovery time. Bailey et.al (2006) states that it’s rather due to the time at which the interruption is presented that determines whether it is disruptive or not. In the past, it was more likely common for receptionists or secretaries to decide when others should get interrupted and how. Today our secretary is our wearable devices (such as smartphone, tablet, smartwatches) and they usually deliver notifications at random times. According to Okoshi et.al (2015), this is one of the great issues that leads us to interruption overload and suggests that we need a detection system who rather decides when the user is “interruptible” and when to deliver notifications. They continue with stating that a deferring of the notifications will lower the impact of the interruption on users’ cognitive load. Through a real-time multi-device breakpoint detection, Okoshi et.al (2015) discovered that when notifications were delivered at the users “interruptible” time, the user’s workload

decreased with 71,8 %, compared to a random delivered notification. Which is a result that supports their statement on cognitive load reduction. While these presented studies delivers different results and claims, they all end up with the same conclusion, that interruption comes with a price. The overall cost of the disruptions was unanimously at the expense of the participants well-being and it led people to more stress and higher cognitive load. Which raises the question; Shouldn't the impact of interruptions then be more important than on our performance?

By the way we are living today we put to much pressure on our working memory and therefore we take advantages from notifications and reminders. In Czerwinski et.al. (2004) study, they discovered that 40% of the task switches were self initiated, meaning that the participants themselves chose to interrupt their primary task. This means that we do also have an individual responsibility to avoid an unnecessary overload. Another challenge we face is our "smartphone addiction" and the pursuit of constant gratification. When activities is not perceived as rewarding enough, the user seek towards other immediate gratifications and interruption overload quickly becomes a fact. This is a concern that Wilmer, Sherman & Chein (2017) summarizes as follows;

" The acute and short-term consequences of having one's attention distracted away from ongoing tasks is an obvious locus of concern in relation to smartphone habits, but there is also growing fear that the increasingly regular interactions we have with smartphones might also have a more lasting impact on the basic capacity for focused and sustained attention".

Some even claim that this constantly crave of immediate gratification and the "rewiring to our brain" that accompanies, is as much of a threat to society as the climate change (Greenfield, 2012). Wilmer et.al (2017) points out that there is limitations of empirical evidence that supports their concern, but one can still observe that these habits of ours; have a huge negative affect on our brain (Wilmer et.al, 2017).

1.3 Problem statement

As shown by the arguments made in the preceding sections, our way of living and interacting with modern technology increases the possibilities of cognitive overload and stress. Due to the fact that stress affects the amount of cognitive resources that a user has available, it means that a person might start with a limited threshold when interacting. If the user interface doesn't respect the limits of human cognitive processing, it will unnecessary force the users to an cognitive overload (Tracy & Albers, 2006). Oviatt (2006) states that by not knowing the user's natural behaviour, the interface might not match with the users capability. This means that even a regular website can contribute to unnecessary overload.

So, how can designer and developers accommodate this?

Blocking us from the outside world and not communicating digitally is not a viable option today, so designers and developers need to understand how users interact dynamically with technology and how it affects them cognitively. Kirsh (200) argues that the user interface can not be designed to increase multitasking and interruptions, but it has to be designed based on the assumption that the user performs things simultaneously or rapidly between one and another (Kirsh, 2000). Today, designers also needs to consider that users have different types of devices, are in different environments and has different preferences, knowledge and abilities.

Halarewich (2016) states that there are many variables in the user's external life and environment that could drain our working memory, but as a designer, this is beyond your control. However, what designers can control is the design variables in which causes cognitive overload (Halarewich, 2016). Another important aspect is the fact that while technology is constantly developing, our brain is still the same as 10,000 years ago (Nielsen, 2009) and haven't evolved to be able to process this amount of information (Levitin, 2015). Therefore in order to decrease cognitive load imposed by the user interface, designer and developers need to understand how the human brain are processing information (attention, perception, short-term memory).

Cognitive psychologists compare our information processing with the computer architecture, saying that in the same way a computer works with input, processing and output, so does the human brain. However, since the capacity of each computer varies,

this also applies to the human brain. How each and one of us perceive and store information depends very much on our individual differences and thereby we can't universally claim when cognitive load is exceeded. What we can do according to Halarewich (2016), is to "isolate the design mistakes that mostly commonly cause it " (Halarewich, 2016).
Yet, how will designers know when cognitive load is high and what the causality is? According to Kramer (1991, as cited in Chen et.al, 2013), the physiological approach for cognitive load measurement is based on the assumption that any changes in human cognitive functioning are reflected in human physiology. The same principle applies to stress measurement and stress response. This assumption is consistent with many of the recent results in the field, where scientists found a strong correlations between the variation of the signals from our autonomic nervous system and the variations in the cognitive load/- and stress measurement. The autonomic nervous system (ANS) is a network of nerves which controls the actions of internal organs and it carries out information to our central nervous system (CNS) whether our body experiences stress or not. The autonomic nervous system (ANS) comprises two sets of nerves, the sympathetic and parasympathetic nervous system (PNS). As mentioned in previous section, our sympathetic nervous system (SNS) is activated when our body experience stress, either by external or internal factors. Our parasympathetic nervous system serves to regulate restoration, i.e decrease our stress state. The changes in our SNS have long been measured by scientists through physiological measurements but even though it has been proven many reliable results on those changes, these measurements has proven to be sensitive both to cognitive load and stress (Setz et.al, 2010), (Shi et.al, 2007) & (Wang & Guo, 2019). This is one of the major challenges with these assessments and also why more research is needed in order to find a reliable method for distinguishing them. Which, to the best of my knowledge, is still unresolved. The purpose of this study is thereby to further contribute to this research area.
Lyu et.al (2015) states that in order to improve sensitivity and reliability of these measures, it requires multiple measures. Gjoreski (2015) claims that stress is not only detected physiologically but also emotionally and behaviorally, in which according to Chen et.al (2013) also applies to cognitive load detection. Thereby, this project proposes a method by using multiple objective measurements along with subjective

measurements. By using this kind of approach it can hopefully help us to better explain the causality and thereby be able to distinguish cognitive load and stress. It has been established that cognitive overload contributes to an increased mental stress and this relationship is often what researchers focus on. Since this study doesn't assume that they have a compulsory relationship, we believe that this is a limitation and thereby the aim of this project is to investigate on how acute stress affects cognitive load, when it's initially low. Another problem is that many similar studies uses stress assessment based on general responses, even though stress differs enormously between people. This either cause unreliable results or failure of stress manipulation (Conway et.al, 2013), & Niculescu et.al (2009)). To address this problem, this project attempt is to induce stress by using a subjective approach. This means that we manipulate stress by allowing each of the participant's to describe a typical stress scenario and thereby creating a person-specific stress template.

1.4 Research questions & Hypothesis

Based on the previous sections, two of the main research questions is as follows:

- In what degree does acute stress affects our cognitive load (CL) while interacting on mobile web, even when CL is initially low?

- How can designers or developers work to reduce cognitive load and stress level among mobile web users?

Sub-questions that this project will attempt to answer is:

- If the performed task has a low load, and the subject is exposed to acute stress, will the cognitive load be increased?

- Will there be any differences when solving the task?.

Based on these questions, the hypothesis in this study is:

A: A stressed situation increases the cognitive load even if the task load is low

B: The users task performance decreases when stress level increases (e.g. the task takes longer to complete, more errors are made).

1.5 Cognitive load

In 1956, Miller developed the “Information processing theory”, which describes how information is received, processed and stored in the human brain (Miller, 1956). This process is usually explained by the relationship between our working memory and long-term memory. As far as we know, long term memory is unlimited and will never be overloaded but our working memory however, has a limited storage capacity of approximately seven elements, also called ”chunks” (Miller, 1956). According to McGrew (2005), our working memory reflects the

“ability to temporarily store and perform a set of cognitive operations on information that requires divided attention and the management of the limited capacity resources of short-term memory” McGrew (2005).

It means that our working memory allow us to retain information while we perform or pay attention to something else and it allows us to remember and respond directly to information, such as during a conversation. In short terms, it is our working memory that enables us to multitask. When new information reaches our working memory, it is processed and stored temporarily before it’s either transmitted to our long-term memory, or decayed and replaced by new information. Cognitive load is described as the amount of mental effort being used in the working memory. If too much information is kept in the working memory at the same time, it causes an cognitive overload (Malamed, 2010). Morrison et.al (2014) imposes three types of cognitive load in the working memory: intrinsic load, extraneous load, and germane load (Morrison et.al, 2014). The intrinsic load is the load imposed by a task, while the load to understand and process content is referred to as the germane load, and the extraneous load is the way the information is presented (Zagermann, Pfeil, Reiterer, 2018).
Paas et.al (2003) states that all these loads are additive and works in tandem, which means that not only are they working together, they are also influenced by one and another. The way the information is presented to the user (extraneous load) and the tasks required for the users to achieve their goals (intrinsic load) are both factors that are relevant to germane cognitive load (Paas et.al, 2003). They continue by stating that if intrinsic load is low, then reducing the extraneous load will be less effective on the

germane load. Gwizdka (2010), on the other hand, concludes that both extraneous load and intrinsic load can be measured separately in order to assess the amount of cognitive load. In his study he compared the presentation of search results in two different interfaces, in which one showed a simple list and the other a list along with categories. Gwizdka (2010) discovered that the mental effort among the users, and the reaction time, highly increased when the presentation of the search result (extraneous load) became more complex and not due to the task itself (Gwizdka, 2010). This is also backed up by Whitenton (2013) who states that in order to minimize the load, designers should strive to eliminate, or at least minimize the extraneous load.

While the main focus for this project is to investigate the effect of acute stress on our cognitive load while interacting, load imposed by the user interface will also be discussed in our final evaluation of the study.

2 Literature review

There are many proposed solutions towards reducing cognitive load but Elting et.al (2002) claims that using the concept of Cognitive Load Theory (CLT) has shown better results both in performance and learning effect among users. The principle of CLT is in short when the amount of information exceeds the limited capacity of our working memory, then our way of learning is impaired. Why CLT has shown to reduce cognitive load is because it’s based around the fact that we process information through multiple memory channels, such as sight, hearing, touch, taste and smell. In Elting et.al (2002) study they used a tourist information system as a test environment, and their aim was to investigate the effectiveness regarding recall-performance among users. By doing so they used modality-combinations both in terms of devices and interaction condition. Their result showed that the most effective modality combination over all device-groups was found when speech was presented with pictures only, which supports the concept behind CLT. However, In terms of user experience and usability, participants found the combination of text, image and speech as most appealing and would rather choose this as a presentation mode, which due to its reundance actually contradicts the concept of CLT. Yet according to Elting et.al (2002), this combination was also rated quite low in load, meaning that if the content (the intrinsic load) is not too complex to begin with, then redundant extraneous load will not contribute to an overload.
Whitenton (2013) however, claims that designers rather need to be focusing on minimizing the extraneous load (the way information is presented) in order to decrease cognitive load. This is an assumption that Gwizdka (2010) also discovered in his study when reviewing and comparing the presentation of search results in two different interfaces. While CLT originally was intended for the field of education and learning purposes, there are many similarities between CLT and the principles of web usability. The idea behind CLT is based the on the Information Processing Theory (IPT) and Miller's concept of "chunking". This means that the capacity of our working memory has a limit of holding approximately seven chunks at a time, plus or minus two depending on the complexity of the information. By “chunks” this could mean words, digits, images e.g. Halarewich (2016) states that both Information Processing Theory (IPT) and CLT can

equally be applied to general user experience design and that many designers has prompted Millers concept of “chunking” both in designing menus (with limited items of seven or nine), sectioning of text (in form of headings, paragraphs), and grouping and organizing web content, all of which have the purpose of reducing cognitive load (Halarewich, 2016). Many uses websites in order to search for information, and even though web browsing is more of a casual activity then formal education, browsing the web in general can imply a high cognitive load (Caldiroli et.al, 2017).

The user must absorb new information, learn how to use a site’s navigation, layout and transactional forms, and even if the site is fairly familiar, the users still has to carry around the information that is relevant to their goal (Whitenton, 2013).

Caldiroli et.al (2017) claims that by using CLT you'll not only minimize cognitive effort, it will also increase usability and accessibility of a website. While this statement might be true in some cases, Elting's (2002) study shows that this could also exclude some of the users from accessing information equally. In their result the most effective modality combination, and the one consistent with CLT, was found when speech was presented with pictures only. If designers had blindly adopted this into the user interface design, this would have excluded people who are deaf in getting the same information, or generally people who are in a loud environment. While the concept of Miller's theory has inspired some of the best-known usability principles (Whitenton, 2013), it would never be able to determine when a website is fully usable or accessible.

2.1 Universal Design

In terms of Information and Communication Technology (ICT), “accessibility” means that everyone can access information and utilize the content equally. One of the issues that Shneiderman & Plaisant (2015) address is that many designers assume that they know and understand the user, and this is not an approach to a successful design (Shneiderman & Plaisant, 2015). A successful design is as Schneiderman et.al (2015) points out, by working towards universal usability, including all the individual differences among the users. This means that you need to design for all regardless of age, gender, cultural or ethnic background or physical and cognitive abilities.

The principle of “design for all” began with the seven principles of Universal Design (UD), which is a concept based around the approach of designing products and environments so that any person, regardless of age or ability can use it. The seven principles are explained as, equitable use, the flexibility of use, simple and intuitive use, perceptible information, tolerance for error, low physical effort, and size and space for appropriate use (Molly Follette Story, 1998). Universal Design was first introduced by the architect Ron Mace and was mainly developed to remove potential barriers for people with disabilities or impairments in built environments. Since then it has extended beyond architecture and into the digital world of both learning and general web usage. Web accessibility can be achieved through several techniques or guidelines, but the most widely known guidelines are the Web Accessibility Initiative - Accessible Rich Internet Applications Suite (WAI-ARIA), which are technical specifications that defines the functionality of Assistive Technology (AT) and Web Content Accessibility Guidelines (WCAG), which are based around design principles towards making a website both accessible and usable for people with or without the usage of AT.

The concept of Universal Design and its principles does not only apply for web usage, it is also used in learning purposes. A person's capacity to learn is directly related to how much cognitive load is used to comprehend the information, and this is also very much individual. This is what Universal Design for Learning (UDL) addresses, to use learning methods that is adjusted for every student’s individual abilities and needs. The goal of UDL is that by providing flexible learning methods, it removes possible barriers to learning and “give all students equal opportunities to succeed” (Morin, n. d). Which is generally what universal design stands for. Guidelines are usually structured to be general enough in order to be applicable to a wide range of products and technologies and it is according to Schneiderman et.al (2015) difficult to see the implication of these design principles since they are more used as guidelines. This is why universal design becomes more of a concept instead of a practice (Schneiderman et.al, 2015) . The reason, as Halbach (2010) claims, is that many guidelines is not established based on concrete examples from the user's real world experiences and therefore it's more difficult to interpret and to further implement. Halbach (2010) states that when not having a reflecting knowledge of the diverse disabilities and the technological implementation, it causes great limitations of evaluating different obstacles and needs of these people, which is why more empirical evidence is needed. This is also supported by Rømen and Svanaes (2010) who discovered that using WCAG alone only identified 38% of the accessibility problems in their heuristic evaluation of selected websites (Rømen and Svanaes, 2010). Based on the websites they reviewed, they claims that this could have been avoided if the guidelines were based on empirical data and validated empirically (Rømen & Svanaes, 2010). Although this project will not add more empirical evidence to this particularly topic, cognitive disabilities and cognitive overload shows a clear resemblance. As mentioned in previous section, Cognitive Load Theory (CLT) is based on the fact that learning is impaired when the amount of information exceeds the limited capacity of our working memory, which is referred to as cognitive load. When information exceeds our limited capacity, it causes cognitive overload and it affects both our memory and learning processes. Long-term stress (chronic) and cognitive overload is what causes these processes to deteriorate and it is also these processes that people with cognitive disabilities struggles with, especially within conditions like ADD, dyslexia and other learning disabilities (WebAim, 2018). It has also been discovered that chronic stress damage our memory functions to such an extent that people risk developing a pre-dementia condition (Dockrill, 2015). With this considered, along with our aging population, and with our digital modernity of draining our cognitive resources, people with cognitive disabilities will no longer be a minority. Although this statement in itself is a good reason for why everything should be universally designed, more and more studies also provide evidence on how this design is beneficial even for people without disabilities. Sandnes and Lundh (2015) compared the effectiveness of two popular calendar types; table view and event list view, with the hypothesis that a table view list would be more cognitively demanding for user's with disabilities, than with a simple event list. The experiment was conducted on 20 participants divided into two groups, one with participants diagnosed with ADHD/ADD and the other group without any conditions. While the amount of cognitive load wasn't established, their results showed that the list view were performed more efficiently and with significantly fewer errors. Although there were no significant differences in the group of people without disabilities, their results still revealed the same and the conclusion leaned toward the same benefits in both of the groups. But given that they had no assessment of the amount of cognitive effort, it becomes difficult to determine how big the difference is between the participants. In Rello and Baeza-Yates (2016) study they investigated the effect of typeface (font) on screen reading performance for people with and without dyslexia. By using fonts specifically designed for dyslexia and other commonly known fonts such as Arial, Helvetica, Verdana, Times New Roman, they analyzed the participants readability and comprehensibility through the participants personal preferences and by eye-tracking technology. What they first discovered was that fonts particularly designed for dyslexia was not giving any better readability. Instead, all the participants reading performance were significantly improved by the same font; Verdana. Their second discovery was that the readability of people with and without dyslexia was not necessarily what they preferred, but rather due to what's socially imprinted than what's actually best in reality. This was discovered by using eye-tracking technology where they measured reading time, fixation duration and the number of fixations. By these variables they were able to analyze the time taken for a participant to read a text, how long the eye rests on a single spot of the text and how often it appears. All of these combined then revealed the participant's readability and the actual fonts to be used for a better usability, which could have been misled if only subjective measures were used in the study (Rello and Baeza-Yates. 2016). Similar results were given in Hill et.al (2011) study where they investigated a library website in order to identify how efficiently users could access to information or whether it needed any improvement in the information searching process. In their survey the usability was evaluated by quantitative rating scores and objectively through eye movements and while the quantitative results revealed usability problem, the eye tracking data gave direct access to the source of the usability flaws. It also provide the entire interaction process from start to finish; not just the end result, which often is summarized by the user's own preferences or experiences. (Hill et.al, 2011). For this project purpose, this is a solution we believe could be a suitable approach in reaching the intended goal.

2.2 Cognitive load & stress in web usability

The user interface provides our way of interacting with technology and it mediates the relation between the user and the technology. There are two terms in user interface design that plays an important role, usability and user experience. Usability refers to the degree to which interactive products are effective, easy to use and easy to learn.
The result of the interaction and the overall satisfaction is known as our user experience (UX). The differences between them can be perceived as confusing and thus treated equally, but the fact is that they have a compulsory relationship where one is always affected by the other. In Ritthiron & Jiamsanguanwong (2017) study they found strong correlations between the tasks with higher percentage of failure and satisfaction scores. The higher percentage of failure, the lower satisfaction scores. Which means that if a website struggles with a usability problem, than it probably affects the entire user experience. However, Nielsen and Norman (n. d.) states that even though the usability of a website is flawless, the site could be missing some features that are important to the user and thereby it affects the whole experience negatively (Nielsen & Norman, n. d.). This means that we can establish that they cooperate but also works individually, which is what this project hope to prove regarding cognitive load and stress. Every aspect in a user interface that impair the user of reaching to their goal, either draws from our cognitive resources or it increases our stress level.
According to Albert and Tullis (2013), stress within ICT typically occurs, when a person perceives that they cannot cope with the current tasks, when they get overwhelmed with too much information or they get the feeling of losing control. They continue by stating that all of these factors can be measured through "user experience metrics".
It can be by measuring the time taken to complete a task, which could be connected to a person's perception of not coping with the current task. It could be by measuring success rate or number of failed mouse clicks, which may be perceived as a way of having or not having control. Or it could be a constant flickering pupil movement, which could be connected to overwhelmed information (Albert & Tullis, 2013).
While the above-mentioned metrics can be adaptable in determine overall user experience and perhaps the amount of stress level, Hill et.al (2011) also discovered usability problems by using these. According to Tracy and Albers (2006), it is also

possible to connect usability problems to the amount of cognitive load that a person experiences (Tracy & Albers, 2006). In view of this, it is difficult to determine whether a user is actually affected by stress or by a high cognitive load or by both simultaneously. That is why this study aims to find a method that distinguishes them.

2.3 Cognitive load and stress assessment

There are many proposed solutions in measuring cognitive load and stress among users. Chen et.al (2013) states that there are four main methods: physiological measures (biosensors), subjective (self-report) measures (where users rank their experienced level of load), performance measures (dual-task method) and behavioural measures (observing patterns of interactive behaviour (mouse-click events etc.).
Chen et.al (2013) claims that the most consisting results has been achieved through subjective measures and according to Lyu et.al (2015) it can also be used in order to distinguish cognitive load and stress. However, Chen et.al (2013) argues that this method is quite impractical in real-time. In subjective measure the user is required to perform a self-assessment of his or her mental demand by answering a set of assessment questions immediately after the task, and not only does this interrupt the task flow, it also adds more load to the potentially overloaded user (Chen et.al, 2013). Physiological measures has by its high accuracy shown to be ideal in providing real-time control indications of experienced load. Many studies have long linked biosensors such as eye-tracking, Electroencephalography (EEG), Blood Volume Pulse (BVP), Heart Rate Variability (HRV) or Galvanic Skin Response (GSR) as methods of assessing stress and cognitive activities. In the Shovon et.al (2015) study, EEG signals was recorded to induce cognitive load while the user engaged in multiple search data. They discovered that the EEG signals varied and different brain areas became more active depending on each performed task. This dynamic changes was also revealed in the cognitive load. Since EEG signals provide a direct measurement of cognitive activity, they were able to find a correlation between the cognitive activities and the immediate interaction (Shovon et.al, 2015). In the traditional evaluation of the site's usability, eye tracking is considered a valid tool because it provides accurate eye movement from the user's point of view. It reveals what the users see and what they ignore when navigating through a user interface. However, our eyes also reveals our cognitive activities. According to Zagermann et.al (2018) & Kruger et.al (2013) the percentage change in pupil diameter, pupil dilation, blink rate, and the number of fixations and saccades per second, all reacts to changes in our cognitive activity. In Zagermann et.al (2018) study they analyzed four eye movements (pupil dilation, fixations, saccades, and blink rate) to identify changes in

cognitive activity related to an increased level of cognitive load. However, individually all the participants responded differently which made it difficult to determine whether the eye movements depended on the level of difficulty or environmental conditions.

The disadvantage as Zagermann et.al (2018) states is that the pupil is more sensitive to light than cognitive demands and the size of the pupil can be affected by numerous external factors. While it's possible to avoid confounds on pupil size by controlling light changes, it is still limited to both indoor environments and certain devices. The same weakness applies for EEG equipment. As the electrical signals is produced in our brain, so are the signals in that can be measured through our skin, also called Galvanic Skin Response. Ayata et.al (2017) found correlations between GSR and arousal with an 85.07% accuracy. Nakasone et al (2005) have successfully used skin conductance for emotion detection. In another study, GSR was used in order to detect stress (Setz et.al, 2010) and in Shi et.al (2007) study they found strong correlations between GSR and cognitive load.

Galvanic skin response has in many studies provided reliable results on our changes in the autonomic nervous system. But with its large area of use, it’s also sensitive to various reactions, especially between cognitive load and stress (Conway, Dick, Li, Wang & Chen, 2013). In Conway et.al (2013) experiment they used three levels of difficulties in order to establish cognitive load and they induced stress by using MIST, a template which consisted of three conditions: rest, control and experimental. In Setz et.al (2010) study, their results showed an accuracy of 82.8% using MIST as a stress induction.

In Conway et.al (2013) results they found out that while extracting and comparing mean GSR against stress, no correlations or common patterns was found between the GSR values and the different stress states. However, when extracting the frequency of high/low peaks a stronger correlation were found between the task difficulties and stress condition. The frequency appeared to be strongest during level one, which had the lowest difficulty and then it gradually went down in level two and three. One of the reason might be due to a “first impression effect”, where the participant gets more and more confident during each session (Niculescu et.al, 2009), or it could mean that a stressed situation affects our cognitive load, even when it’s initially low. Which is what this project will address further. While Conway et.al (2013) managed to dissociate cognitive load from stress in their experiment, the credibility may still be questioned with the use of stress measurement. Even though our body respond biologically the same towards stress, it is a variety of how each person reacts and responds to stress. This response can be different due to past experience or as Murgia mentions, it can come genetically from your ancestors and how they handled stress (Murgia, 2015). Conway et.al (2013) do point out the fact that stress varies and through the results from previous research they used the most commonly paradigm for stress induction, with conditions such as: failure at a task, with feelings of lack of control, and the feeling of being evaluated by others. To determine stress among the participants they investigated the responses through behavioural and performance measure. While these do manage to reflect our stress symptoms, using a physiological measurement might serve a higher credibility, as it did with the assessment of cognitive load.

Lyu et.al (2015) states that researchers often need to combine multiple measures in order to improve the sensitivity and reliability in the measurement. With a physical measurement as an index, linked to the behaviour and performance measurement, then maybe stress condition could be established with an even higher certainty.

Gjoreski (2015) claims that in order to measure stress you need to develop a stress detection system which monitor three stress responses; emotional, behavioral and physiological (Gjoreski, 2015). This kind of approach was adapted in Niculescu et.al (2009) study where both physiological, behavioural/performance and subjective measures was deployed in order to validate the induced stress and cognitive load levels. To induce stress among the participants they used six different parameters (e.g background noises etc.) and two parameters for the cognitive load manipulation (task complexity and presentation format). The attempt of distinguishing cognitive load and stress was by alternating both high/low levels in four different trials. Their results showed a successful manipulation of cognitive load in all of the trials while stress manipulation only succeeded in one of the them. In the result they also claim that the levels of both factors were better indicated by the behavioural/performance and subjective measurements than by the physiological measurements. Niculescu et.al (2009) states that it could be due to their own lack of experience and interpretation ability of the physiological measurement. One can argue for the fact that their physiological measurements also were performed in parallel and by the lack of an individual index, the causality between the values and the factors becomes more difficult to comprehend. In their experiment Niculescu et.al (2009) used Galvanic Skin Response (GSR) and Heart Rate Variability (HRV) as for the physiological measurement and both of these sensors has shown promising results in the assessment of cognitive load and stress. In Shi et.al (2007) study they conducted a user experiment in order to assess potential cognitive load indicators in a multi-modal interface, this by using GSR as a control measure. Their results showed that the GSR levels increased significantly when task cognitive load level increased. Mark et.al (2014) discovered Heart Rate Variability (HRV) as a valid indicator of stress while investigating multitasking and computer usage. During a 7-days study HRV was monitored 24/7 and by logging the computer data they were able to show direct relationship between computer duration and stress. The result revealed that as time spent on the computer increased, stress increased, and with more window switches, the higher the stress (Mark et.al, 2014). While many of these studies show positive results in stress assessment, none of them takes a subjective approach to induce stress. Gjoreski (2015) states that in order to determine when the participant's experience stress, you need to construct a person-specific stress template, use this to induce a stress scenario and further investigate the responses in a real-life scenario. By investigating in the actual field and in the users natural surroundings, the more reliable results in finding real-world problems, as highlighted by Fuglerud et.al (2015). This project has therefore chosen to step away from an controlled laboratory environment and into realistic environments and as a part of the experiment we induced stress by developing an individual user test based on the outcome from a questionnaire. In our project, Heart Rate Variability acts as an index of stress and Galvanic Skin Response as an index of cognitive load. With the advancement of mobile communication technologies HRV monitoring is almost available in every modern wearable devices. Thereby the HRV measurement doesn't require any heavy equipment and it is not limited to certain environments such as EEG or Eye Tracking. This also applies for the GSR sensors. Unfortunately, GSR sensors are not as available on the market as HRV and among those who distribute GSR, it's either a long delivery time to the Nordic countries or too expensive, at least for this project budget. Therefore, this study had to develop its own version of a GSR device.

3 Prototype

Galvanic Skin Response is one of the physiological measurements whose task is to measure the electrical conductance of the skin. These values then varies depending on how much sweat our body produces. In our experiment a GSR device was developed using a Grove Sensor which was controlled by an Arduino UNO and further coordinated by an Arduino IDE software on a laptop. While most common GSR devices is used to measure the conductance of our skin, our device was unfortunately not able to perform this kind of measurement. What our device measure is the electrical resistance through our body and detects how much resistance our skin emits. While this is a limitation to our device, Lunn & Harper (2010) demonstrates that an increased sweating (increased conductivity), also increases resistance, which means that they act as an inverse to each other. Because of this we could then assume that when the resistance was high, the subject should produced more sweat. Our GSR device has two leads attached to two fingers, index and middle finger, where one of the leads sends the current reading and the other measure the difference. The Grove sensor has an operating voltage of 3.3 / 5 voltage, with an input signal of resistance and an output signal of analog readings. As illustrated in figure 1.1, the Arduino is powered by an USB cable (which in turn is connected to a laptop) and the board has several different kinds of pins, each of which is labeled and used for different functions.

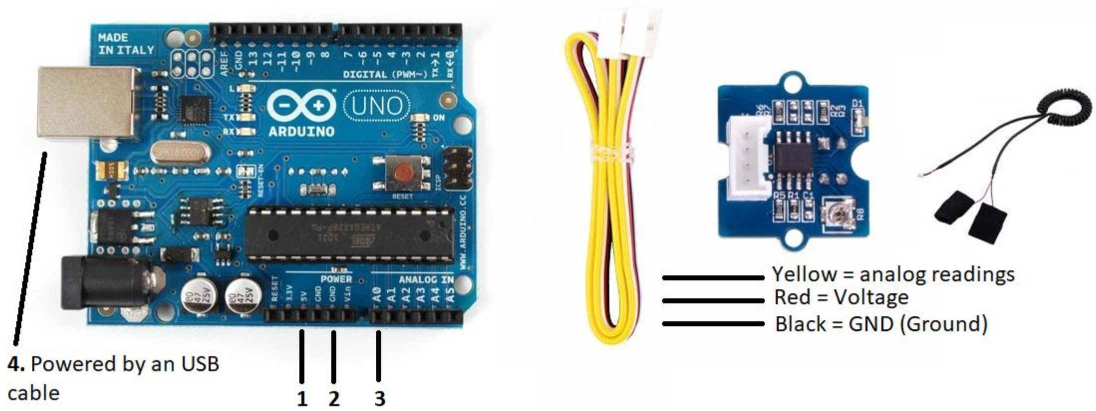

Figure 3.1. Illustration of the connection between Arduino Uno and Grove Sensor.

The Grove Sensor is connected to Arduino Uno by three colored wires. The yellow wire represents the analog signals from the sensor and it's connected to Arduino's analog inputs. Essentially there is no functionality unless there is a negative and a positive electrical pole and for the Grove Sensor, the red cable is connected to the Arduino 5 voltage pin and the black cable is connected to the Arduino ground pin (see figure 3.1). The unit of electrical resistance is measured in ohm (represented by Ω) and in order to get the human resistance of our GSR into values in ohm, the sensor was calibrated using a formula compatible with the Arduino board and then converted into digital readings in the Arduino code. The potentiometer was used on the Grove Sensor board to adjust the resistor value (See figure 3.2).

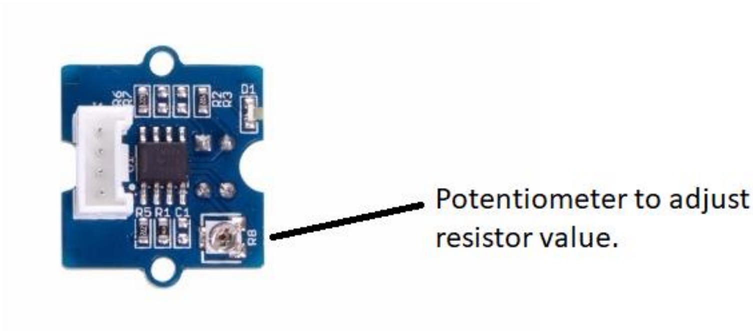

Figure 3.2. Grove Sensor board illustrating potentiometer.

All the data from the sensor was stored as a txt file on a memory card by using a memory card reader implemented on the Arduino board.

3.1 Programming behind the GSR device

To control our sensor through the board we've used Arduino's own software, IDE, whose task is to send instructions to the microcontroller on the board. These instructions basically works as reading inputs (from e. g a sensor) and turns it into an output (printing out measuring values). The programming language is based on something called processing, which is very similar to C/C++ coding language. The structure of the Arduino program runs basically by two parts, preparation and execution. This consist usually by two blocks, or functions; void setup () and void loop (). Both of these functions are required for the program to work. When a function is called, we initialize statements that will be executed. The setup () function starts the communication with the computer and here we decides the date rate in bits per second for the serial data transmission. One common rate is 9600 and it’s declared by the Serial.begin () statement (Figur 3.3).

```
void setup() {
    // Initialize the pushbutton pin as an input:
    pinMode(buttonPin, INPUT);


    // Setting baud rate
    Serial.begin(9600);
    while (!Serial); // Wait for serial port to connect. Needed for native USB port only

    Serial.print(F("Initializing SD card..."));

    if (!SD.begin(4)) {
        Serial.println(F("Initialization failed!"));
        while (1);
    }
    Serial.println(F("Initialization done."));

}
```

Figure 3.3. Illustration of the setup function in the Arduino code

It is also where we initializes the SD card and set one of the pins of the Arduino board as an input. In the loop we actively control the Arduino boards and it's where we set the initial values and initializes the reading inputs and triggering outputs etc. Our program starts with menu options which calls upon other functions to further be executed (See figure 3.4).

```
void loop() {
    GetMainMenuInput();
}
```

Figure 3.4. Illustration of the loop function in the Arduino code.

In the menu function there are four different cases; AutoSetThreshold, StartRecordingValues, PrintRecordedValues and EraseRecordedValues. All of these options are again different blocks or functions that are called upon while choosing one of these from the menu (See Appendix A). The one most frequently used and the most essential for our measurement is the function of StartRecordingValues (). This function writes the results on an SD-card and sets the time interval on the measurements that are retrieved from another function called RecordValue (). It is in the RecordValue () function all the reading signals from the sensor are retrieved and interpreted. Since the Grove Sensor isn't designed to deliver µS (microSiemens) values, which is the unit of conductance, the ohm values was converted into microSiemens (See figure 3.5).

```
float resistance = ((1024 + 2 * adcValue) * 10000.0) / (512 - adcValue); // In ohms
float conductance = 1 / resistance;
float conductanceMS = conductance * 1000000; //conductance in micro-Siemens
```

Figure 3.5. Illustration of the conversion from ohm to microSiemens in Arduino code.

The human resistance was set by using the recommended formula of this sensor (See figure 3.6).

```
According to http://wiki.seeedstudio.com/Grove-GSR_Sensor/:

Human Resistance = ((1024+2*Serial_Port_Reading)*10000)/(512-Serial_Port_Reading),
unit is ohm, Serial_Port_Reading is the value display on Serial Port(between 0~1023)
```

Figure 3.6. Illustration of the recommended formula of the sensor calibration.

The output data from the sensor displays time, adc value, resistance and conductance in the serial monitor and the same is written to the SD-card (See figure 3.7).

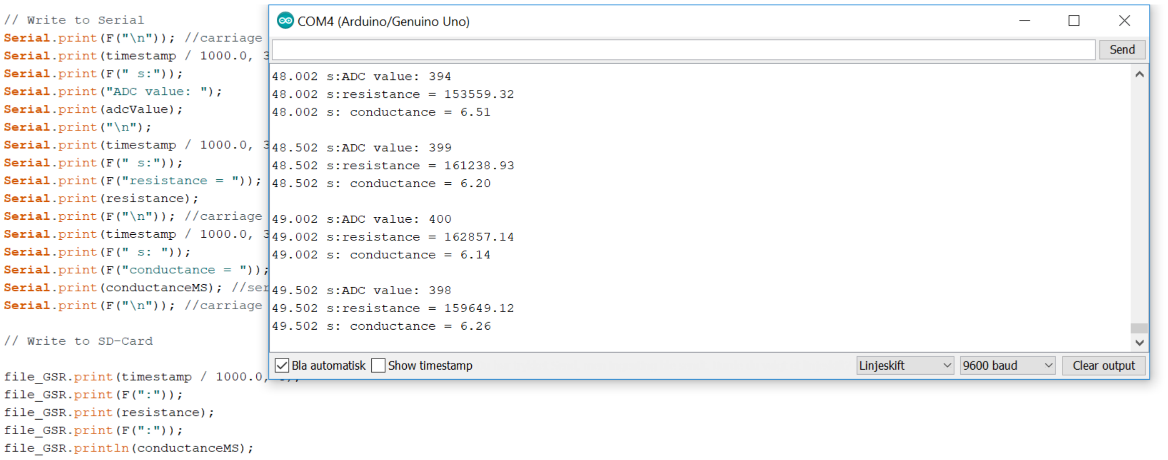

Figure 3.7. Illustration of the output data in the serial monitor, in real-time.

Even though resistance and conductivity is corresponding to each other, conductance gave us an easier overview and therefore these values were chosen to be the only one highlighted in our analysis.

3.2 Stress measurement

As for the assessment of stress levels, this project have used Heart Rate Variability (HRV) as a physiological measurement. The device that has been used is a Polar H10 Heart Sensor chest strap along with the compatible software application Elite HRV (See figure 3.2.1).

The advantage of using Elite HRV application is that they provide the calculations needed for HRV measurement. When people are reacting to stress, our sympathetic nervous system responds by e. g increasing our heart rate (HR). Although this might be true in many cases, multiple researchers have discovered that it is rather due to the variation between each heartbeat that indicates when a person is stressed, and not initially an high HR. Heart Rate (HR) is measured in beats-per-minute and the time between the each beat is called R-R Intervals, which is measured in milliseconds (ms). The variation in the R-R intervals and these small fluctuations of the heart is what is known as the Heart Rate Variability. HRV has proven to be a powerful observator of the balance between our sympathetic and parasympathetic nervous system and in several studies they've demonstrated many positive results by using this type of measurement in revealing stress reactions (Yu, Funk, Feijs, 2017 & Hjortskov et.al, 2004).

However, in order to increase accuracy and reliability of the assessment, we’ve also used behavioural measures and some of the subjective measures to determine the response of stress.

4 Methods

Within scientific research there are many methods of conducting and gathering data to your study. For similar studies, researchers use either quantitative, qualitative or a mixed method approach, where the latter is often seen as the most beneficial. Penta & Tamburri (2017) points out that it is because it's difficult to find causality when using them individually, and thereby a mixed method is usually preferred. They claim that quantitative insights usually offers statistical significance about an observed phenomena, but these alone doesn't explain why and what has caused this phenomena. From this perspective, these two approaches are rather more complementary than competitive, which also has been proven in many of the mentioned studies in this report, when combining subjective and objective methods (Shi et.al, 2007 & Nourbakhsh, Wang & Calvo, 2017 & Zagermann, Pfeil & Reiterer, 2018). While a qualitative approach in this study might answer; Does acute stress affects our cognitive load while interacting on mobile web? A quantitative approach could answer; In what degree does acute stress affects our cognitive load? While the objective methods gives us the information of when the cognitive load and stress level is low/high and how it variates, the subjective methods explains the causality. Similar studies often conduct their experiment in a controlled laboratory environment and this can be very limited when finding people's real-life challenges in relation to human-computer-interaction. To accommodate this, one uses empirical methods, meaning that you derive knowledge from actual experience and real-life scenarios rather than from theory. The theoretical background may be formed throughout the literature review, but the primary data is collected through empirical evidence and the conclusions of the study is strictly drawn from concrete verifiable evidence. Due to the nature of this project being based around investigating how acute stress affects our cognitive load while interacting on mobile web in a real-life scenario, it was first decided that an empirical method would be most optimal. However, since the purpose of this study is to investigate individual stress assessment, in addition to a limited time frame, we had to simulate a realistic stress scenario in order to make the experiment feasible. Thereby our methodology can not be claimed as fully empirical but rather seen as a research based on empirical methods. An empirical research can be conducted and analyzed using qualitative or quantitative methods and for the above reasons, a mixed approach is used in this study.

One of the quantitative methods in empirical research are experimental research, which means that the data is collected through an experiment and then compared against a hypothesis. In an experiment a situation is normally created where variables are manipulated and tested against each other. These are divided into independent variable (IV) and dependent variable (DV) and the query is usually; what happens to a variable (IV) when another variable (DV) is removed or altered. There are different types of experimental design but one most commonly linked to empirical studies is “natural experiment”, which means that an experiment is conducted in a real-life scenario but without the researchers intervention and without any controlled or manipulated variables. Here, the participant is observed in a real-life situation but without the knowledge of being observed and a single variable is influenced by nature or factors outside the researcher’s control. Another known experimental approach is “field experiment”, which in contrary to natural experiment is conducted in a real-life situation but with the researcher being actively involved and with controlled and manipulated variables. Since this study controls and manipulates a real scenario, it was also logical that the variables are treated equally and thereby we can argue that our experiment is more closely to a “field experiment” than a "natural experiment".
In our experiment the dependent variable (DV) is defined as acute stress and the independent variable (IV) as cognitive load. The cause and the effect of these variables can be described by our hypotheses; “A stressed situation (cause) increases the cognitive load (effect) even if the task load is low” and “The user’s task performance decreases (effect) when stress level increases (cause).

One of the qualitative methods in empirical research is observational method, which means that a researcher observes the participants ongoing behavior. It reveals how people react or behave when they are put in a real-life situation or in natural surroundings and for this project, such a method seemed imperative in order to link the cause/effect of the variables with the user's interaction behavior. To achieve the highest accuracy of stress and cognitive load detection, Gjoreski (2015) refers to the analysis of three components; physiological, emotional and behavioral. In our observational method we therefore used behavioural and performance measures which was observed using a mobile screen recorder. Through the recording we are then able to extract parameters such as response competition time and number of errors related to the given task, and behavior patterns such as verbal feedback, breaks or hesitations. To determine the participant's own views and opinions it was also decided to employ a subjective method with interviews and questionnaires, mainly because the subjective responses could be used against the results of the physiological and performance and behavioral measures but it also gave us some directions in how to induce a stressed situation for the experiment.

In summary, since we do not completely use one type of methodology in this study, our method can be considered as a hybrid approach. We have conducted a research using "field experiment" and based on empirical methods we have collected and analyzed data both quantitatively and qualitatively (e.g mixed method).

4.1 Procedure

Like Niculescu et al (2009), our experiment was conducted on a focus group with limited number of participants. This decision was mainly due to the in-depth data gathering that was required and due to the time constraints of this project. It was also based on making the project realistic to complete within the outlined timeframe. For this study there were 9 participants, all healthy females in the ages between 26-34. One full-time student, one part-time student along with a part-time job, two with a part-time job and the latter five had a full-time job. All of the participants signed a declaration of consent which was retrieved from the Norwegian Center of Research Data (See Appendix D). Two of the pariticpants were excluded from the analysis due to technical failure of the equipment and inconclusive results from the GSR measurement.

The amount of stressors in a person's life often correlates with the amount of stress a person experiences and therefore, all participants had to fill out a stress detection questionnaire (See Appendix C). The participants was also required to state certain diseases, mental conditions and label any other events that possibly could affect the physiological measurements (e.g medicine, caffeine, alcohol etc.) Even though our body respond biologically the same towards stress, it is a variety of how each person reacts and responds to stress. Sierra et al. (2011) proposes a solution of creating a stress template in where they model the behaviour of the participants both during a stressful and non-stressful situation. This was done by inducing tests that included hyperventilation and talk preparation (Sierra et.al, 2011).

For this project it was a high priority to study the participants in natural surroundings. An attempt to achieve that was by creating a stress detection by using the questionnaires to construct a person-specific stress template and then use this to induce a non-stressed situation and a real-life stress scenario. Here the participants were asked to answer following questions; “What does stress mean to you?” and “Please describe what scenarios that typically makes you stressed?”. This means that the experiment consists of two user tests, developed individually based generally on the outcome of these questions. After the completion of each session, task difficulty, stress assessment, performance and user experience was collected through interviews (See Appendix B). In terms of subjective report, a common method is to allow participants to respond immediately after each completed task, but since this approach interrupts task flow, it seemed impractical in real-time (Chen et.al, 2013). It is also a risk of adding more load to the user and thereby interfere with the reliability of our physiological measurement. In the same way as Niculescu et al (2009), a similar method was thereby chosen after each completed trial. Unlike theirs who used the NASA-TLX report to determine the participant's mental and physical demands, own performance, effort and frustration, we opted a semi-structured interview. This was chosen as it allows a blend of closed- and open-ended questions and hence, more in-depth answers from each of the participants. It was also argued that such an interview would most effectively address any possible usability problems in the UI's since the participants could communicate freely about the user experience, and not only by using a point scale. The results of the user tests was then examined based on the hypotheses in our research.

4.2 Experiment

Interacting with our smartphones while additionally doing something else is not unfamiliar. Like Niculescu et.al (2009) study where they simulated a real scenario based on commonly occurring crisis situations, our experiment was designed on the basis of our multitasking society. Our experiment was divided into two sessions, both in which had two different user tests and three additional phases, “baseline”, “non-stress state” and “stress state”. The idea of using two different sessions was partly by testing the theory of an individual stress assessment yet also by comparing the effect of cognitive load when it’s reasoned to be higher.

Before starting the first user test, which consistently was the “baseline” and the “non-stress condition”, all of the participants was instructed to fill out a stress detection questionnaire. The idea of this was to be able to assess the participant current emotional state and how they consider themselves to react and respond to stress.

The participants were also told to describe which scenarios that normally makes them stressed and further explain what stress means to them in general and how often they experience it (See Appendix C). The scenarios they describe and the answer to what stress means to them individually, build the structure of user test number two, “stress condition”. The answers of the individual reactions to stress, both physically and mentally, were further used in the analysis of mobile interaction (emotional and behavioral). In the first trial, four of the participants stressed situation was developed by their individual response, but since it turned out to be a common pattern of what was considered as a stressed situation, it was possible to conduct each of the user test quite similar between the participants. These responses gave us parameters such as, walking while interacting, environments with other people, noises and time pressure.

This discovery laid the foundation of testing our theory with the need of an individual stress detection and thereby, it was set to examine whether the remaining three participants also experienced stress through these common patterns, even if they responded differently in the questionnaires. By this we were able to determine if all the participants experience stress by the same reasons or if it’s highly individual.

The non-stressed situation was always the first test in each of the trials and it started with a recording of the participants “baseline” where the participants HRV and GSR

values were measured without any interactions. In order to get reliable HRV values this measurement had to be approximately 4 minutes (Elite HRV, n. d.).
The non-stressed situation in trial one was consistently conducted at the participant's home, which everyone chose as the least stressed environment. The exact location inside the house and their position on the other hand, varied during the tests. Some of them chose to sit at the kitchen table, another in the living room sofa and one even chose their own bed. For the stressed situation, the environment was changed based on the location of the participant's home and was carried out while either walking on a bus, tram or grocery store. In trial number two, the non-stressed situation was conducted in other calming and quiet environments suggested by each of the participants, which varied from office room to a hotel lounge. The stressed situation was either conducted in a shopping mall or while walking in the street.

For the attempt of conducting the experiment in a real-life scenario, the basic idea was also to provide given tasks with certain realistics. One of the challenges as Niculescu et al (2009) highlights is the "first impression effect", meaning that participants might react more during their first trial due to uncertainty or unfamiliarity of the user interfaces and therefore, the tasks were designed based upon a scenario that probably many of us has experienced, namely booking train/-or bus tickets online. Another advantages of using this kind of websites is that they usually have similar user interfaces and thus we could hopefully minimize the change of provoking this reaction. Since our aim of the project is to investigate whether a stressed situation affects our cognitive load when it is initially low, the task couldn't be designed with too much mental effort. However, by the intent of distinguishing cognitive load and stress, the load must be able to be determined and thereby it can't be extensively low either. For the tasks in session one, four of the participants were instructed to find the cheapest ticket and the fastest public transport in a specific time interval, this by using three different UI's. Here the participants was forced to both multi-task between each of the UI's but also memorise numbers and calculate traveling time. By that, cognitive load was induced by trying to increase the memory set. When this trial was completed the participants had to open notes in the mobile interface and write down the correct answer, which was the time, price and company. In order to determine the effect of acute stress on an initially low cognitive load we had to try increasing the amount of load in another sequence and then further compare the differences between them.

This attempt was done in session two were the three remaining participants was instructed to perform the same tasks as the previous but along with more parameters. In addition to the above-mentioned tasks, the participants in session two was also instructed to find the most expensive ticket in the same specific time interval and finally through mental arithmetic, calculate the differences between these and the cheapest ticket. As by the first trial, all the participants had to open notes in the mobile user interface and write down the correct answer. This included the time, price and company for both the cheapest and most expensive ticket, and of course the difference. For the interaction all the participants were provided with an Android smartphone and had to use Google Chrome as browser. While the participants were performing the tasks, their performance data and physiological data was recorded simultaneously and after the completion of the task, task difficulty, stress assessment, performance and user experience was collected through semi-structured interviews.

4.3 Interviews

As this study aims to investigate the individual response of stress and cognitive load, it was desirable to hear the participants independent thoughts and therefore a semi-structured interview was decided to use for this project. It was chosen as we could prepare questions ahead but given the flexibility of asking follow-up questions when needed. The participants was able to freely express their views in their own terms and it provided us with the opportunity of hopefully gaining more in-depth knowledge of the topic in hand. One of the disadvantages of using a semi-structured interview is that open-ended questions are difficult to analyze and to compare, but the information obtained from an semi-structured interview, we still believe can be served to a greater advantage. Based on the interviews we investigated how the participants experienced the test in general, their own performance after completion of the task and their personal reactions during the non-stress and stressed situation. Three of the questions were related to the user interfaces where the participants were encouraged to elaborate the answers as much as possible, either by naming any missing functions or highlighting what they believed worked well in the user interface. The reasoning behind this was the possibility of addressing any usability problems but also as a desire of increasing our knowledge about cognitive load imposed by the user interface. It was also important to establish whether the participants experienced any discomfort using the physiological equipment and that is because a discomfort may affect the values during the measurement, therefore a question excluding this, seemed required (See Appendix B).

4.4 Apparatus

The Galvanic Skin Response was recorded using a Grove Sensor along with an Arduino Uno were two leads were attached on the participants left hand's fingers. The GSR Signals was collected and synchronised by an 14" ASUS E403SA laptop. Heart Rate Variability was recorded using a Polar H10 Heart Rate Sensor chest strap along with the application Elite HRV Version 4.2.2. An Huawei MYA-L41 and an Samsung Galaxy S7 was used for the mobile interaction and the interaction was recorded with a 720p resolution by using the screen recorder application DU Recorder Version 1.7.5.1.

5 Analysis methods

In order to assess and compare Heart Rate Variability (HRV) you usually need to look deeper and analyze time or frequency domains. The time domain measurement for detecting any ANS activity in short-term duration is usually by looking at the rMSSD (Root Mean Square of Successive Differences) or the SDNN (standard deviation of all R-R intervals) of the “HRV score”. The rMSSD is influenced by the parasympathetic nervous system (PNS) which means that a lower rMSSD is correlated with a higher score of sympathetic activity (e.g. higher stress level) and SDNN reflects and represents the total variability of all the responsible cyclic components in the heart rate (Goldstein et.al, 2011). In the frequency domain analysis, the spectral measures of HRV shows either high frequency (HF) values or low frequency (LF) values. HF has been argued as the marker of the parasympathetic activity were LF measures sympathetic activity. The ratio of LF/HF is what is known as the reflection of the balance between the sympathetic and parasympathetic nervous system, where a higher value indicates a dominance of sympathetic activity. This has been closely correlated with both increased workload as well as the amount of stress level (Healey & Picard (2005), and Cinaz et.al (2013)). According to Elite HRV (2018) the measurement needed to be longer than 4 minutes in order to get reliable LF and LF/HF results and they also suggested an open reading as the most useful HRV measure during a short-term activity. This was mostly because the other type of readiness was more situational, e. g morning readiness, exercise or meditation. After completing the tests, the values of time domain, frequency domain and heart rate results were exported to a Microsoft Excel document, because Elite HRV did not offer any comparison function between the different results, which made it difficult to analyze. Instead, Excel was used as a visualization method and the results were demonstrated through bar charts (See figure 5.1).

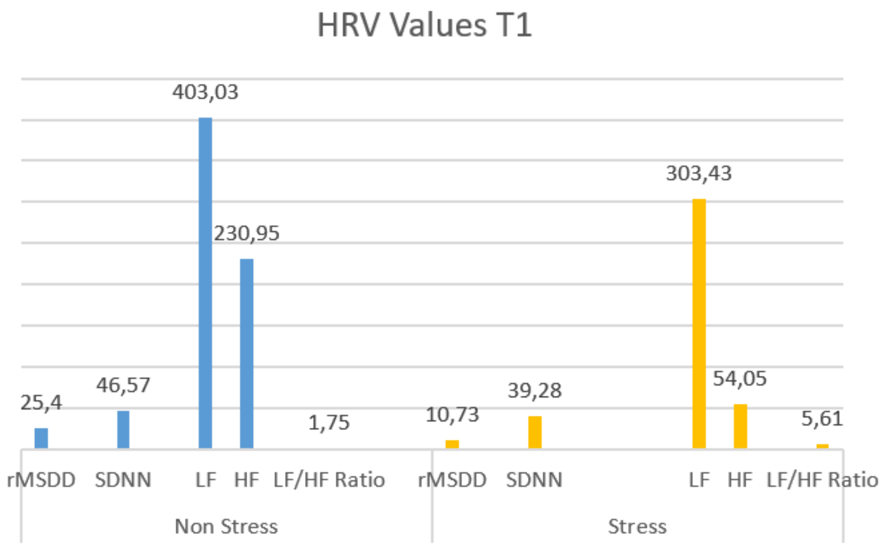

Figure 5.1. Example of HRV visualization of participant T1 across the states using bar-charts in Microsoft Excel.

These results were then set against the Galvanic Skin Response (GSR) results, whose data also was exported into an Excel document. The GSR data was measured at a sampling rate of 2Hz, meaning that the values were recorded every 500 ms. For the GSR analysis mean was extracted, as well as the max and min value in each of the test sessions. In the comparison of the different values in each of the tests, the entire sequences were visualized using a line diagram (See figure 5.2), mainly to easier detect high/low peaks or any other possible patterns in the measurement.

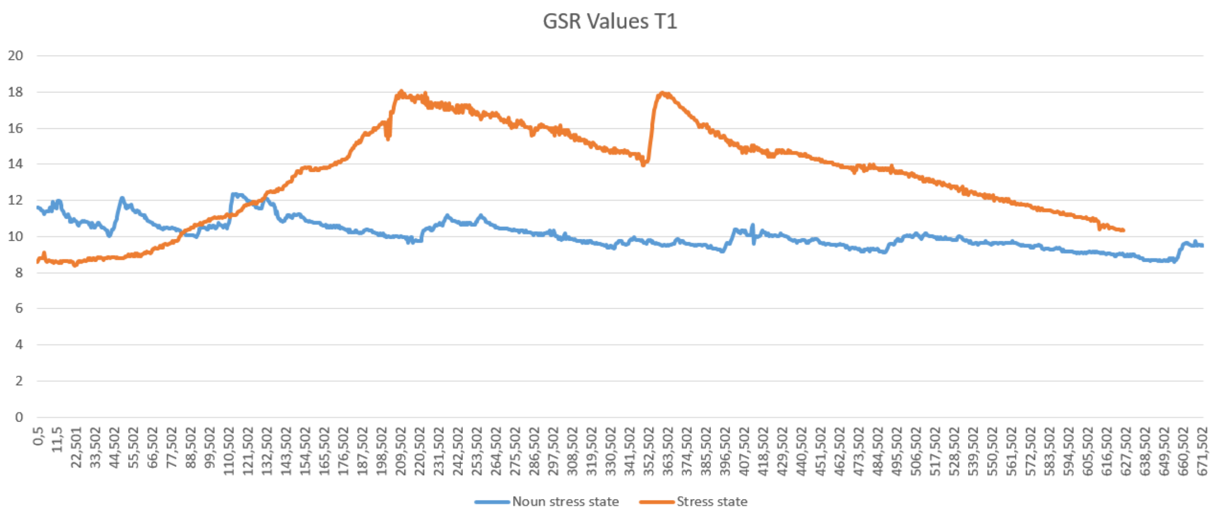

Figure 5.2. Example of GSR value visualization across the states using line diagram in Microsoft Excel.

These high/low peaks were mainly those used in our review of the GSR and HRV values, which means that when GSR suddenly increased or decreased, it was matched against the HRV values. The HRV values in this case was the heart rate (HR) and the variation in the RR interval and this is because both time and frequency domains were only summarized after completion of the measurement and not during real-time. However, since the heart rate and its rhythm are under the control of our autonomic nervous system, the variation can reveal the balance effects of the sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS) and when the SNS is maximally stimulated (e.g higher stress), the RR-Interval usually decreases and HR increases (Hye-Geum et.al, 2018).

As mentioned in earlier section, objective measurement does not alone explain the reason for a suddenly high GSR or HRV value, it also requires a subjective measurement. This means that in addition to our high/low peaks surveys, interview responses and interaction behavior with behavioral and performance measures were also analyzed and compared in this study. In that way we had a greater chance of explaining any sudden changes in the measurement or common patterns. The whole interaction during the tests was recorded by using a mobile screen recorder. This gave us the ability to carefully investigate the participant's interaction behaviour, in parallel with the absorbed sounds from the participant's response (eg word expression, exhale etc.), which again could be connected to their current emotional state. Essentially, the performance measure was inspired by Ritthiron & Jiamsanguanwong (2017) calculation method and by Albert and Tullis (2013) "user experience metrics". In Rhittiron et.al (2017) study they used success rate scores and percentage of efficiency as part of their measurement of web usability evaluation but unlike their study in which both focused on success rate and failure, our study chose failure rate as the only parameter of our measurement. The reason for this was mainly because we wanted to investigate whether the amount of errors correlated with the stressed situation, or by any other reasons. The efficiency were measured by comparing the completion time of the performed task between the different scenarios. The pre-time was calculated by the time taking to perform test in non stress situation and the post-time was calculated in the stressed situation. To compute the percentage of efficiency we used Rhittiron et.al

(2017) calculation by dividing the completion time in minutes and then multiply the answer by 100. By that we were able to get a percentage difference between the performed tasks in each session (See figur 5.3).

$$\text{Percentage of effiency} = \frac{\text{Pre-Time}}{\text{Post-Time}} \times 100$$

Figure 5.3. Calculation method of the performance measure

The number of errors was analyzed in the video recording and it was calculated by the number of clicks which either led the user to another user interface or feature, actions that gave the user incorrect results, or generally clicks that interfered with the performed task. In the behavioral measures we used hesitations and breaks as parameters and these were calculated by the number of breaks in the interaction and by hesitations such as; switching back and forth, rewriting inputs or when the user makes one choice but repents. In addition we extracted verbal feedback for the emotional responses. Notes were written down manually in the observations, but it turned out to be virtually impossible to connect these with the exact time interval from the video recording and the physical measurements. Thereby, the emotional responses from the participant's were only analyzed based on the sound expressions and feedback from the interviews, and not by any facial expressions or body language. The time between the high/low GSR peaks was compared to the time in the HRV measurement and then along with the time sequence in the video recording. Between the GSR and the HRV recording, a difference of +- a second was taken into account in the analysis. As for the final result of the user tests, the questionnaire was used as a baseline for the expected physiological measurement of each participant.

6 Ethical consideration

Our experiment was approved by the Norwegian Center of Research Data (NSD) and performed in the accordance with NSD guidelines and the inspiration of Bryman and Bell’ s (Research Methodology, 2018) ten most important principles related to ethical considerations. A fundamental right in the Norwegian Law is that the researcher is obligated to inform the participants about the personal data they are going to collect. The main rule is that you have to gain consent where the importance lies in the fact that the participants understands what the participation involves, what they are consenting to, and which consequences consenting to participate in the research will have (Norwegian Center of Research Data, 2018). In this project, this was done by using the recommended template from the NSD. This sheet had to be signed before starting the experiment but despite this, the participation is voluntary, which means that the participants can withdraw from the study at any time. A signature on a consent form alone does not guarantee the research ethical consent, it is actually in the researcher's responsibility to treat data accordingly. The data in this project has been treated both accordingly to NSD recommendations but also with inspiration from Bryman and Bell's (Research Methodology, 2018) ten most important principles regarding ethical aspects of dissertations. Since this project assess some form of personal data like age, gender, health condition and biometric data, the protection of privacy was of high importance. To ensure both privacy and anonymity, all raw data from the GSR was saved on an external hard drive. The HRV data was saved in an application on a smartphone and protected by a four-digit code and login credentials. Data from questionnaires and interviews were collected through Microsoft Word and further stored in a separate folder on a laptop, which also was protected by login credentials. None of the names of the participants was disclosed, instead they were assigned a letter followed by a number as an identification. This identification was consistently used in all of the data collections in the project.

7 Results

The subjective measures was conducted by a questionnaire and then by interviews after each session. The interviews gave feedback to the evaluation and in trial one and two, the questionnaire provided the basis for the development of the user tests. In trial two our aim was to test the theory of an individual stress assessment yet also by comparing the effect of cognitive load when it's reasoned to be higher, and thereby the user tests in trial two was conducted equally regardless of the individual responses. To distinguish cognitive load and stress we used feedback from the interviews and examined the relation between the objective measures and interaction behavior from the video recordings.

7.1 Subjective measurement

In order to gain more understanding of how each of the participants deals with stress in their daily lives we wanted to know how often they experience it. Two of the participants experience stress once a week, two of them every now and then, one experience it several times a week, another daily and the latter experience stress rarely. For their personal stress reaction, all of the participants except one stated "irritation" as one of the main psychological response and sweaty or increased heartbeat as the physiological response. The remaining participant stated anxious, controlled, short patience and jumpy. In trial number one, T1-T4 had all similar responses of describing a stressed situation. In trial number two one of the participant response (T9) could be corresponded along with the others but the remaining two participants (T6 & T7) had only similar descriptions among themselves and not in correlation with the conducted stressed situation in the user test. Both T6 and T7 admitted that they were more stressed before the test and explained further that it was mainly due to a reaction of not knowing what to do or not having any control, which is backed up by their description of a stressed situation. Due to previous experiences both participants perceived the tasks as easy and claimed that the most challenging of the test was the unfamiliarity of using an Android smartphone. This was also stated by T4. In comparison between the non-stress (test one) and stress situation (test two) T6 and T9 stated that the mental arithmetic were more challenging in test number two but in T6's personal response, she felt more stress in test number one. T7 stated that the task were even easier to perform in test number two and personally felt more calm than in test number one. This is also backed up by T2 who stated that the tasks were easier to perform in test number two because she became more comfortable with the user interfaces and with the assignment. However, T2 claimed that she was more irritated and had less patience in test number two, which corresponds with her individual description of stress reaction, and thereby T2 was more stressed in test number two. This also turned out to be true for participant T3 who did not have any previous experience with these user interfaces and thereby perceived the task as easier in test number two, yet she became more stressful and anxious due to more things to manage and to keep track of.

Both T1, T3 and T9 experienced that the task itself was not affected by the simulated stressed situation but they unanimously agreed that test number two was more challenging due to the interaction while walking and the amount of other external factors or distractions. In the personal response T9 stated that her own performance was affected especially during the mental arithmetic and T1 became more confused and disoriented in test number two. T1, T3 and T4 had the least experience and knowledge of the different user interfaces used in the experiment and this can probably have made an impact on the "first impression effect", which has been previously mentioned as a potential problem when measuring the amount of cognitive load.

In summary, more than half of the participants were more negatively affected in their personal responses during the stressed situation, but perceived the task as easier due to increased self-confidence and more familiality. Finally, all of the participants had no prior experience using an android smartphone and this was highlighted during many of the interview responses.

7.2 Objective measurement

In our second user test where we contextually simulated a realistically stressed scenario, all the participants were instructed to walk while performing the tasks. This we assume could have an influence on the overall results of the HRV measurement since HRV both can be affected by physical activity and other internal and external circumstances (Hye-Geum et.al, 2018). In addition, HRV is highly affected by individual preferences, such as age, the overall body functions, as well as lifestyle, and that's why it is also unique and slightly different among people. The same individuality applies for GSR measurement.

7.2.1 Heart Rate Variability

As expected, both an increased average heart rate and a lowered HRV value were shown in the results of the stressed situation, but the values of time & frequency domains produced widespread individual differences. Some researchers claims that both mental effort and stress are particularly shown in the Low Frequency (LF) (Harris et.al, 2017 & Lyu et.al, 2015), while others states that there are not enough strong evidence in the assumption of this relationship (Billman, 2013 and Heather, 2014) and that you should rather be looking at the High Frequency (HF) band in order to assess stress. While Dongsoo et.al (2014) states that workload increases the power of LF and decreases HF, Harris et.al (2017) claims that mental effort rather reduces the LF due to an increased sympathetic regulation, which further was discovered in Tozman et.al (2015) study. In Goldstein et.al (2010) study they demonstrated that LF had no representation of sympathetic activity. Dongsoo et.al (2014) continues by stating that acute stress is also predicted to be associated with an increase of the LF/HF ratio, which further has been proven by Healey & Picard (2005), and Cinaz et.al (2013) studies. Shaffer & Ginsberg (2017) states that the RMSSD is correlated with HF power and that a lower HF is correlated with stress.

In our study, all the participants showed an increased LF/HF ratio and a decreased rMSDD (See Figure 7.1), which is a result consistent with Healey & Picard (2005), and Cinaz et.al (2013) studies.

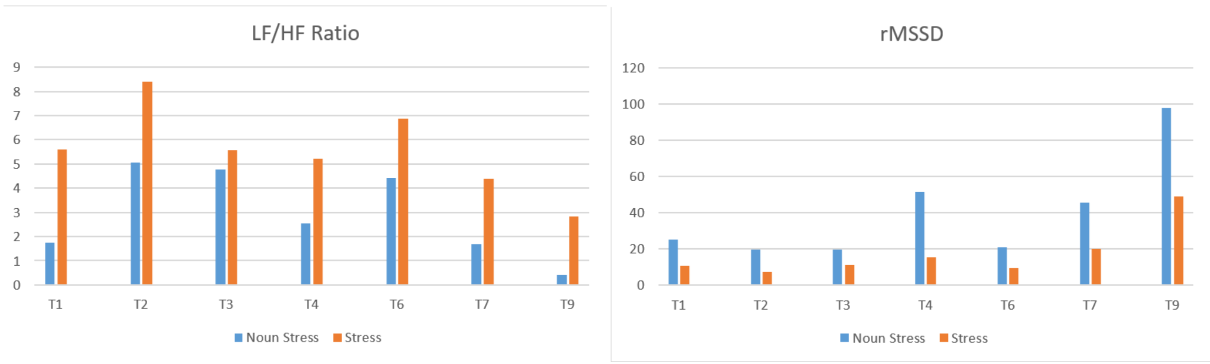

Figure 7.1.This illustration shows that the LF/HF ratio increased and rMSSD decreased and during the simulated stress scenario.

Four participants (T1, T2, T3 & T9) showed a decreased LF during stress state, one (T4) had lower LF during non-stress and the remaining two participants (T6 & T7) during the baseline. In addition, these results had a great correspondence to the participant's subjective results.

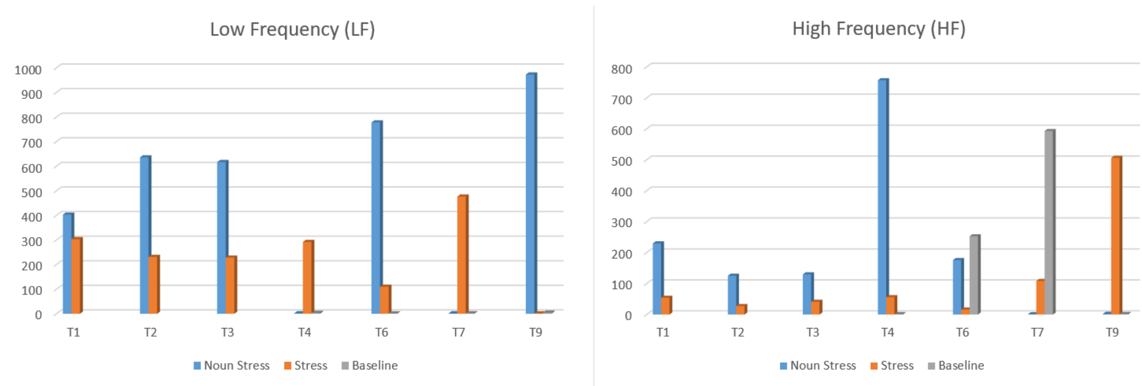

Figure 7.2. This illustration shows that LF is in corrospondence with HF, which means that when LF decreases, HF decreases.

As for the time-domain variables, SDNN has strongly been correlated with LF and for that reason it should reveal corresponding results in the measurement with a decreased value during stress state. However, this turned out to be only consistent in four of the participant's results. Participant T3 who scored lower on the Low-Frequency (LF) band in stress state, had lower SDNN values in non-stress state, participant T7 who scored lower on the LF during baseline, had lower SDNN values in stress state and participant T4 who scored lower on the LF during non-stress state, had lower SDNN values in stress state

(See Figure 7.3).

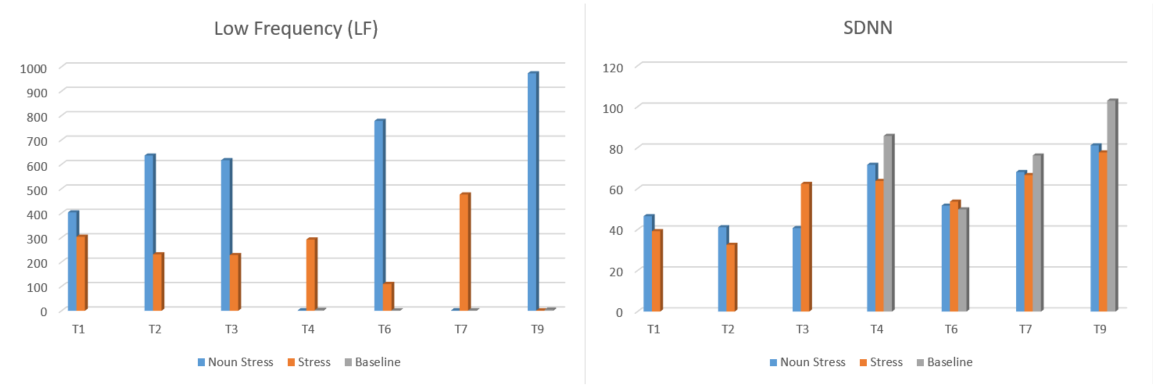

Figure 7.3. This illustration shows that LF is only in corrospondence with SDNN for participants T1, T2, T6 and T9.

The same was revealed in the correspondence between rMSSD and a reduced HF (See Figure 7.4).

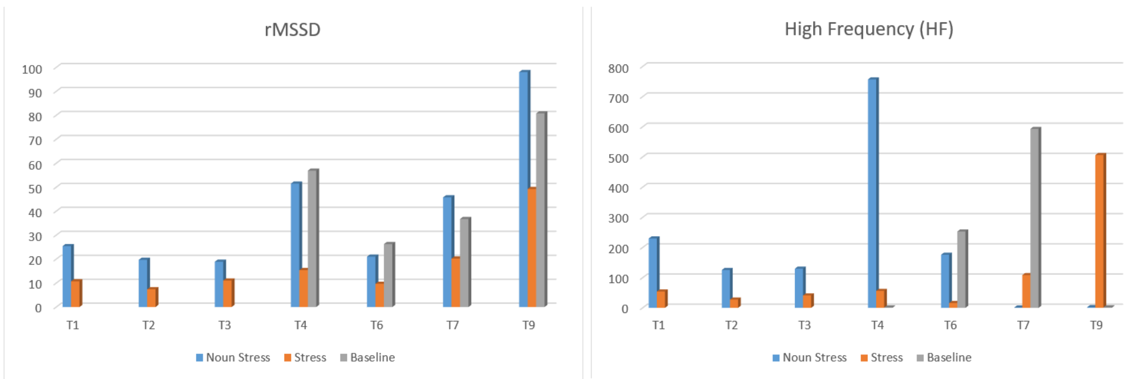

Figure 7.4. This illustration shows that HF is only in corrospondence with rMSDD for participants T1, T2, T3 and T6.

This is one of the biggest challenges when measuring and analyzing HRV results, as Hye-Geum et.al (2018) emphasizes, by being able to connect these values into an understandable reasoning, in which they claim is due to inconsistent results. This was proven in their own survey where they discovered that the presence of significant variation in HRV variables was inconsistent in several other studies (Hye-Geum et.al, 2018). In summary, our HRV results demonstrates that the rMSSD and the LF/HF ratio are most consistent, but the low frequency domain (LF) seems to be the one with strongest connection to the sympathetic activities (e. g stress level).

Another approach is to examine the relationship between the RR interval and stress, which has previously provided reliable results. In Sloan et.al (1994) study they discovered that by examining the relationship between RR Interval and the response of stress, they were able to find that a reduced RR interval was the index of a higher stress level. In our analysis, the variation of the heart rate and the RR-Interval were cross-examined against the variation in the GSR values.

7.2.2 Galvanic Skin Response

In the GSR measurement we examined both the average of the GSR (mGSR), maximum value and the number of peaks. Four of the participants (T1, T2, T4 & T9) had higher scores on the mGSR and max values during stress state and significantly lower scores in the non-stress state, even though T4 scored higher on the number of peaks in the non-stress state. For the remaining three participants (T3, T6 & T7), GSR were significantly higher in the Baseline (T6) and in the non-stress state (T3 & T7), even though it hypothetically was supposed to be lower (See Figure 7.5 & Figure 7.6).

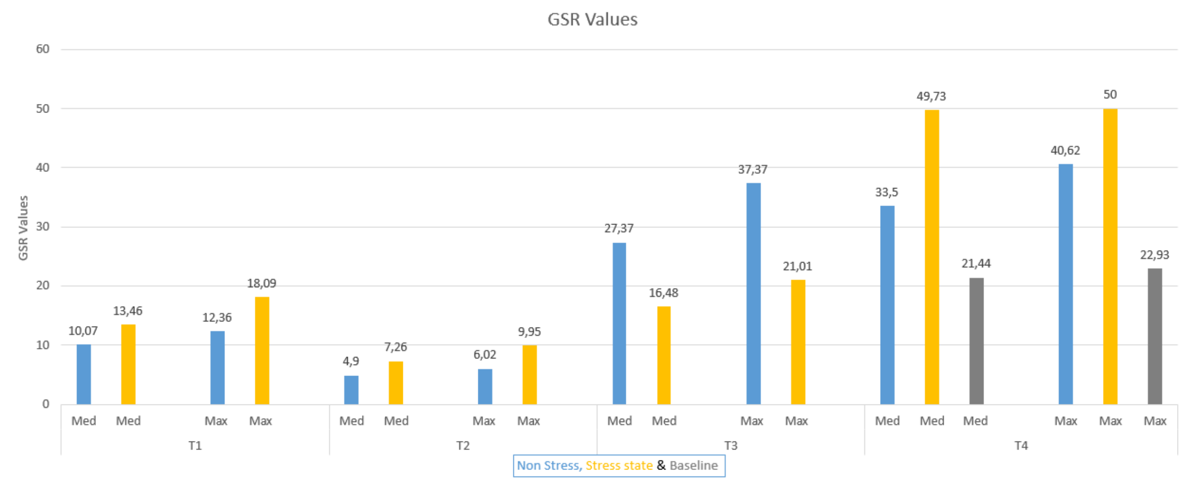

Figure 7.5. This illustration shows the mean (med) and max GSR Values of participants T1 - T4.

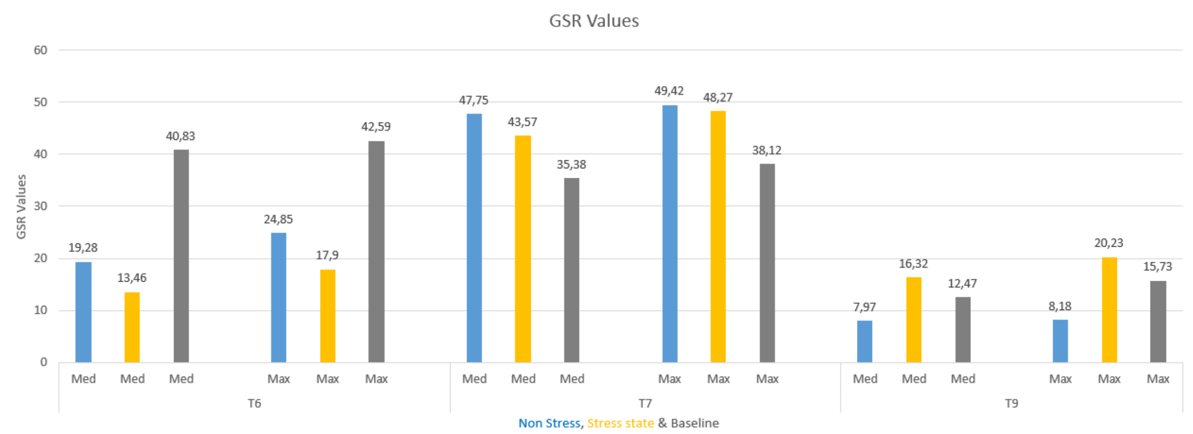

Figure 7.6. This illustration shows the mean (med) and max GSR Values of participants T6 - T9.

After cross-examining this GSR data along with the video recording of the subject performing the task and from the subjective measurement, we reason that this could be corresponded with one of the subject's unfamiliarity of the user interfaces (T3) and due to the fact that both T6 and T7 claimed that they were more stressed before and during test one. The number of peaks varied between the participants but for the majority the variation was higher in non-stress state or baseline. In trial one, the number of high/low peaks showed a greater variation for all of the participants (T1-T4) in non-stress state, but the degree of the increased values between these peaks was significantly higher in stress state, especially for T1 and T2. For participants T3, both number of peaks and high values were shown in non-stress state while T4 generally scored higher on the GSR values in stress state (See Figure 7.7 & Figure 7.8).

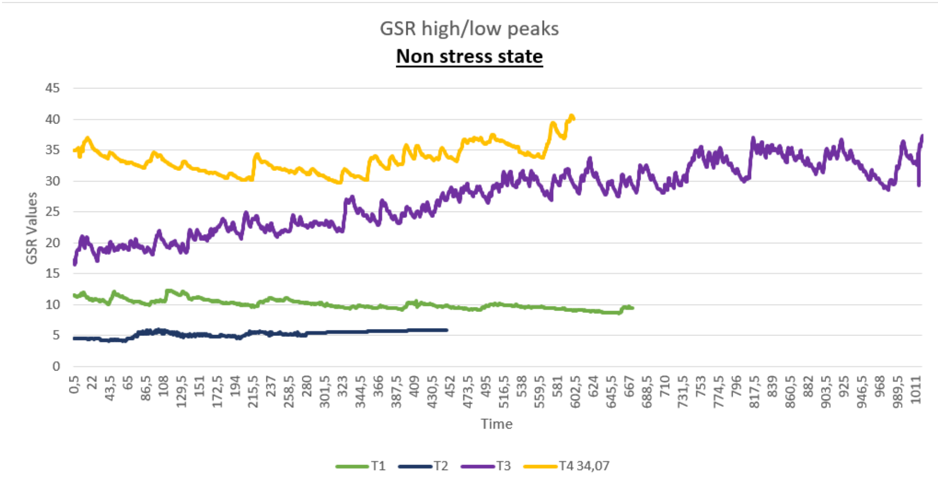

Figure 7.7. This illustration shows the participant T1-T4 high/low peaks during non-stress state. The time is illustrated every twenty-second due to the lack of space.

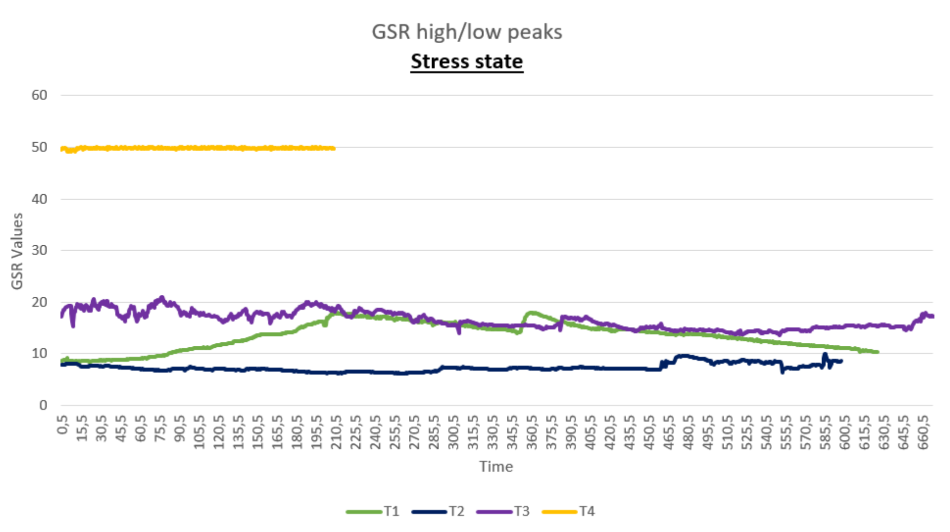

Figure 7.8. This illustration shows the participant T1-T4 high/low peaks during stress state. The time is illustrated every twenty-second due to the lack of space.

After reviewing this along with the video recording and feedback from the interviews, T4's response could also be justified for the same reason as T3, which is a lower experience of the different user interfaces used in the performed tasks.

In trial number two, T6 had the largest number of peaks and a higher degree of GSR values during the baseline while T7 and T9 scored higher in stress state. However, both T7 and T9 scored significantly higher on the number of peaks in baseline than in non-stress state and even though they had higher GSR values in stress state, T9 actually scored higher in the baseline then in non-stress state, were it should rationally be lower (See Figure 7.9, Figure 7.10 & Figure 7.11). One reason that may have contributed to this effect is that this participant had to perform test one twice and by a second attempt T9 probably got more confident and the task became less cognitively demanding.

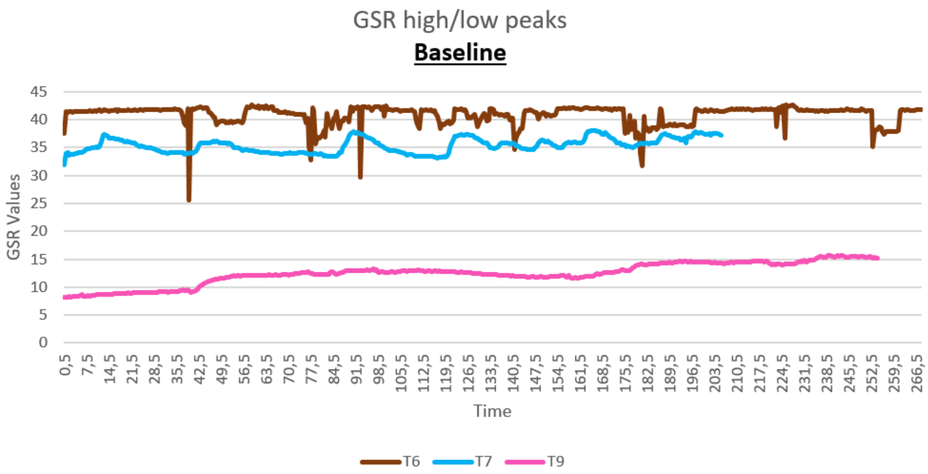

Figure 7.9. This illustration show participant T6-T9 high/low GSR peaks during baseline. The time is illustrated every twenty-second due to the lack of space.

.

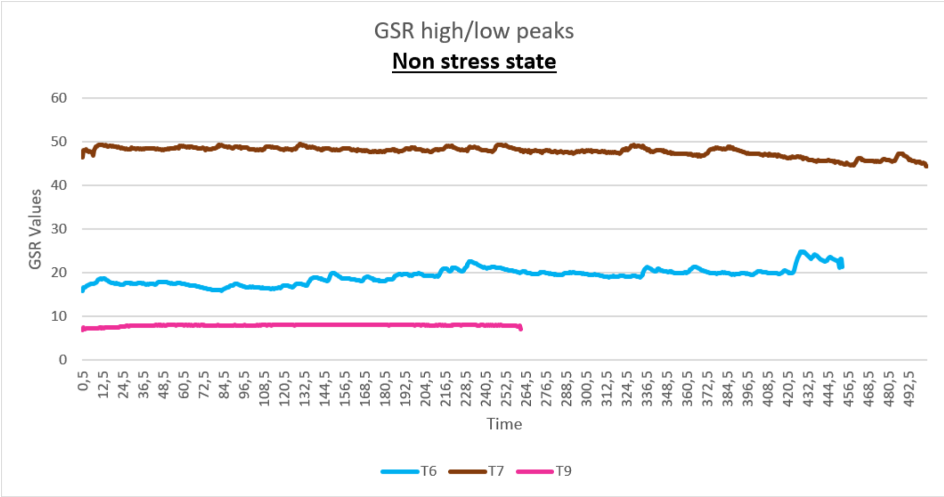

Figure 7.10. This illustration show participant T6-T9 high/low GSR peaks during noun-stress state. The time is illustrated every twenty-second due to the lack of space.

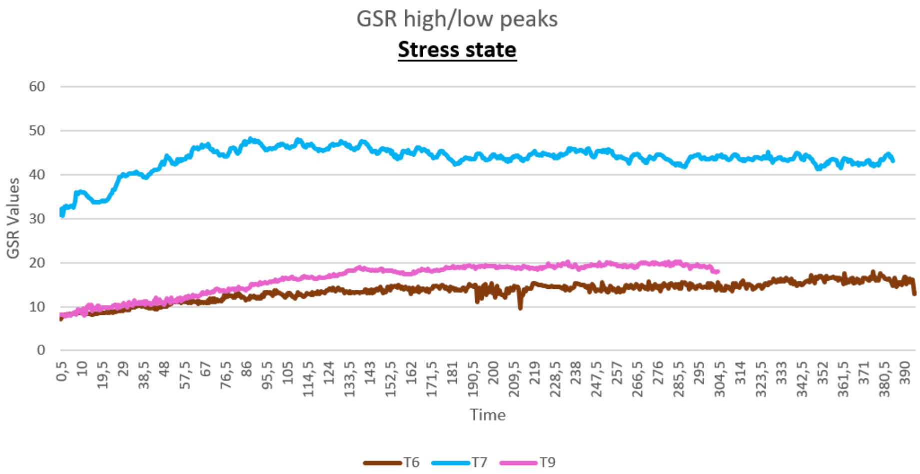

Figure 7.11. This illustration show participant T6-T9 high/low GSR peaks during stress state. The time is illustrated every twenty-second due to the lack of space.

In summary, the GSR had higher values when the participant stated higher stress but for the majority, the variation of high/low peaks was higher variation during the non-stress state or baseline. This confirms our assumption that the participants probably were affected by a "first impression effect".

7.2.3 Cross analysis

In our cross-analysis of the GSR and HRV values, we focused mostly on the high/low peaks in both of the measurements. For the GSR, a deeper analyzed was made when the GSR values became significantly higher or lower and at the exact time of the event, we matched those values up against the variation of the heart rate and the RR-Interval. The overall result revealed that GSR often corresponds to the variation in the HRV values, especially during test two which was the simulated stressed situation. During the baseline measurement with no mobile interaction, the GSR values showed to be clearly affected by the participant’s emotional response like laughter, deep inhalation or when talking and while cross-examining the behavioral measures in non-stress and stress state, this turned out to be true even in these conditions. GSR alone seemed to be only affected when the participant interacted with the UI's in silence and with no heavy breathing. Another discovery was that T9’s GSR values increased in trial number two where the participant performed the mental arithmetic and the variation appeared larger in the GSR measurement then in HRV. For the latter participants in the second trial (T7 & T9), both GSR and HRV had high values and variations during this session. When analyzed deeper, a reasoning of the low HRV variation for participant T9 might have been caused by a standing position while T6 and T7 were on-the-go during this sequence, and thereby causing a higher HRV variation. Another finding was that the GSR values decreased during breaks in the interaction and during total silence, which of course we were hoping for, yet otherwise there were more occasions where only HRV was affected or both simultaneously. In several occasions it was an increased variation in the HRV when the participants had a higher degree of multitasking, but without any verbal feedback. If the participants talked during this time, the GSR values increased.

7.3 Performance results

Performance-based approaches usually measures cognitive load and stress level by monitoring the achievements of the subjects, such as completion time and the rate of correct answers, in performing an activity (Nourbakhsh et.al, 2017). In our experiment we analyzed the performance measure by dividing the completion time in minutes and then multiply the answer by 100. By doing this we were able to get a percentage difference between the performed tasks in each session. Five of the participants performed the task faster in the simulated stressed situation, where 3 of them had over 50% performance increase and two of them around 30%. The remaining two participants performed the task quicker in the non-stressed situation and performed the task 12% slower in the stressed situation (See figure 7.12).

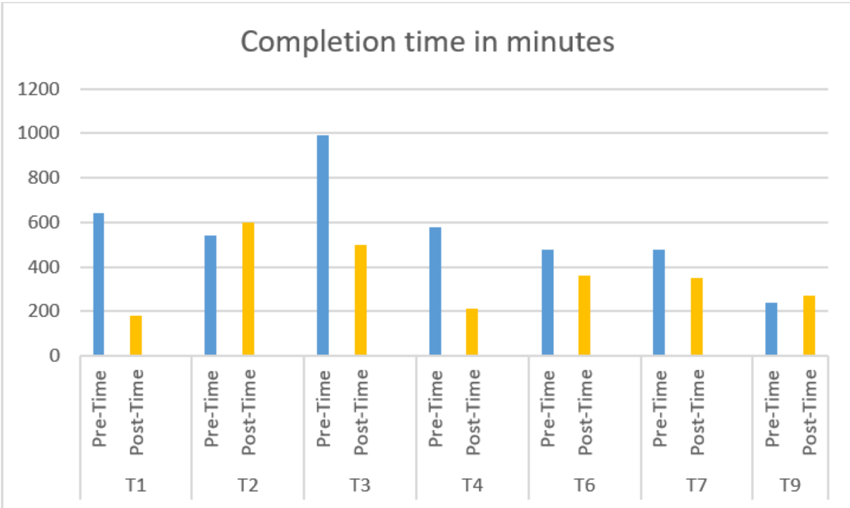

Figure 7.12. Completion time in minutes between the participants across non stress state (blue color) & stress state (orange color).

While 4 of the participants (T1,T3,T6 & T7) had higher number of errors in non-stress state, only one (T2) had significantly higher score in stress state and the remaining participants (T4 & T9) had equal number of errors (see Figure 7.13).

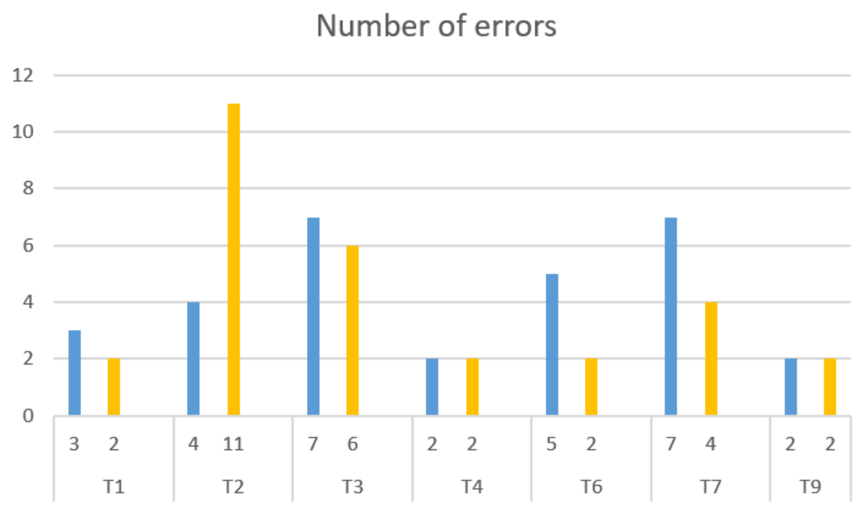

Figure 7.13. Number of errors across non stress state (blue color) & stress state (orange color).

It turned out to be an increased number of hesitations and breaks during non-stress state (see Figure 7.14 & Figure 7.15) but looking at the overall results of the performed task, 5 participants got the wrong answer, especially during the stressed situation.

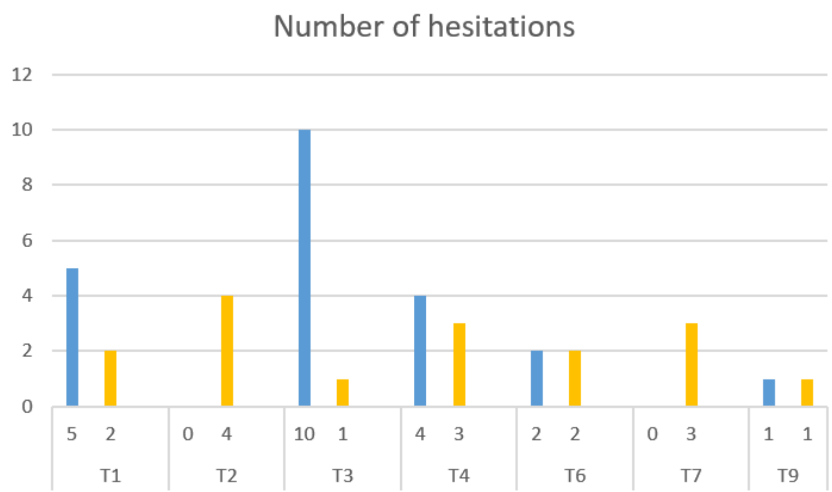

Figure 7.14. Number of hesitation across non stress state (blue color) & stress state (orange color).

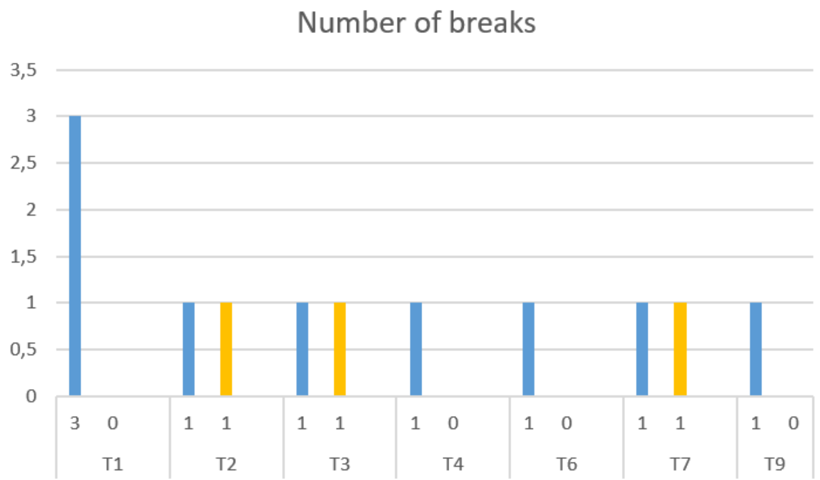

Figure 7.15. Number of breaks across non stress state (blue color) & stress state (orange color).

7.4 User interface evaluation

Elting et.al (2002) states that if the content is not too complex, then how we present the information will not contribute to an overload. Whitenton (2103) claim the opposite and states that the only way of reducing load is by looking at the visual representation of the content, which is further proven by Gwidzka (2010) and Niculescu et.al (2009). In our study we used three different UI's while the participant performed the tasks and since one of these UI's actually change their design between the first and second trial, we had the opportunity to investigate whether these changes might have affected the objective and subjective measures. In trial number one, all of the participant's GSR values seems to repeatedly show an increase while they are entering one of the UI's for the first time, and this pattern also appears during the second trial. However, the amount of variation and the quantity of the GSR increase appeared to be larger in trial number one than in the second trial where one of the UI's had been redesigned. Although the GSR values did not have a significant difference between them, this result could also be linked to the subjective results where many of the participants only complained about this particularly interface in trial one and not during the second trial. At this point, this UI had the least resemblance between them and didn't get the same degree of similarity in the structure and the design as the others, until it had been redesigned during the second trial (see figure 7.16).

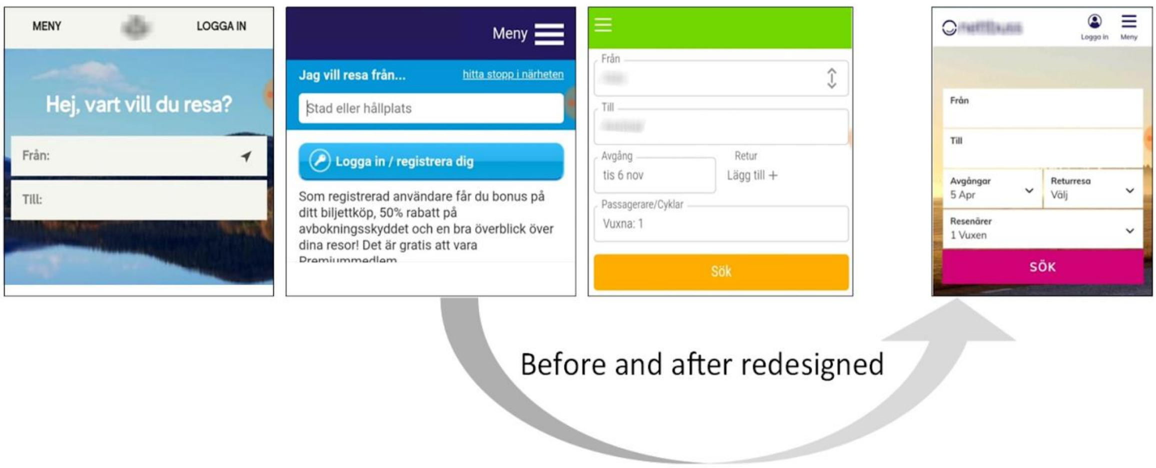

Figure 7.16. Illustrations of the user interfaces used in the experiment, where one was re-designed in trial two.

In Gwidzka's (2010) study, they discovered that all the participants mental effort increased as the extraneous load became more complex and not due to the task itself. In this case, the complexity of the UI is probably not an issue, but the result of our GSR measurement still indicates that a resemblance decreases our effort. Which might be a discovery that could act as a resource when designing user interfaces in the future. Halarewich (2016) highlights that many designers has prompted Miller's concept of "chunking" to enhance user experience and reducing workload, especially when designing menus and organizing web content (Halarewich, 2016). Looking at the user interfaces in our experiment, one can clearly observe that the ideas behind this concept has been used, particularly in the search results (see figure 7.17).

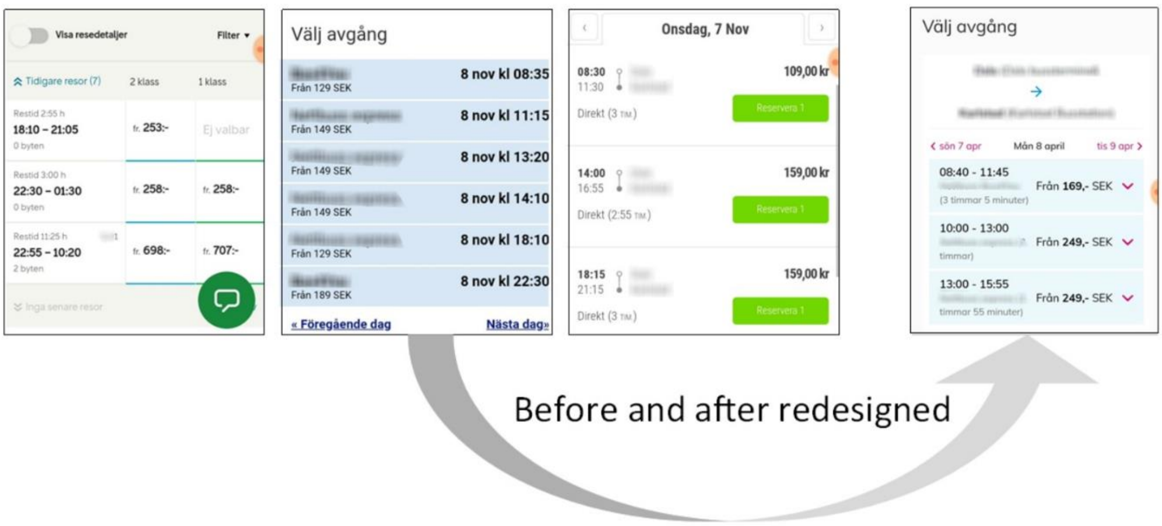

Figure 7.17. Illustration of the search results between the three user interfaces.

According to Miller's concept, the limited capacity of our working memory is approximately by holding 7 items at a time and this has clearly been adopted in one of the UI's in our experiment. In the event of an exceeded limit, they solved this by hiding the remaining items using links. However, as none of the participants in our study clicked on the given links, this solution also contributed to failure of the performed tasks in the user tests. As for the GSR measurement, the values seemed to be decreased when the content revealed fewer results but this were only more evident in comparison directly between the results of the same UI (see figure 7.18). Otherwise, the GSR values showed no major differences in the total amount or variation.

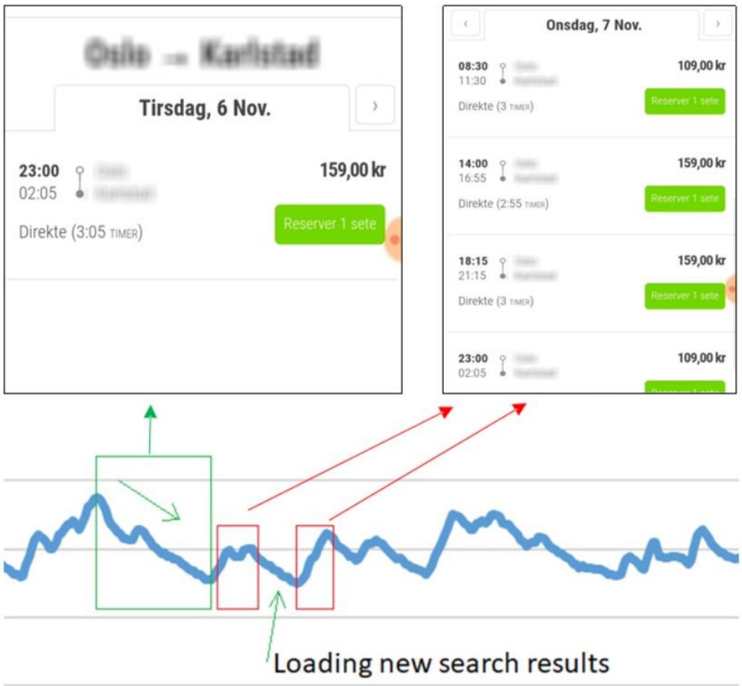

Figure 7.18. Illustration of GSR values increasing with fewer search results and increasing when larger amount.

From the evaluation of the subjective and performance measurements, all of the participants claimed that the mobile user interface was the biggest issue during the tests. As part of the tasks, one of them was to find notes in the mobile interface and when analyzing the video recording, this step also seemed to be the most challenging among the participants. In this case, we assumed that this could possibly contribute to an increase in the GSR but in fact, there was no significant changes at all during this interaction and two of the participants values actually decreased at this point, which again verifies that a subjective point of view does not always correspond to how you respond physically or mentally.

8 Discussion

As expected, both an increased average heart rate (HR) and a reduced HRV value were shown in the results of the stressed situation, but the values of time & frequency domains all produced inconsistent results except the low frequency band (LF), which also had strong connectedness to the participant's subjective results. As these metrics are computed and summarized after completion of the measurement, we explored the differences between the R-R Interval and heart rate (HR) and discovered that a lowered R-R and a high HR was linked to an increased stress level but also by emotional states (e. g arousal, laughter). A greater variability between heart beats, e. g a higher HRV usually means that the body has a strong ability to tolerate stress and here all of the participants had lower HRV during the simulated stress scenario (See figure 9.1), which indicates that our manipulation of creating a stress template was successful.

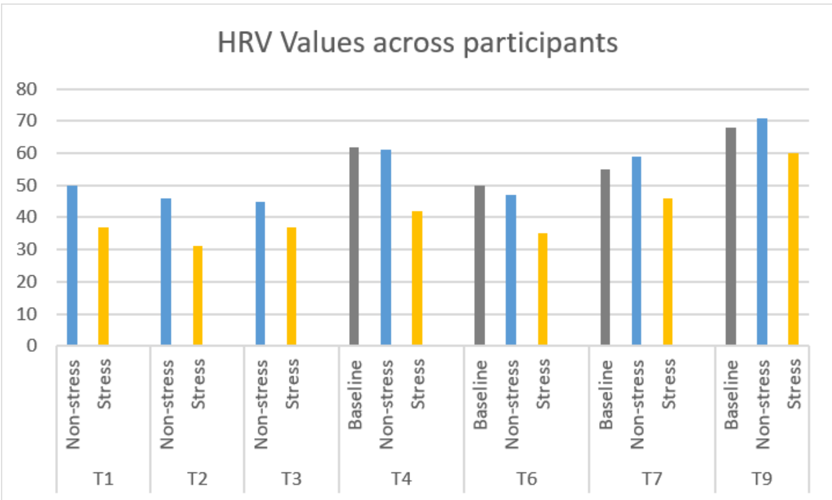

Figure 8.1. Illustration of the HRV Values across the participant's.

Although this result may have been affected due to the large differences between test one and two, where one was sedentary and the other on-the-go, all of the participants except T6 and T7 experienced a higher stress level in their internal responses, which was verified by the subjective feedbacks. For T6 and T7, who participated in a stressed scenario that was not based on their questionnaire, their results showed higher stress level either during baseline or in non stress state, which strengthen our hypothesis that a subjective approach against stress assessment is necessary.

We hypothesized that a stressed situation would increase the cognitive load even if the task load is low and that the user's task performance would decrease when stress level increases. In our result, all the participants except one (T3) had higher GSR scores during the experienced stressed scenario, which in this case might have been affected by this subject's unfamiliarity of the user interfaces, and thereby required more of the mental effort. Even though the overall score of mGSR and maxGSR was higher during stress state, the variation of GSR high/low peaks turned out to be larger in trial one during non-stress state. During trial number two, all of the participants had larger variations during the simulated stress scenario, and two of them had higher GSR values during the simulated non stressed scenario. In the results of our performance measure, 72 % of the participants performed the task more poorly during the non stressed scenario, which means that they had higher scores of errors, breaks and hesitations, but the results also revealed that the same amount failed with the given task mainly in the simulated stress scenario and not during the non stressed condition. The participants didn't perceive the interaction as more difficult during the simulated stress situation, instead they were more affected by their internal responses. In Niculescu et.al (2009) study, all of the participants felt more frustrated and insecure during the first trial than the second and as shown in their results, our stress manipulation was clearly disturbed by a "first impression effect" where the majority of our participants perceived the second user test as more easy, even though they were more rushed. Although the participants in our study was provided with a brief introduction about the given task, this effect could have been avoided with a conducted training session before starting the experiment, which is also posited by Nisculescu et.al (2009). In addition, this approach could also have prevented the participants from asking questions during the actual task, which in our case affected the GSR values repeatedly and made it difficult when distinguishing the amount of load and the causality of the experienced stress level. We therefore suggest this kind of an approach in future experiments.
It has been previously confirmed that GSR has been used to detect mental state, emotions and stress (Ayata et.al (2017), Nakasone et al (2005) & Setz et.al, (2010)) and therefore one of the challenges in this project was the likelihood that cognitive load and stress would reflect on the internal reponses equally and by our results, GSR clearly shows high sensitivity to various reactions. However, it did appear as the GSR values

decreased while the participants stopped interacting and was stand-alone in those cases when interacting in silence and with no emotional responses. This suggests that the GSR can act as an index for cognitive load while interacting with technology, but when carefully controlled. With respect to extraneous load, we also believed that one of the UI's would pose lower GSR values when redesigned and this was confirmed yet with low statistical significance. Based on this project research question, our findings confirmed that the GSR was increased while interacting on mobile web during a higher stress level, but it’s currently uncertain to what extent these values are affected.

9 Conclusion

Since the increase in the GSR value corresponded to the perceived feeling of stress, yet without any increased mental effort, we can conclude that this confirms our first hypothesis saying; a stressed situation increases cognitive load even if the workload is low. Our second hypothesis was the assumption that the user's task performance would decrease when stress level increases and here our results showed that the number of errors, hesitations and breaks was rather more affected by the "first impression effect" than the level of stress. While the stressed situation did not contribute to a decrease in performance, it still contributed to the fact that the majority of the participant's performed the task incorrectly. Thereby we can confirm that a stressed situation in the end adversely affects our performance and interaction behavior negatively. However, what we cannot determine in this survey is in what degree acute stress affects our cognitive load (CL) while interacting on mobile web, which is very likely due to our own interpretability but also the lack of knowledge towards other statistical software, which probably could have been served as a better usage for this kind of survey. Therefore, this project is not able to answer the first research question with accuracy and hence not successfully. Our second research question has been covered throughout this thesis with multiple scientific evidence but also with recommended guidelines to minimize the risk of a cognitive overload, mainly imposed by the user interface. In our own evaluation, we found out that "similarity" seems to contribute to such a reduction.

It has also been established that it requires multiple measures to distinguish cognitive load and stress and it has been confirmed that an individual template is needed for stress assessment. Although it can be assumed that humans are aware of our cognitive processes and mental effort, a subjective measurement alone can not give the information of which processes that causes the perceived amount of cognitive load. However, it would not have been possible to determine the causality between the objective measurements and the experienced load / stress if we had not used a subjective measurement as a additional measure. Our explorations has shown a promising result of using GSR as an index of cognitive load and HRV as an index of stress assessment, but only when controlled carefully and along with additional measures.

While this paper contributes to the creation of an individual stress template and an experiment with empirical methods, more experimentation should be conducted for further survey of the differences between the two physiological measurement and the causality. We acknowledge that this study has some limitations both in terms of our GSR equipment but also our own interpretation ability of the physiological equipment, as we have no previous experiences in this field. The GSR equipment was of low cost and despite the fact that resistance acts as an inverse to skin conductivity, accuracy and sensitivity can still be questioned by the lack of this functionality. No baseline was performed on the first three participants, which also limits the possibility of verifying a low cognitive load with certainty. This was a function who were discovered later in the project. The biggest challenge in this study was mainly to find a useful method to distinguish cognitive load and stress but also to develop an experiment that gave us reliable results. After completion of this project, we learned that a more thorough investigation is needed in the results of other resources. By doing this properately, we had gained access to important information much earlier in the project and thus also avoided making the same mistake.

For future work it would be advisable to consider the following recommendations: Use an individual stress template along with multiple measures in order to assess cognitive load and stress, have training sessions to avoid "first impression effect", measure the baseline after the completion of the test, as it has been shown that the participant experiences least tension during this period and as a final suggestion: increase the number of point scales in the semi-structured interview in order to increase the percentual degree of the perceived load/stress. In the regardance of simplicity for the analysis, we also propose more experimentation to explore different wordings of the questions in the survey. Although the study has been able to prove certain patterns and results, it would also serve a greater statistical significance with more participants.

In summary, the current experimental design of this study can be considered as a good starting point and we hope that this thesis can serve as a resource for those conducting further research in this field.

Acknowledgement

I would like to thank all the people who participated in this survey for their effort and time, and a special thanks to my supervisor Pietro Murano who has guided me throughout the project. I would also like to thank MakerSpace at Oslo Metropolitan University for their technical expertise and realization of our GSR equipment.

10 References

Albert, W., Tullis, T. (2013). Measuring the User Experience. Collecting, Analyzing, and Presenting Usability Metrics. Second Edition. Chapter one: What is User Experience, page 4. Retreived from Google E-books https://books.google.no/books?hl=sv&lr=&id=bPhLeMBLEkAC&oi=fnd&pg=PP1&dq=bad+user+experience+causing+stress+among+users&ots=R9LhfvRYwN&sig=rUiDBOZeDJUax7KKrxBRqVWFzc&redir_esc=y#v=onepage&q&f=false

Ayata, D., Yaslan, Y., Kamasak, M. (2017) COMPARISON OF MACHINE LEARNING ALGORITHMS AND FEATURE EXTRACTION METHODS. Faculty of Computer and Informatics Engineering, Istanbul Technical University, Istanbul, Turkey. Retrieved from http://dergipark.gov.tr/download/article-file/272082

Bailey, P.B., Konstan, A.J. (July, 2006). On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. Computers in Human Behaviour. Volume 22, Issue 4. Pages 685-708. https://doi.org/10.1016/j.chb.2005.12.009

Billman, E.G. (20.februar, 2013). The LF/HF ratio does not accurately measure cardiac sympatho-vagal balance. https://doi.org/10.3389/fphys.2013.00026

Caldiroli, L.C., Garbo, R., Pallavicini, F., Antonietti, A., Mangiatordi, A and Mantovani, F. (2017). How Web Design influences User Experience: a Multi-modal Method for Real-Time Assessment during Web Browsing. 14th IEEE Annual Consumer Communications & Networking Conference (CCNC). Pages: 1063-1066.

Cinaz, B., Arnrich, B., La Marca, R., and Trøster, G. (2013). Monitoring of mental workload levels during an everyday life office-work scenario. Pers Ubiquit Comput (2013) 17:229–239 DOI 10.1007/s00779-011-0466-1. Retrieved from: http://doc.rero.ch/record/321225/files/779_2011_Article_466.pdf

Chen, F., Ruiz, N., Choi, E., Epps, J., Khawaja, M. A., Taib, R., Yin, B., and Wang, Y. (2013). Multimodal behavior and interaction as indicators of cognitive load. ACM Trans. Interact. Intell. Syst. 2, 4, Article 22 (January 2013), 36 pages. DOI=http://dx.doi.org/10.1145/2395123.2395127

Conaill, B., Frohlich, D. (1995). Timespace in the workplace: Dealing with interruptions. Mosaic Of Creativity. Retrieved from: https://interruptions.net/literature/O_Conail-CHI95-p262-o_conaill.pdf

Conway, D., Dick, I., Li, Z., Wang, Y., Chen, F. (2013). The Effect of Stress on Cognitive Load Measurement. 14th International Conference on Human-Computer Interaction (INTERACT), Sep 2013, Cape Town, South Africa. Springer, Lecture Notes in Computer Science, LNCS-8120 (Part IV), pp.659-666, 2013, Human-Computer Interaction – INTERACT 2013. <10.1007/978-3-642-40498-6_58>.

Chen, F., Ruiz, N.., Choi, E., Epps, J., Khawaja, A.M.,Taib, R., Yin, B., and Wang, Y.(2013). Multimodal behavior and interaction as indicators of cognitive load. ACM Trans. Interact. Intell. Syst. 2, 4, Article 22 (January 2013), 36 pages. DOI=http://dx.doi.org/10.1145/2395123.2395127

Czerwinski, M., Horvitz, E., Wilhite, S. (2004). A diary study of task switching and interruptions. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '04). ACM, New York, NY, USA, 175-182. DOI: https://doi.org/10.1145/985692.985715

Dockrill, P. (15.december, 2015). Stress could increase the risk of cognitive impairment preceding Alzeimer’s study finds. ScienceAlert. Retrieved from: https://www.sciencealert.com/stress-could-increase-the-risk-of-cognitive-impairment-preceding-alzheimer-s-study-finds

Dongsoo, K., Hyojin, K., Wooil, L., Minsu, K. (2014). Application and Limitation of Frequency Domain, LF/HF Component in Heart Rate Variability as an Acute Stress Index. Proceedings of the International Conference on Biomedical Engineering and Systems Prague, Czech Republic, August 14-15, 2014 Paper No. 128. Retrieved from: https://avestia.com/ICBES2014_Proceedings/papers/128.pdf

Elite HRV. (16.august, 2018). What are HRV Frequency Measurements (LF, HF, LF/HF). Retrieved from https://help.elitehrv.com/article/159-hrv-frequency

Elting, C., Zwickel, J., and Malaka, R. 2002. Device-dependant modality selection for user-interfaces: an empirical study. In Proceedings of the 7th international conference on Intelligent user interfaces (IUI '02). ACM, New York, NY, USA, 55-62. DOI: https://doi.org/10.1145/502716.502728

Friedman, L. T. (2006). The age of interruption. The New York Times. Retrieved from https://www.nytimes.com/2006/07/05/opinion/05friedman.html

Fuglerud, K.S., Halbach, T., Tjøstheim, I. (2015). Cost-benefit analysis of universal design (Report No. 1032). Retrieved from Norsk Regnesentral website: https://www.nr.no/sites/default/files/files/NR1032-Cost-benefit%20analysis%20of%20unversal%20design-final.pdf

Gjoreski, M. (2015). Stress Detection Using Smartphone and Wearable Sensors. Retrived from http://kt.ijs.si/markodebeljak/Lectures/Seminar_MPS/2012_on/Seminars_2014_15/Gjoreski%20Martin/Seminar1-StressDetection-Gjoreski-final.pdf

Goldstein. S.D., Bentho, O., Park, M-Y., and Sharabi, Y. (6.september, 2011) Low-frequency power of heart rate variability is not a measure of cardiac sympathetic tone but may be a measure of modulation of cardiac autonomic outflows by baroreflexes. Experimental Physiology. The physiological society. https://doi.org/10.1113/expphysiol.2010.056259

Greenfield, S. (25.januar, 2012). The internet and 'mind-change'. Youtube video. Retrieved from: https://youtu.be/ri4_CW9P41s

Gwizdka, J. (2010). Using stroop task to assess cognitive load. In Proceedings of the 28th Annual European Conference on Cognitive Ergonomics(ECCE '10). ACM, New York, NY, USA, 219-222. DOI=http://dx.doi.org/10.1145/1962300.1962345

Halarewich, D. (9. sept, 2016). Reducing Cognitive Load For A Better User Experience. Smashing Magazine. Retrieved from: https://www.smashingmagazine.com/2016/09/reducing-cognitive-overload-for-a-better-user-experience/

Halbach, T. (2010). Towards Cognitively Accessible Web Pages. Advances in Computer-Human Interactions, 2010, 19 - 24, DOI: 10.1109/ACHI.2010.34 IEEE Conference Publications

Harris, J.D., Vine, J.S., Wilson, R. M. (2017) 2.3.1 Psychophysiological Measures of Attentional Effort. Chapter 12: Neurocognitive mechanisms of the flow state. Progress in Brain Research. Volume 234. Pages 221-243. ScienceDirect.

Healey, J.A. and Picard, R.W. (2005). Detecting stress during real-world driving tasks using physiological sensors. Retreived november 2018 from https://affect.media.mit.edu/pdfs/05.healey-picard.pdf

Heathers, A.J. (7.may, 2014) Everything Hertz: methodological issues in short-term frequency-domain HRV. https://doi.org/10.3389/fphys.2014.00177

Heisler, Y. (2016). Mobile internet usage surpasses desktop usage for the first time in history. BGR Media. Retrieved from http://bgr.com/2016/11/02/internet-usage-desktop-vs-mobile/

Hill, L.R., Dickinson, A., Arnott, L.J., Gregor, P., and McIver, L. (2011). Older web users' eye movements: experience counts. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11). ACM, New York, NY, USA, 1151-1160. DOI: https://doi.org/10.1145/1978942.1979115

Hjortskov, N., Rissèn, D., Blangsted, K.A., Fallentin, N., Lundberg, U., and Søgaard, K. (2004). The effect of mental stress on heart rate variability and blood pressure during computer work. Eur J Appl Physiol (2004) 92: 84–89 DOI 10.1007/s00421-004-1055-z. Retrieved from:

https://www.researchgate.net/profile/Karen_Sogaard/publication/7449722_The_Effect_of_Mental_Stress_on_Heart_Rate_Variability_and_Blood_Pressure_During_Computer_Work/links/00b49533940a084363000000.pdf

Hye-Geum, K., Cheon, E-J., Bai, S-D., Lee, H-Y., and Koo, H-B. (2018). Stress and Heart Rate Variability: A Meta-Analysis and Review of the Literature. doi: 10.30773/pi.2017.08.17. Retrieved from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5900369/

Jiménez, E., Márquez, S., Moreno, F., Coret, J., Alcantud, F. (2012). Analysis of research on web usability for people with cognitive disability from 2002 to 2011. In Proceedings of the 13th International Conference on Interacción Persona-Ordenador (INTERACCION '12). ACM, New York, NY, USA, Article 25, 2 pages. DOI=http://dx.doi.org/10.1145/2379636.2379661

Kirsh, D. (2000). A Few Thought on Cognitive Overload. Intellectica, 2000/1, 30, pp. 19-51. Retrieved from: http://intellectica.org/SiteArchives/archives/n30/30_01_Kirsh.pdf

Klemm, R W. (29.dec, 2016). Thwart Stress Effects of Memory. Forced retrieval protects against stress-induced impairment of memory. Psychology Today. Retrieved from https://www.psychologytoday.com/us/blog/memory-medic/201612/thwart-stress-effects-memory

Kruger, J-l., Hefer, E., Matthew, G. (2013). Measuring the impact of subtitles on cognitive load: eye tracking and dynamic audiovisual texts. In Proceedings of the 2013 Conference on Eye Tracking South Africa (ETSA '13). ACM, New York, NY, USA, 62-66. DOI=http://dx.doi.org/10.1145/2509315.2509331

Levitin, J. D. (2015). Why the modern world is bad for your brain. Science. The Guardian. Retrieved from https://www.theguardian.com/science/2015/jan/18/modern-world-bad-for-brain-daniel-j-levitin-organized-mind-information-overload

Levitin, J. D. (23.september, 2015). Why it's so hard to pay attention, explained by science. Retrieved from: https://www.fastcompany.com/3051417/why-its-so-hard-to-pay-attention-explained-by-science

Lunn, D., and S, Harper, S. (2010). Using galvanic skin response measures to identify areas of frustration for older web 2.0 users. In Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A) (W4A '10). ACM, New York, NY, USA, Article 34, 10 pages. DOI: https://doi.org/10.1145/1805986.1806032

Lyu, Y., Luo, X., Yu, C. Z. J., Miao, C., Wang, T., Shi, Y., and Kameyama, I. K. (2015). Measuring Photoplethysmogram-Based Stress-Induced Vascular Response Index to Assess Cognitive Load and Stress. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15). ACM, New York, NY, USA, 857-866. DOI: https://doi.org/10.1145/2702123.2702399

Malamed, C. (2010). 20 Facts You Must Know About Working Memory. Cognition. The Elearning Coach. Retrieved from: http://theelearningcoach.com/learning/20-facts-about-working-memory/

Mark, G., Gudith, D., & Klocke, U. (2004). The Cost Of Interrupted Work: More Speed And Stress. Retrieved from: https://www.ics.uci.edu/~gmark/chi08-mark.pdf

Mark, G., Wang, Y., Niiya, M. (2014). Stress and multitasking in everyday college life: an empirical study of online activity. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14). ACM, New York, NY, USA, 41-50. DOI: https://doi.org/10.1145/2556288.2557361

McEwen, S.B., Sapolsky, M. R. (1995). Stress and cognitive function. Retrieved from https://pdfs.semanticscholar.org/4b4b/3edde5162ed5cc142f999be2cb1d818dc909.pdf

McGrew, K. S. (2005). The Cattell–Horn–Carroll theory of cognitive abilities. In D. P. Flanagan, & P. L. Harrison (Eds.), Contemporary intellectual assessment: Theories, tests, and issues (pp. 136−181)., 2nd ed. New York: Guilford Press. Retrieved from: https://onlinelibrary.wiley.com/doi/full/10.1002/9781118660584.ese0431

Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81-97. [Available at http://www.musanim.com/miller1956]

Molly Follette Story M.S. (1998). Maximizing Usability: The Principles of Universal Design, Assistive Technology, 10:1, 4-12, DOI: http://dx.doi.org/10.1080/10400435.1998.10131955

Morin, A. (unknown). Understood. Educational strategies: Universal Design for Learning (UDL): What You Need to Know. Retrieved from: https://www.understood.org/en/learning-attention-issues/treatments-approaches/educational-strategies/universal-design-for-learning-what-it-is-and-how-it-works

Morrison, B.B., Dorn, B., and Guzdial, M. (2014). Measuring cognitive load in introductory CS: adaptation of an instrument. In Proceedings of the tenth annual conference on International computing education research (ICER '14). ACM, New York, NY, USA, 131-138. DOI: http://dx.doi.org/10.1145/2632320.2632348

Murgia, M. (9.november, 2015). How stress affects your brain. Youtube video. Retrieved from https://youtu.be/WuyPuH9ojCE

Nakasone, A., Prendinger, H., and Ishizuka, M. (2005). Emotion Recognition from Electromyograhpy and Skin Conductance. Retrieved from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.104.4937&rep=rep1&type=pdf

Nielsen, J., and Norman, D. (unknown). Nielsen and Norman Group. The definition of User Experience (UX). Retrieved from: https://www.nngroup.com/articles/definition-user-experience/

Nielsen, J. (7.december, 2009). Short-Term Memory and Web Usability. Psychology and UX. Human-Computer-Interaction. Nielsen Norman Group. Retrieved from: https://www.nngroup.com/articles/short-term-memory-and-web-usability/

Niculescu, I.A., Nijholt, A.**Nie**, & Cao, Y. (2009). Manipulating Stress and Cognitive load in Conversational Interactions with a Multimodal System for Crisis Management Support. Human Media Interaction, University of Twente. Development of Multimodal Interfaces: Active Listening and Synchrony (pp.134-147) DOI: 10.1007/978-3-642-12397-9_11. Retrieved from: https://www.researchgate.net/publication/220716921_Manipulating_Stress_and_Cognitive_Load_in_Conversational_Interactions_with_a_Multimodal_System_for_Crisis_Management_Support

Nourbakhsh, N., Wang, Y., Chen, F., and Calvo, A. R. (2012). Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks. In Proceedings of the 24th Australian Computer-Human Interaction Conference (OzCHI '12), Vivienne Farrell, Graham Farrell, Caslon Chua, Weidong Huang, Raj Vasa, and Clinton Woodward (Eds.). ACM, New York, NY, USA, 420-423. DOI=http://dx.doi.org/10.1145/2414536.2414602

Nourbakhsh, N., Wang, Y., Chen, F., and Calvo, A. R. (2017). Detecting Users' Cognitive Load by Galvanic Skin Response with Affective Interference. ACM Trans. Interact. Intell. Syst. 7, 3, Article 12 (September 2017), 20 pages. DOI: https://doi.org/10.1145/2960413

Okoshi, T., Ramos, J., Nozaki, H., Nakazawa, J., K. Dey, A., Tokuda, H. (2015). Reducing users' perceived mental effort due to interruptive notifications in multi-device mobile environments. In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15). ACM, New York, NY, USA, 475-486. DOI: https://doi.org/10.1145/2750858.2807517

Oviatt, S. (2006). Human-centered design meets cognitive load theory: designing interfaces that help people think. In Proceedings of the 14th ACM international conference on Multimedia (MM '06). ACM, New York, NY, USA, 871-880. DOI: https://doi.org/10.1145/1180639.1180831

Paas, F., Renkl, A., Sweller, J. (2003), Cognitive Load Theory and Instructional Design: Recent Developments. Educational Psychologist, 38 (1), 1-4. Retrieved from:https://s3.amazonaws.com/academia.edu.documents/36183118/Educational_Psychologist_paas2.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1556737830&Signature=BZPHHZM912FgBEmE2eVG4d52

q14%3D&response-content
disposition=inline%3B%20filename%3DCognitive_load_theory_and_instructional.pdf

Penta, D.M., Tamburri, A.D. (2017). Combining quantitative and qualitative studies in empirical software engineering research. In Proceedings of the 39th International Conference on Software Engineering Companion (ICSE-C '17). IEEE Press, Piscataway, NJ, USA, 499-500. DOI: https://doi.org/10.1109/ICSE-C.2017.163

Ritthiron, S., and Jiamsanguanwong, A. (2017). Usability Evaluation of the University Library Network's Website Using an Eye-Tracking Device. In Proceedings of the International Conference on Advances in Image Processing (ICAIP 2017). ACM, New York, NY, USA, 184-188. DOI: https://doi.org/10.1145/3133264.3133294

Rello, L., Baeza-Yates, R. (2016). The Effect of Font Type on Screen Readability by People with Dyslexia. *ACM Trans. Access. Comput.* 8, 4, Article 15 (May 2016), 33 pages. DOI: http://dx.doi.org/10.1145/2897736

Rømen, D, Svanæs, D. (2010). Validating WCAG 1.0 and WCAG 2.0 through Usability Testing with Disabled Users. Unitech 2010.
Retrieved from: http://www.iu.hio.no/~frodes/unitech10/021-Romen/index.html

Sandnes.E, F., Lundh.V, M. (2015) Calendars for Individuals with Cognitive Disabilities: A Comparison of Table View and List View. In Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility (ASSETS '15). ACM, New York, NY, USA, 329-330. DOI=http://dx.doi.org/10.1145/2700648.2811363

Setz, C., Arnrich, B., Schumm, J., La Marca, R., Tröster, G., & Ehlert, U. (2010). Discriminating Stress From Cognitive Load Using a Wearable EDA Device. IEEE Transactions on Information Technology in Biomedicine. Volume 14, Issue 2. Pages 410-417. Retrieved from: https://ieeexplore-ieee-org.ezproxy.hioa.no/stamp/stamp.jsp?tp=&arnumber=5325784

Shaffer, F., & Ginsberg, Jp. (2017). An Overview of Heart Rate Variability Metrics and Norms. Frontiers in Public Health. 5. 258. 10.3389/fpubh.2017.00258. Retrieved from: https://www.researchgate.net/publication/320078994_An_Overview_of_Heart_Rate_Variability_Metrics_and_Norms

Shi, Y., Ruiz, N., Taib, R., Choi, E., and Chen, F. (2007). Galvanic skin response (GSR) as an index of cognitive load. In CHI '07 Extended Abstracts on Human Factors in Computing Systems (CHI EA '07). ACM, New York, NY, USA, 2651-2656. DOI: https://doi.org/10.1145/1240866.1241057

Shovon, H, I, Md., Nandagopal, D., Du, T. J., Vijayalakshmi, R., Cocks, B. (2015). Cognitive Activity during Web Search. In Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '15). ACM, New York, NY, USA, 967-970. DOI: http://dx.doi.org/10.1145/2766462.2767784

Shneiderman, B., and Plaisant, C. (2015). Designing the user interface. Strategies for effective human.computer-interaction. Fourth edition. Retrieved from: http://steinhardtapps.es.its.nyu.edu/create/courses/2015/reading/shneiderman1b.pdf

Sierra, de.S.A, Ávila, S.C., Casanova, G.J., Bailador del Pozo, G. (2011). Real-Time Stress Detection by Means of Physiological Signals. Recent Application in

Biometrics. Retrieved from:
http://cdn.intechopen.com/pdfs/17036/InTech-Real_time_stress_detection_by_means_of_physiological_signals.pdf

Sloan, R.P., Shapiro, P.A., Bagiella, E., Boni, S.M., Paik, M., and Bigger, J.T. (1994) Effect of mental stress throughout the day on cardiac autonomic control. Biol Psychol. 1994;37:89–99. Volume 32, Issue 2. DOI: https://doi.org/10.1016/0301-0511(94)90024-8

Tozman, T., Magdas, E.S., MacDougall, H.G., Vollmeyer, R. (2015). Understanding the psychophysiology of flow: A driving simulator experiment to investigate the relationship between flow and heart rate variability. Comput Hum Behav. 52, 408-418. DOI: https://doi.org/10.1016/j.chb.2015.06.023

Tracy, P. J., and Albers, J. M. (2006). Measuring Cognitive Load to Test the Usability of Web Sites. ResearchGate. Retrieved from: https://www.researchgate.net/profile/Michael_Albers/publication/253713707_Measuring_Cognitive_Load_to_Test_the_Usability_of_Web_Sites/links/55ef096608ae199d47bff6cd.pdf

Wang, C., Guo, J. (2019). A data-driven framework for learners' cognitive load detection using ECG-PPG physiological feature fusion and XGBoost classification, Procedia Computer Science, Volume 147, 2019, Pages 338-348, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2019.01.234. Retrieved from: http://www.sciencedirect.com/science/article/pii/S187705091930256X

Whitenton, K. (22.december, 2013). Minimize Cognitive Load to Maximize Usability. Nielsen Norman Group. Retrieved from:https://www.nngroup.com/articles/minimize-cognitive-load/

Wilmer, H.H., Sherman, E.L., and Chein, M.J. (25.april, 2017). Smartphones and Cognition: A Review of Research Exploring the Links between Mobile Technology Habits and Cognitive Functioning. doi: 10.3389/fpsyg.2017.00605. Retrieved from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5403814/

Yu, B., Funk, M., Hu, J., and Feijs, L. (2017). StressTree: a Metaphorical Visualization for Biofeedbackassisted Stress Management. Retrieved from: https://www.drhu.eu/publications/2017-DIS-StressTree.pdf

Zagermann, J., Pfeil, U., Reiterer, H. (2018). Studying Eye Movements as a Basis for Measuring Cognitive Load. In Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (CHI EA '18). ACM, New York, NY, USA, Paper LBW095, 6 pages. DOI: https://doi.org/10.1145/3170427.3188628

Zijlstra, H.R.F., Roe, A.R., Leonora, B.A., Krediet, I. (16.december, 2010). Temporal factors in mental work: Effects of interrupted activities. https://doi.org/10.1348/096317999166581

Appendix

Appendix A

```
#include <SPI.h>

#include <SD.h>

#include "Wire.h"


const uint8_t GSR = A0;

const uint8_t buttonPin = 2;     // the number of the pushbutton pin

const uint8_t ledPin =  13;      // the number of the LED pin

File file_GSR;


// Record ADC values if: adcValue < lowerThresholdValue || adcValue >
upperThresholdValue

uint16_t lowerThresholdValue = -1;

uint16_t upperThresholdValue = 0;


void AutoSetThresholds_MinMax() {

// Set lower threshold value to the lowest value found, and upper

// threshold to the highest value found.

const uint16_t numSamples = 500;

const uint8_t samplingPeriod = 5; // In milliseconds

lowerThresholdValue = -1;

upperThresholdValue = 0;


for (uint16_t i = 0; i < numSamples; ++i) {

uint16_t sample = analogRead(GSR);

if (sample > upperThresholdValue) {

upperThresholdValue = sample;}

if (sample < lowerThresholdValue) {

lowerThresholdValue = sample;} delay(samplingPeriod);}}
```

```
void AutoSetThresholds_LowPassFilteredMinMax() {
   // LP-filter by averaging samples. Set lower threshold value to the lowest
   // value found, and upper threshold to the highest value found.
const uint16_t numSamplesPerAverage = 50;
const uint16_t numAverages = 50;
const uint8_t samplingPeriod = 1; // In milliseconds
lowerThresholdValue = -1;
upperThresholdValue = 0;
for (uint16_t i = 0; i < numAverages; ++i) {
uint32_t average = 0;
for (uint16_t j = 0; j < numSamplesPerAverage; ++j) {
average += analogRead(GSR);}

average /= numSamplesPerAverage;
if (average > upperThresholdValue) {
upperThresholdValue = average;}
if (average < lowerThresholdValue) {
lowerThresholdValue = average;} delay(samplingPeriod);}
}
void AutoSetThresholds_Manual(){
Serial.flush();
Serial.println("Enter lower threshold");
lowerThresholdValue = Serial.parseInt();
Serial.println("Enter upper threshold");
upperThresholdValue = Serial.parseInt();
}
void AutosetThresholds() {
// Open a menu to select autoset method.
Serial.flush();
Serial.print(F("1.  Autoset using min/max sampled values.\n"));
```

```
Serial.print(F("2.  Autoset using min/max low-pass filtered sampled values.\n"));

while (!Serial.available());
char input = Serial.read();
Serial.flush();
Serial.print(F("\n\n"));
switch (input) {
case '1':
AutoSetThresholds_MinMax();
break;
case '2':
 AutoSetThresholds_LowPassFilteredMinMax();
break;
case '3':
AutoSetThresholds_Manual();
break;
default:
Serial.println(F("Invalid menu selection."));
break;
}

Serial.print(F("Autoset complete!\n"));
Serial.print(F("Lower threshold value: "));
Serial.print(lowerThresholdValue);
Serial.print(F("\n"));
Serial.print(F("Upper threshold value: "));
Serial.print(upperThresholdValue);
Serial.print(F("\n\n"));
}
void RecordValue(uint16_t adcValue, uint32_t timestamp) {
/*According to http://wiki.seeedstudio.com/Grove-GSR_Sensor/:
```

```
Human Resistance = ((1024+2*Serial_Port_Reading)*10000)/(512-Serial_Port_Reading),
unit is ohm, Serial_Port_Reading is the value display on Serial Port(between 0~1023*/
// adcValue=512 will cause division by zero
if (adcValue == 512) {
return;}
float resistance = ((1024 + 2 * adcValue) * 10000.0) / (512 - adcValue); // In ohms
float conductance = 1 / resistance; // In siemens
float conductanceMS = conductance * 1000000;

// Write to Serial
Serial.print(F("\n")); //carriage return for easier display
Serial.print(timestamp / 1000.0, 3);
Serial.print(F(" s:"));
Serial.print("ADC value: ");
Serial.print(adcValue);
Serial.print("\n");
Serial.print(timestamp / 1000.0, 3);
Serial.print(F(" s:"));
Serial.print(F("resistance = "));
Serial.print(resistance);
Serial.print(F("\n")); //carriage return for easier display
Serial.print(timestamp / 1000.0, 3);
Serial.print(F(" s: "));
Serial.print(F("conductance = "));//***this is the value you need to refer to for GSR
mapping***
Serial.print(conductanceMS); //serial printing conductance in micro-Siemens
Serial.print(F("\n")); //carriage return for easier display

// Write to SD-Card
file_GSR.print(timestamp / 1000.0, 3);
file_GSR.print(F(":"));
file_GSR.print(resistance);
```

```
file_GSR.print(F(":"));
file_GSR.println(conductanceMS);}

void StartRecordingValues() {
// Starts recording values. Stops recording when the button is pushed.
file_GSR = SD.open("test_GSR.txt", FILE_WRITE);

if (!file_GSR) {
Serial.println(F("Error opening test_GSR.txt"));
return;}
Serial.print("Started recording.\n\n");
const uint32_t interval = 500;
uint32_t startTime = millis();
uint32_t previousMeasurementTimestamp = startTime;
float minMeasuredValue = -1;

while (!digitalRead(buttonPin)) {
// We use a timer instead of delay() to prevent an unresponsive button:
if (millis() - previousMeasurementTimestamp >= interval) {
previousMeasurementTimestamp = millis();
uint16_t adcValue = analogRead(GSR);
if (adcValue <= lowerThresholdValue || adcValue >= upperThresholdValue) {
RecordValue(adcValue, previousMeasurementTimestamp - startTime);}
}}
file_GSR.close();}

void PrintRecordedValues() {
// Prints the contents of test_GSR.txt.
file_GSR = SD.open("test_GSR.txt", FILE_READ);
// If the file didn't open, print an error:
if (!file_GSR) {
Serial.println(F("Error opening test_GSR.txt"));
```

```
return;}
// Read from the file until there's nothing else in it:
while (file_GSR.available()) {
Serial.write(file_GSR.read());}
// close the file:
file_GSR.close();}

void EraseRecordedValues() {
SD.remove("test_GSR.txt");
Serial.print("Erased values.\n\n");}

void ClearSerialBuffer(){
while (Serial.available()){
Serial.read();}
}

void GetMainMenuInput() {
// Empty serial buffer:
ClearSerialBuffer();
Serial.print(F("1.  Autoset thresholds.\n"));
Serial.print(F("2.  Start recording values.\n"));
Serial.print(F("3.  Print recorded values.\n"));
Serial.print(F("4.  Erase recorded values.\n"));
// Wait for input:
while (!Serial.available());
char input = Serial.read();
Serial.print(Serial.peek());
ClearSerialBuffer();
Serial.print(F("\n\n"));
switch (input) {
case '1':
AutosetThresholds();
```

```
break;
case '2':
StartRecordingValues();
break;
case '3':
PrintRecordedValues();
break;
case '4':
EraseRecordedValues();
break;
default:
Serial.println(F("Invalid menu selection."));
break;}
}

void setup() {
// Initialize the pushbutton pin as an input:
pinMode(buttonPin, INPUT);
// Setting baud rate
Serial.begin(9600);
while (!Serial); // Wait for serial port to connect. Needed for native USB port only
Serial.print(F("Initializing SD card..."));
if (!SD.begin(4)) {
Serial.println(F("Initialization failed!"));
while (1);}
Serial.println(F("Initialization done."));
}
void loop() {
GetMainMenuInput();
}
```

Appendix B

Interviews

after each test session

1. **How stressed have you felt today? (1-5)**

1<----------------------------3---------------------------->5

2. **How stressed have you felt today compared to yesterday? (1-5)**

1<----------------------------3---------------------------->5

3. **How did you experience the test in general?**

4. **Did you feel any discomfort using the measuring equipment?**

5. **How was your experience of solving the tasks?**

6. **What do you think was the most challenging of the test?**

7. **Was there anything in the different UI's that affected you negatively during this test?**

   ____

   ____

   ____

8. **Was there anything in the UI's that could have been different (better, easier etc), given the situation you were in now?**

   ____

   ____

   ____

   ____

9. **What did work well in the UI? (features, design, structure etc)**

   ____

   ____

   ____

   ____

10. **Did you experience any differences in your interaction between the different UI's in test one and test two?**

    ____

    ____

    ____

    ____

    ____

11. **Did you experience any differences in your personal response between test one and test two?**

    ____

    ____

    ____

Appendix C

Questionnaire Stress detection

Gender:

Age:

Health Conditions:

Do you have any of the following diseases?

- Diabetes
- Heart or vascular disease
- High Blood Pressure
- High cholesterol
- Mental disorder (if so, what kind:______________________________________)
- Neither

Are you taking any medication at the moment that may affect your stress response? (anxiety pills, cardiac medications etc)

- No
- Yes:____

Are you currently suffering from depression or anxiety disorder?

- No
- Yes:____

Are you currently diagnosed with fatigue syndrome or burnout?

- No
- Yes

Have you recently experienced a major life change? (pleasant like a job promotion, studies or unpleasant like death in the family, break-up etc)

- No

- Yes:____

  ____

At this moment, which emotion state to you consider yourself to be in?

- worried
- sad
- depressed
- nervous
- anxiety
- happy
- excited
- cool
- neither of them
- another emotion state: ________________________

What does stress mean to you?

____

____

Please describe what scenarios that typically make you stressed? (e.g. time pressure on tasks, exams/other deadlines, loud noises, crowded places, claustrophobia or other phobias etc)

____

____

____

____

____

What is typically your reaction to a stressed situation?

psychological:

- irritated

- anxious
- confused
- oblivious
- calm
- controlled
- no particular reaction
- another reaction:____

physiological:

- Increased heartbeat
- sweaty
- jumpy
- hyperventilation
- stomach hurts
- butterflies in the stomach
- no particular reaction
- another reaction:____

What is your experiences of stress?

- Experience it daily
- Experience it several times a week
- Experience it at least once a week
- Experience it every now and then
- I rarely experience it
- I have never been stressed
- Other experiences:____

Appendix D

Declaration of consent.

Do you want to participate in our research project:
Universal design of smartphone user interface interaction in stressful situations

Aim and goals

____

Who is responsible for the research project?

____

Why are you asked to participate?

____

What does your participation mean to the project?

____

Is the participation voluntary?

____

Your privacy - how we store and use your information

____

What happens to all the information & data when the project is completed?

____

Your rights

____

What gives us the right to process personal information about you?

____

We process information about you based on your consent.

____

On behalf of OsloMet, The Norwegian Center for Research Data (NSD) AS has considered that the processing of personal data in this project is in accordance with the privacy regulations.

Where can you find more information about this project?

____

If you have any questions about the study, please contact:

Saja Andersson, project manager
Email: s198539@oslomet.no
Tlf:

Pietro Murano, supervisor
Email: Pietro.Murano@oslomet.no
Tlf:

NSD – Norsk senter for forskningsdata AS, email: (personvernombudet@nsd.no) or telephone: (+47) 55 58 21 17.

Best regards,

Project Manager

Declaration of consent

I have received and understood all the information about the project "Universal design of smartphone user interface interaction in stressful situations", and I consent to:

- to participate in the questionnaire
- to participate in the interviews
- to participate in two experiments (e.g user tests)
- that the project manager can give personal information about me to the project
- that my personal information can be stored and further used for the purpose of any other research or follow-up studies.

Hereby I consent to my information being processed until the project is completed, approximately by June 2019

(Signed by project participant, date)